CODEN: JASMAN

The Journal of the Acoustical Society of America

ISSN: 0001-4966

Vol. 125, No. 2 February 2009

JASA EXPRESS LETTERS

LETTERS TO THE EDITOR

AEROACOUSTICS, ATMOSPHERIC SOUND [28]

UNDERWATER SOUND [30]

(Continued)

CONTENTS—*Continued from preceding page*

(Continued)

CONTENTS—*Continued from preceding page*

(Continued)

CONTENTS—*Continued from preceding page*

A generalized Cramér-Rao lower bound for moving arrays

Edmund J. Sullivan
EJS Associates, 46 Lawton Brook Lane, Portsmouth, Rhode Island 02871
paddypriest@aol.com

Abstract: By properly including the forward motion of the array in the signal model, improved bearing estimation performance for a towed line array can be obtained. The improvement is a consequence of utilizing the bearing information contained in the Doppler. In this paper, it is shown by use of the Cramér-Rao lower bound that, as the array moves forward, the variance on the bearing estimate for an array of pressure sensors decreases, and that if an array of pressure-vector sensors is used, a significant improvement over that obtained for the array using pressure sensors only is obtained.

PACS numbers: 43.60.Fg, 43.30.Wi [JC]
Date Received: June 30, 2008 **Date Accepted:** October 8, 2008

1. Introduction

The fact that the passive synthetic aperture effect has been demonstrated, both theoretically and experimentally,[1,2] along with the emerging technology of the vector sensor, has presented the need to generalize the Cramér-Rao lower bound (CRLB) (Ref. 3) on the bearing estimator for the moving towed array.

The CRLB is the "best" (lowest) value that can be theoretically obtained for the variance of an estimate, and when an estimator attains this bound, it is said to be *efficient*. It is found as follows. Given $p(\mathbf{y}/\mathbf{x})$, the likelihood function for the measurement vector $\mathbf{y}=[y_1,y_2,\ldots,x_M]^T$, conditioned on the parameter vector $\mathbf{x}=[x_1,x_2,\ldots,x_N]^T$, with T indicating the transpose, the CRLB on the estimate of x_i, the ith element of $\mathbf{x}$, is given by

$$\sigma_{ii}^2 > (F^{-1})_{ii}, \tag{1}$$

where F is Fisher's information matrix, defined as

$$F_{ij} = -E\left\{\frac{\partial}{\partial x_i}\frac{\partial}{\partial x_j}\ln p(\mathbf{y}/\mathbf{x})\right\}. \tag{2}$$

The E denotes the expected value.

For a fixed line array with N equally spaced elements, and a signal model $p_n(\mathbf{x})$ with additive white Gaussian noise of variance σ^2 at the element level, the log likelihood function is given by

$$L = \ln p(\mathbf{y}/\mathbf{x}) = -\frac{1}{2\sigma^2}\sum_{n=0}^{N-1}[y_n - p_n(\mathbf{x})]^T[y_n - p_n(\mathbf{x})]. \tag{3}$$

If the estimator is unbiased, Eq. (3) results in

$$F_{ij} = -E\left\{\frac{\partial}{\partial x_i}\frac{\partial}{\partial x_j}L\right\} = E\left\{\frac{1}{\sigma^2}\sum_{n=0}^{N-1}\frac{\partial p_n^T}{\partial x_i}\frac{\partial p_n}{\partial x_j}\right\}. \tag{4}$$

2. The moving array

Consider a line array with elements on the x-axis of a Cartesian coordinate system. If the array is moving along the x-axis with speed $\pm v$, the narrow-band signal model for the pressure is

$$p_n(\omega_0,t) = A\cos((\omega_0/c)nd\sin\theta \pm \beta\omega_0 t), \tag{5}$$

with the angle θ measured clockwise positive from the y-axis, and

$$\beta = 1 \pm (v/c)\sin\theta. \tag{6}$$

This means that the source frequency ω_0 enters the problem as an additional unknown, so that there are now two unknowns to deal with. For this case, Eq. (4) takes the form

$$F_{\theta\theta} = \frac{1}{\sigma^2}\sum_{n=0}^{N-1}\frac{\partial p_n^T}{\partial\theta}\frac{\partial p_n}{\partial\theta},$$

$$F_{\theta f_o} = \frac{1}{\sigma^2}\sum_{n=0}^{N-1}\frac{\partial p_n^T}{\partial\theta}\frac{\partial p_n}{\partial f_o},$$

$$F_{f_o f_o} = \frac{1}{\sigma^2}\sum_{n=0}^{N-1}\frac{\partial p_n^T}{\partial f_o}\frac{\partial p_n}{\partial f_o}. \tag{7}$$

From Eq. (5), we have

$$\frac{\partial p_n}{\partial\theta} = -2\pi A(f_o/c)[nd \pm vt]\cos\theta\sin((\omega_0/c)nd\sin\theta + \beta\omega_0 t), \tag{8}$$

and

$$\frac{\partial p_n}{\partial f_o} = -2\pi A(1/c)[(nd \pm vt)\sin\theta + ct]\sin((\omega_0/c)nd\sin\theta + \beta\omega_0 t). \tag{9}$$

Substituting these into Eq. (7), it follows that

$$F_{\theta\theta} = \rho\sum_{n=0}^{N-1}(2\pi f_o/c)^2[nd \pm vt]^2\cos^2\theta,$$

$$F_{\theta f_o} = \rho\sum_{n=0}^{N-1} f_o(2\pi/c)^2[(nd \pm vt)^2\sin\theta + (nd \pm vt)ct]\cos\theta,$$

$$F_{f_o f_o} = \rho\sum_{n=0}^{N-1}(2\pi/c)^2[(nd \pm vt)^2\sin^2\theta + 2(nd \pm vt)ct\sin\theta + (ct)^2], \tag{10}$$

where $\rho = (A^2/2\sigma^2)$ is the element-level signal-to-noise ratio.

Since the results of Eq. (10) are time dependent, there are two approaches one could take. The CRLB could simply be evaluated at some chosen time, or it could be averaged over the observation time T. To be more specific would require that some information on the estimation procedure to be used be known beforehand. Here, we wish to compare the moving array to the stationary array, avoiding any specific estimation scheme. In the spirit of simplicity then, we choose to compute the time average.

We begin by recalling that $L = (N-1)d$ is the physical aperture, and make the following definitions:

$$R = vT/L, \tag{11}$$

$$K = 2\rho(2\pi/c)^2, \tag{12}$$

$$\left(\frac{1}{T}\right)\int_0^T dt\left\{\sum_{n=0}^{N-1}[nd \pm vt]^2\right\} = L^2\left[\frac{N}{6}\frac{(2N-1)}{(N-1)} \pm NR + \frac{1}{3}(N-1)R^2\right] = L^2X(R), \quad (13)$$

$$\left(\frac{1}{T}\right)\int_0^T dt\left\{\sum_{n=0}^{N-1}(nd \pm vt)ct\right\} = L^2\left(\frac{c}{v}\right)\left[\frac{N}{4}R \pm \frac{1}{3}(N-1)R^2\right] = L^2Y(R), \quad (14)$$

$$\left(\frac{1}{T}\right)\int_0^T dt\left\{\sum_{n=0}^{N-1}(ct)^2\right\} = L^2\frac{1}{3}\left(\frac{c}{v}\right)^2(N-1)R^2 = L^2Z(R). \quad (15)$$

In the last three equations we have used the following identities:

$$\sum_{n=0}^{N-1} n = (N/2)(N-1), \quad (16)$$

$$\sum_{n=0}^{N-1} n^2 = (N/6)(2N-1)(N-1). \quad (17)$$

R is the ratio of the virtual aperture, i.e., the aperture traced out by the motion of the array, to the physical aperture. Using the definitions in Eqs. (13)–(15), Eqs. (10) become

$$F_{\theta\theta} = KL^2f_o^2\cos^2\theta X(R),$$

$$F_{\theta f_o} = KL^2f_o\cos\theta[X(R)\sin\theta + Y(R)],$$

$$F_{f_of_o} = KL^2[X(R)\sin^2\theta + 2Y(R)\sin\theta + Z(R)]. \quad (18)$$

We now have the Fisher matrix elements in a form convenient for numerical evaluation.

3. Classical CRLB calculation for the moving array

The CRLB on the bearing estimate follows from Cramer's rule as

$$\sigma_\theta^2 > (F^{-1})_{11} = \frac{1}{F_{\theta\theta} - (F_{\theta f_o}^2/F_{f_of_o})}. \quad (19)$$

It is easy to show that for the stationary array, the CRLB is given by

$$\sigma_\theta^2 > (F^{-1})_{11} = \frac{1}{F_{\theta\theta}}, \quad (20)$$

with $R=0$.

The results are shown in Fig. 1, where the standard deviation of the bearing estimate is shown as a function of R, where R is defined in Eq. (11). Here, it can be seen that when jointly estimating the bearing and source frequency, as opposed to the bearing alone, the moving array underperforms[4] the stationary array. This result was first shown by Stergiopoulos[5] and later, in a more general sense, by Edelson.[6] It is interesting to note that, at early times (i.e., small R), the down-Doppler case underperforms the up-Doppler case. As we shall see in the next section, to obtain improved performance, the source frequency must be known *a priori* to a reasonably high accuracy.

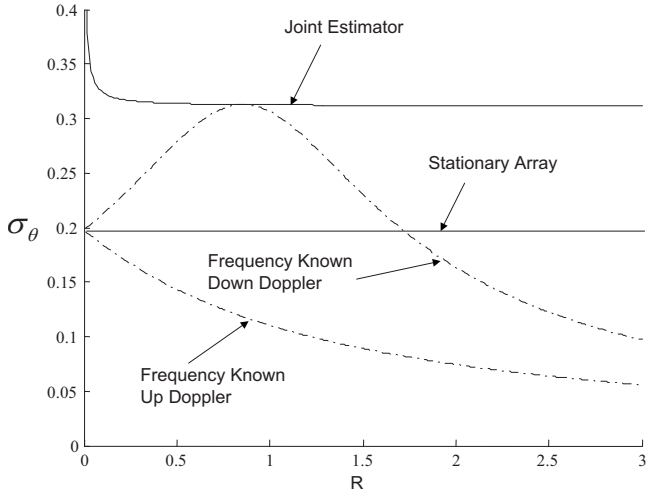


Fig. 1. Comparison of the classical CRLB for a two-element stationary array to that of the same array when it is moving for the case of zero (broadside) bearing. The spacing is $\lambda/2$, the frequency is 300 Hz, and the speed of the array is 2.5 m/s. σ_θ is in radians and the element level SNR is −10 dB. Here, the CRLB calculation implies that jointly estimating the bearing and source frequency is not sufficient for obtaining the passive synthetic aperture effect, and it will actually underperform the stationary array in this case.

4. Bayesian CRLB calculation for the moving aray

The Bayesian CRLB is defined as the expected value of the *joint* probability of the measurements and the model parameters. This is in contrast to the definition in Eq. (2) for the classical CRLB, where the conditional probability density function is used. Thus, replacing $p(\mathbf{y}/\mathbf{x})$ with $p(\mathbf{y}/\mathbf{x})p(\mathbf{x})$ in Eq. (2), the relevant Fisher matrix for the Bayesian evolves as

$$F_B = F_C + F_P = \begin{bmatrix} (F_C)_{11} & (F_C)_{12} \\ (F_C)_{21} & (F_C)_{22} \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 0 & (F_P)_{22} \end{bmatrix}. \tag{21}$$

Here, the elements of F_C are identical to those given in Eqs. (18) and F_P is referred to as the "prior" Fisher matrix.[7] Thus, the Bayesian Fisher matrix is the sum of the classical Fisher matrix and the prior Fisher matrix. In order to evaluate this, we first observe that assuming a Gaussian prior on the frequency leads to

$$(F_P)_{22} = 1/\sigma_f^2. \tag{22}$$

For this case, Eq. (19) generalizes to

$$\sigma_\theta^2 > (F^{-1})_{11} = \frac{1}{(F_C)_{11} - ((F_C)_{12}^2/[(F_C)_{22} + 1/\sigma_f^2])}. \tag{23}$$

The results for the Bayesian CRLB are shown in Fig. 2, where we see that the inclusion of prior information has a highly significant effect.

5. The moving pressure-vector sensor array

The jump from the pressure sensor only case to the pressure-vector (PV) sensor case introduces a significant increase in the complexity of the problem. This arises from the fact that the covariance matrix between the sensors is quite complex, even in the Gaussian case. Not only is the matrix not diagonal in the case of $\lambda/2$ spacing, but it also depends upon the parameters of interest. Thus, Eq. (4) generalizes to[3]

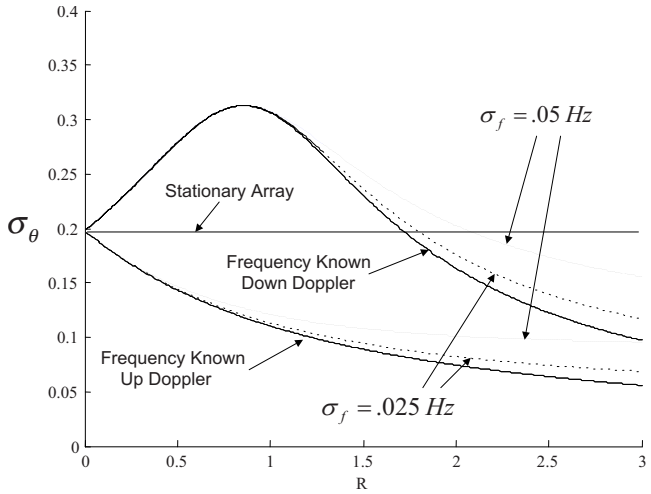


Fig. 2. (Color online) Comparison of the Bayesian CRLB for a two-element moving array with that of the case for the frequency known *a priori* for several values of the variance of the Gaussian prior. The bearing is at broadside. The spacing is $\lambda/2$, the frequency is 300 Hz, and the speed of the array is 2.5 m/s. σ_θ is in radians and the element level SNR is −10 dB. The impact of prior knowledge of the source frequency is clear.

$$F_{ij} = \left[\frac{\partial \mathbf{P}}{\partial \theta}\right]^T R^{-1}\left[\frac{\partial \mathbf{P}}{\partial \omega}\right] + 1/2tr\left[R^{-1}\frac{\partial \mathbf{R}}{\partial \theta}R^{-1}\frac{\partial \mathbf{R}}{\partial \omega}\right]. \tag{24}$$

Here, **P** is the signal vector. We stay with a two-element array, and in order to further reduce the complexity of the problem, we use a two-dimensional vector sensor in the $x-y$ plane. Thus, the signal vector is 6×1 and is given by

$$\mathbf{P} = [p_1 u_1^p u_1^n p_2 u_2^p u_2^n]^T. \tag{25}$$

In Eq. (25), the subscript labels the element and the superscripts n and p indicate normal and parallel to the array to the array, where the array is on the x-axis. The signal is arriving in the $x-y$ plane and the angle θ is measured positive clockwise from the y-axis. The narrow-band signals at the outputs of the receiver elements have the following form.[9]

$$p_1(\omega_0, t) = \cos(\beta\omega_0 t), \tag{26}$$

$$u_1^n(\omega_0, t) = \cos(\beta\omega_0 t)\cos(\theta), \tag{27}$$

$$u_1^p(\omega_0, t) = \cos(\beta\omega_0 t)\sin(\theta), \tag{28}$$

$$p_2(\omega_0, t) = \cos((\omega_0/c)nd\,\sin\,\theta + \beta\omega_0 t), \tag{29}$$

$$u_2^n(\omega_0, t) = \cos((\omega_0/c)nd\,\sin\,\theta + \beta\omega_0 t)\cos(\theta), \tag{30}$$

$$u_2^p(\omega_0, t) = \cos((\omega_0/c)nd\,\sin\,\theta + \beta\omega_0 t)\sin(\theta). \tag{31}$$

As before,

$$\beta = 1 \pm (v/c)\sin\,\theta. \tag{32}$$

The covariance matrix for the spherically isotropic noise case is constructed from the interelement correlations. These are given by Naulai and Lauchle,[10] as

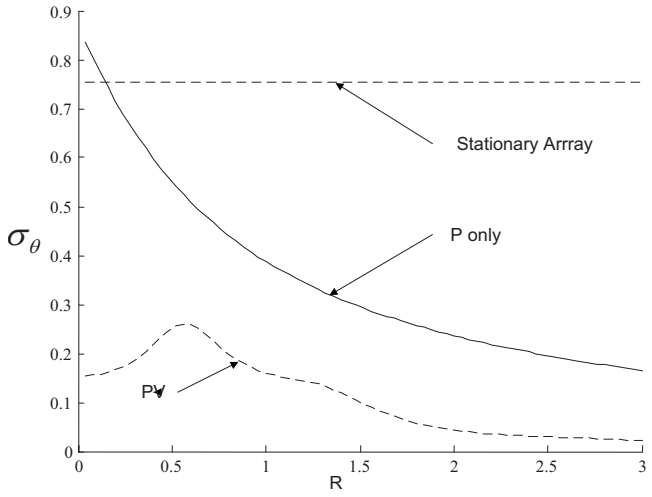


Fig. 3. Comparison of the Bayesian CRLB For the pressure-only case to that of the pressure-vector case for a two-element array. This is an up-Doppler case for an array speed of 2 m/s. The frequency is 150 Hz, the element spacing is $\lambda/2$, the SNR is +3 dB, and the bearing is 75°. The horizontal dotted line is the result for a stationary array of pressure sensors only. σ_θ is in degrees.

$$R_{ij}^{pp} = P\frac{\sin(\omega\alpha)}{\omega\alpha}\cos(\tau), \tag{33}$$

$$R_{ij}^{pu^p} = \frac{P}{\sqrt{3}}\frac{\sin(\omega\alpha)-(\omega\alpha)\cos(\omega\alpha)}{(\omega\alpha)^2}\sin(\tau), \tag{34}$$

$$R_{ij}^{u^pu^p} = \frac{P}{3}\frac{[(\omega\alpha)^2-2]\sin(\omega\alpha)-2(\omega\alpha)\cos(\omega\alpha)}{(\omega\alpha)^3}\cos(\tau), \tag{35}$$

$$R_{ij}^{u^nu^n} = \frac{P}{3}\frac{\sin(\omega\alpha)-(\omega\alpha)\cos(\omega\alpha)}{(\omega\alpha)^3}\cos(\tau). \tag{36}$$

Here, $\alpha=d/c$, P is the noise power at a given location on the array,[11] and $\tau=\omega\alpha\sin(\theta)$. Thus, we see that the covariance matrix is a function of both ω and θ. Based on the order of the signals in Eq. (25), the covariance matrix has the following form:

$$R = \begin{bmatrix} P & 0 & 0 & PR_{14}^{pp} & (P\sqrt{3})R_{15}^{pu^p} & 0 \\ 0 & P/3 & 0 & (P/\sqrt{3})R_{24}^{pu^p} & (P/3)R_{25}^{u^pu^p} & 0 \\ 0 & 0 & P/3 & 0 & 0 & (P/3)R_{36}^{u^nu^n} \\ PR_{41}^{pp} & (P/\sqrt{3})R_{42}^{pu^p} & 0 & P & 0 & 0 \\ (P/\sqrt{3})R_{51}^{pu^p} & (P/3)R_{52}^{u^pu^p} & 0 & 0 & P/3 & 0 \\ 0 & 0 & (P/3)R_{63}^{u^nu^n} & 0 & 0 & P/3 \end{bmatrix} \tag{37}$$

Since the complexity of the problem does not lend itself to a closed-form solution, we show the results based on a numerical evaluation of Eqs. (21)–(24). Figure 3 shows an up-Doppler case for a bearing of 75°, and Fig. 4 shows a down-Doppler case for a bearing of 45°.

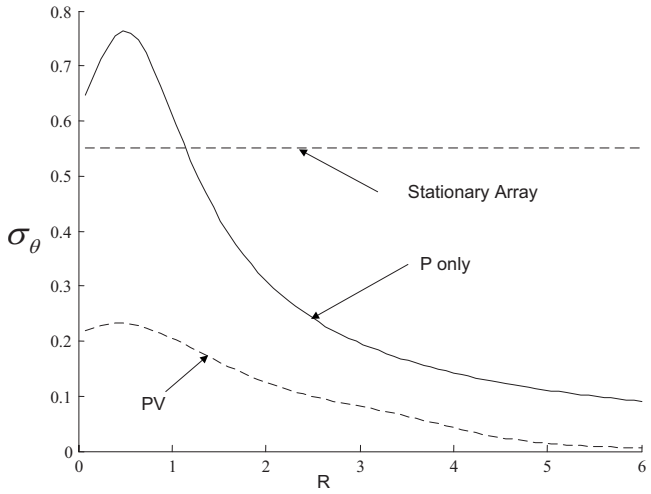


Fig. 4. Comparison of the Bayesian CRLB For the pressure-only case to that of the pressure-vector case for a two-element array. This is a down-Doppler case for an array speed of 2 m/s. The frequency is 150 Hz, the element spacing is $\lambda/4$, the SNR is +3 dB, and the bearing is 45°. The horizontal dotted line is the result for a stationary array of pressure sensors only. σ_θ is in degrees.

In both cases, the variance of the Gaussian prior was taken to be the observed frequency. Although these examples are by no means exhaustive, they clearly exhibit two things. A passive synthetic aperture can provide a significant enhancement to towed array processing, and in those cases where flow noise can be avoided (e.g., short arrays on AUVs), the vector-sensor synthetic aperture will provide a significant improvement over the pressure-only case.

References and links

[1]E. J. Sullivan and J. V. Candy, "Space-time array processing: A model-based approach," J. Acoust. Soc. Am. **102**, No. 5, 2809–2820 (1997).
[2]E. J. Sullivan, J. D. Holmes, W. M. Carey, and J. F. Lynch, "Broadband passive synthetic aperture: Experimental results," J. Acoust. Soc. Am. **120-EL**, 49–52 (2006).
[3]S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory* (Addison-Wesley, Reading, MA, 1991), Chap. 3.
[4]It should be pointed out that in practice, the joint estimate can still outperform the stationary array. This indicates that the CRLB is a highly conservative measure.
[5]S. Stergiopoulos, "Optimum bearing resolution for a moving towed array and extension of its physical aperture," J. Acoust. Soc. Am. **87**(5), 2128–2140 (1990).
[6]G. S. Edelson, "On the Estimation of Source Location Using a Passive Towed Array," Ph.D. dissertation, University of Rhode Island, 1993.
[7]Since we are assuming that the bearing is deterministic, it has no contribution to the prior Fisher matrix. Bell and Van Trees (Ref. 8) refer to this case as the "hybrid" case.
[8]K. L. Bell and H. L. Van Trees, "Posterior Cramér-Rao Bound for Tracking Target Bearings," in Proceedings of the Adaptive Sensor Array Processing Workshop, MIT Lincoln Labs, 7–8 June (2005).
[9]The velocity signals do not explicitly contain the acoustic impedance, since we assume that the measured velocities have already been scaled by the impedance, thus converting them to their equivalent pressure units.
[10]N. K. Naulai and G. C. Lauchle, "Acoustic intensity methods and their applications to vector sensor use and design," Penn State Report. No. 2006-01 (2006).
[11]The factors of 3 and $\sqrt{3}$ are a consequence of the inherent directivity of the vector sensors.

Melody recognition by two-month-old infants

Judy Plantinga[a] **and Laurel J. Trainor**
McMaster University, Hamilton, Ontario L8S 4K1, Canada
judy.plantinga@utoronto.ca, ljt@mcmaster.ca

Abstract: Music is part of an infant's world even before birth, and caregivers around the world sing to infants. Yet, there has been little research into the musical abilities or preferences of infants younger than 5 months. In this study, the head turn preference procedure used with older infants was adapted into an eye-movement preference procedure so that the ability of 2-month-old infants to remember a short melody could be tested. The results show that with minimal familiarization, 2-month-old infants remember a short melody and can discriminate it from a similar melody.

PACS numbers: 43.75.Cd, 43.66.Hg, 43.66.Mk, 43.66.Lj [QJF]
Date Received: October 8, 2008 **Date Accepted:** December 4, 2008

1. Introduction

Although 6-month-old infants do not demonstrate knowledge of culturally specific characteristics of either musical pitch structure (Lynch *et al.*, 1990; Trainor and Trehub, 1992) or metrical structure (Hannon and Trehub, 2005), their perception of melody is adult-like in many ways. They can remember musical pieces for weeks (Ilari and Polka, 2006; Trainor *et al.*, 2004; Saffran *et al.*, 2000), discriminate single note changes to a short melody (e.g., Trehub *et al.*, 1985; Trainor and Trehub, 1992), and recognize melodies in transposition (e.g., Chang and Trehub, 1977; Trainor and Trehub, 1992; Plantinga and Trainor, 2005). The presence of these skills early in development suggests that humans begin life with a predisposition to process music. However, little is known about the music processing abilities of infants younger than 5 months, although experience with sound begins before birth and very young infants are exposed to music through infant-directed singing, the musical qualities of infant directed speech, and music in the general environment (from television, radio, etc.). The ability of 2-month-old infants to remember a short melody immediately following familiarization was tested in this study.

At birth or soon after, infants appear to have the perceptual capabilities required to process the pitch and timing information necessary for the perception of music. With respect to pitch, by 35 weeks gestational age (GA), the fetus responds to pure tones at frequencies ranging from 100 to 3000 Hz (Hepper and Shahidullah, 1994). At birth infants can discriminate upward from downward pitch contours (Carral *et al.*, 2005). By 2 months after birth, infants show cortical EEG responses to a change of one semitone in the pitch of piano tones (He *et al.*, 2007). By 3 months, frequency resolution approaches that of adults for all but high frequencies (Werner and VandenBos, 1993) and is finer than is required for musical purposes (Trehub, 2001). With respect to timing, at 2 months infants can discriminate tempo changes to an isochronous tone sequence (Baruch and Drake, 1997) and can discriminate simple rhythmic patterns (Demany *et al.*, 1977).

Further evidence that in the first months after birth infants have the prerequisites to perceive music comes from auditory research using speech stimuli. At birth, infants are sensitive to pitch contours in speech and can use contour information to categorize words (Nazzi *et al.*, 1998). Neonates can use prosodic information (characteristic pitch and rhythm patterns) to distinguish between their native language and an unfamiliar one with different prosodic structure (Mehler *et al.*, 1988; Moon *et al.*, 1993). Third trimester fetuses

[a]Judy Plantinga is now at the University of Toronto, Mississauga, Ontario, Canada.

(DeCasper *et al.*, 1994) and neonates (DeCasper and Spence, 1986) can recognize a poem or story that their mothers read repeatedly during the last trimester of pregnancy.

The present study examines whether 2-month-old infants can remember a brief melody after a short familiarization of 15 repetitions. A conditioned head-turn procedure is often used to test older infants, but 2-month-olds do not have good control of head movements (Kemler-Nelson *et al.*, 1995). For the present study, an eye-movement preference procedure was developed in which infants controlled how long they listened to the familiarized melody and how long they listened to a novel melody through their eye movements rather than their head movements. Memory for the familiarized melody was expected to manifest as a preference for that melody over the novel melody on the basis of previous studies with speech stimuli (De Casper *et al.*, 1994; De Casper and Spence, 1986; Moon *et al.*, 1993; Mehler *et al.*, 1988).

2. Method

2.1 Participants

Sixteen healthy full-term infants (eight females, eight males) between 2 and 3 months of age (average age 84 days) with no known hearing impairments participated in the study. Data from another 11 infants were not used due to low correlation between raters' judgments of looking behavior (six) or because the infants did not finish testing due to fussiness (five).

2.2 Stimuli and apparatus

The stimuli consisted of the first phrase of each of two old English folk songs, "Country Lass" and "Painful Plough," following studies using these as stimuli in 6-month-olds (Trainor *et al.*, 2004; Plantinga and Trainor, 2005). Both songs are simple in structure, but unlikely to be familiar to the infants, and both are in a similar folk song style. The excerpts were equated for number of notes (14) and playing time (5.8 s). The songs differ in meter (6/8 and 4/4) and mode (G major and G minor).

The stimuli were produced using the acoustic piano instrumentation in the Cakewalk program on a personal computer with a Sound Blaster AWE64 Gold sound card, and recorded using Cool Edit. For familiarization and testing, the sounds were presented by a Macintosh G5 computer, an NAD C352 stereo integrated amplifier, and two audiological GSI speakers. The speakers were located inside a large sound attenuating booth (Industrial Acoustics Co.).

2.3 Procedure

After the procedure was explained to the parent and a consent form was signed, the parent and the infant were taken into the sound-attenuating booth. The parent placed the infant in a car seat which was facing a 23-in. Apple flat computer screen. A camera located under the computer screen was connected to a Macintosh G4 computer so that the experimenters outside the booth could view an image of the infant. The parent was seated behind the infant in the booth so that he/she could remain with the infant, and wore headphones over which masking music was played during the entire procedure. Once the parent and infant were settled and comfortable, the experimenter left the booth and closed the door. During familiarization, the infant was presented with 15 repetitions of either the Country Lass phrase or the Painful Plough phrase, with random assignment to this familiarization melody. Then an eye movement preference procedure was used for testing the infant's memory. An animated stimulus appeared in the middle of the computer screen in front of the infant to attract the infant's attention. Two observers watched the infant on the computer screen outside the booth and independently judged where the infant was looking. Each observer had a keypad that was connected to the Macintosh G5 computer and was unaware of the response of the other observer. Once both observers judged that the infant was looking at the center of the screen (indicated independently with a key press on each key pad), the animated stimulus disappeared and a still visual target appeared on either the left or the right side of the computer screen. When both observers indicated by pressing a second key on each pad that the infant was looking at the visual target, the music for that trial began. Once both observers indicated that the infant had looked away (by releasing the key for more than 1 s), the

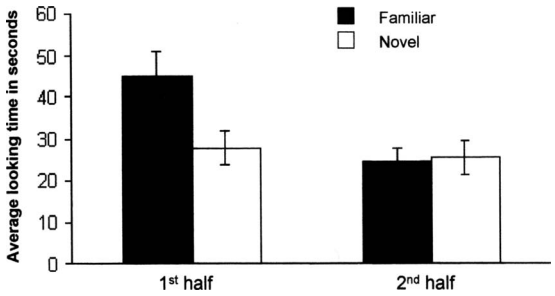


Fig. 1. Infants preferences as measured by the amount of time they chose to listen to the familiar compared to the novel melody. In the first half of the trials, infants preferred the familiar compared to the novel melody, indicating that they recognized the familiar melody. In the second half of the trials, looking times decreased overall, and the preference disappeared, suggesting that habituation had occurred.

trial ended, the music stopped, and the visual target disappeared. The animated stimulus reappeared in the center of the screen and the procedure was repeated, with the visual target appearing on the opposite side of the monitor screen from the previous trial. The infant was presented with the familiar melodic phrase on trials with the visual target on one side and the novel melodic phrase on trials with the visual target on the other side. First song (novel, familiar) and first side of presentation (left, right) were counterbalanced across participants. The observers were unable to hear the sounds, so they were unaware of which song was being presented on any particular trial. The experiment ended when 5 min had elapsed from the start of testing. The computer kept track of the time that each observer pressed the key for each trial, calculated the time the infant spent listening to each stimulus on each trial (the longer of the two observers' times), and calculated the correlation between the two raters' judgments of looking times over the experiment. The correlation between raters' judgments was required to be greater than 0.8 for the data from an infant to be included in the final sample.

3. Results

Infant preference shifts from familiar to no preference to novel with increased exposure to a stimulus (Rose *et al.*, 1982). To investigate whether this shift occurred over the trials of the testing, the looking time for each half of the total number of trials for each infant was computed and an ANOVA, with novel/familiar melody and test half as independent measures and mean listening times as the dependent measure, was performed. It revealed a significant effect of half, $F(1\ 15)=6.46$, $p=0.02$, with listening times decreasing from the first to second half. The only other effect was a significant interaction of novel/familiar × half, $F(1,15)=13.26$, $p=0.003$ (Fig. 1). Dependent-sample t-tests on novel and familiar listening times for each half showed that listening times to the familiar phrase were significantly greater than listening times to the novel phrase in the first half, $t(15)=3.42$, $p=0.004$, but not in the second half, $t(15)=0.21$, $p=0.83$ (Fig. 1). The disappearance of the preference in the second half is likely because by this time the "novel" phrase had been heard a number of times during the testing and was therefore no longer novel. Indeed, listening times decreased significantly from the first to second half, suggesting that the melodies were becoming less interesting in general with repetition.

4. Discussion

The prevalence across human cultures of singing to infants and talking to infants using "musical" infant-directed speech has been noted many times (e.g., Trehub, 2001; Trehub and Trainor, 1998; Trehub and Schellenberg, 1995; Trainor and Schmidt, 2003). The results of the present study show that infants as young as 2 months remember and discriminate a familiar from a novel melody after minimal exposure, so they are sensitive to the sequential pattern information in melodies.

The eye-movement preference methodology developed here provides a means with which to explore a number of critical aspects of infant musical processing. For example, after the minimal exposure to the melodies of the present study, infants showed a familiarity preference. However, after sufficient exposure, young infants would be expected to demonstrate a novelty preference, as is the case with older infants (Plantinga and Trainor, 2005). Determining the amount of exposure needed for familiarity and novelty preferences between 2 and 6 months of age would allow exploration of the development of musical memory across this age span.

Another critical question concerns relative pitch processing (i.e., the pitch distances between tones as opposed to the absolute pitches of the tones). A number of studies have shown that infants 6 months and older readily process relative pitch in that they recognize melodies when transposed to higher or lower pitch levels (Chang and Trehub, 1977; Plantinga and Trainor, 2005; Trainor and Trehub, 1992). At the same time, it has been suggested that only absolute pitch is available early in life, and that relative pitch develops with exposure to music (Saffran and Griepentrog, 2001; Sergeant and Roche, 1973). In order to understand the development of melodic pitch processing, it is therefore critical to test when infants are able to recognize melodies in transposition.

In summary, the present study developed a new methodology for testing discrimination in infants younger than 5 months of age and used this methodology to show that infants as young as 2 months can remember a short melody and can discriminate between the familiar and a novel melody.

Acknowledgments

This research was supported by a National Science and Engineering Research Council Graduate Scholarship to J.P. and a National Science and Engineering Research Council Grant to L.J.T. The authors thank Janice Wright for her help in testing participants.

References and links

Baruch, C., and Drake, C. (**1997**). "Tempo discrimination in infants," Infant Behav. Dev. **20**, 573–577.
Carral, V., Huotiainen, M., Ruusuvirta, T., Fellman, V., Näätänen, R., and Escera, C. (**2005**). "A kind of auditory "primitive intelligence" already present at birth," Eur. J. Neurosci. **21**, 3201–3204.
Chang, H. W., and Trehub, S. E. (**1977**). "Auditory processing of relational information by young infants," J. Exp. Child Psychol. **24**, 324–331.
DeCasper, A. J., Lecanuet, J. P., Busnel, M. C., Granier-Deferre, C., and Maugeais, R. (**1994**). "Fetal reactions to recurrent maternal speech," Infant Behav. Dev. **17**, 159–164.
DeCasper, A. J., and Spence, M. J. (**1986**). "Prenatal maternal speech influences newborns' perception of speech sounds," Infant Behav. Dev. **9**, 133–150.
Demany, L., McKenzie, B., and Vurpillot, E. (**1977**). "Rhythm perception in early infancy," Nature (London) **266**, 718–719.
Hannon, E. E., and Trehub, S. E. (**2005**). "Metrical categories in infancy and adulthood," Psychol. Sci. **16**, 48–55.
He, C., Hotson, L., and Trainor, L. J. (**2007**). "Mismatch responses to pitch changes in early infancy," J. Cogn Neurosci. **19**, 878–892.
Hepper, P. G., and Shahidullah, B. S. (**1994**). "Development of fetal hearing," Arch. Dis. Child **74**, F81–F87.
Ilari, B., and Polka, L. (**2006**). "Music cognition in early infancy: Infants preferences and long-term memory for Ravel," Int. J. Music Educ. **24**, 7–20.
Kemler Nelson, D. G., Jusczyk, P., Mandel, D. R., Myers, J., Turk, A., and Gerken, L. A. (**1995**). "The head-turn preference procedure for testing auditory perception," Infant Behav. Dev. **18**, 111–116.
Lynch, M. P., Eilers, R. E., Oller, D., and Urbano, R. C. (**1990**). "Innateness, experience, and music perception," Psychol. Sci. **1**, 272–276.
Mehler, J., Jusczyk, P., Lambertz, G., Halsted, N., Bertoncini, J., and Amiel-Tison, C. (**1988**). "A precursor of language acquisition in young infants," Cognition **29**, 143–178.
Moon, C., Cooper, R. P., and Fifer, W. P. (**1993**). "Two-day-olds prefer their native language," Infant Behav. Dev. **16**, 495–500.
Nazzi, T., Floccia, C., and Bertoncini, J. (**1998**). "Discrimination of pitch contours by neonates," Infant Behav. Dev. **21**, 779–784.
Plantinga, J., and Trainor, L. J. (**2005**). "Memory for melody: Infants use a relative pitch code," Cognition **98**, 1–11.
Rose, S. A., Gottfried, A. W., Melloy-Carminar, P., and Bridger, W. H. (**1982**). "Familiarity and novelty preferences in infant recognition memory: Implications for information processing," Dev. Psychol. **18**, 704–713.
Saffran, J. R., and Griepentrog, G. J. (**2001**). "Absolute pitch in infant auditory learning: Evidence for

developmental reorganization," Dev. Psychol. **37**, 74–85.
Saffran, J. R., Loman, M. M., and Robertson, R. R. W. (**2000**). "Infant memory for musical experiences," Cognition **77**, B15–B23.
Sergeant, D., and Roche, S. (**1973**). "Perceptual shifts in the auditory information processing of young children," Psychol. Music **1**, 39–48.
Trainor, L. J., and Schmidt, L. A. (**2003**). "Processing emotions induced by music," in *The Cognitive Neuroscience of Music*, edited by I. Peretz and R. Zatorre (Oxford University Press, Oxford), pp. 310–324.
Trainor, L. J., and Trehub, S. E. (**1992**). "A comparison of infants' and adults' sensitivity to Western musical structure," J. Exp. Psychol. **18**, 394–402.
Trainor, L. J., Wu, L., and Tsang, C. D. (**2004**). "Long-term memory for music: Infants remember tempo and timbre," Dev. Sci. **7**, 289–296
Trehub, S. E. (**2001**). "Musical predispositions in infancy," Ann. N.Y. Acad. Sci. **930**, 1–16.
Trehub, S. E., and Trainor, L. J. (**1998**). "Singing to infants: Lullabies and playsongs," Adv. Infancy Res. **12**, 43–77.
Trehub, S. E., and Schellenberg, E. G. (**1995**). "Music: Its relevance to infants," in *Annals of Child Development Vol. II*, edited by R. V. Vlasta (Jessica Kingsley Publishers, New York), pp. 1–24.
Trehub, S. E., Thorpe, L. A., and Morrongiello, B. A. (**1985**). "Infants' perception of melodies: Changes in a single tone," Infant Behav. Dev. **8**, 213–223.
Werner, L. A., and VandenBos, G. R. (**1993**). "Developmental psychoacoustics: What infants and children hear," Hosp. Community Psychiatry **44**, 624–626.

Transient motion of a circular plate after an impact

Thomas R. Moore, Daniel W. Zietlow, Christopher W. Gorman, Donald C. Griffin, Connor P. Ballance, and David J. Parker
Department of Physics, Rollins College, Winter Park, Florida 32789
tmoore@rollins.edu, dzietlow@rollins.edu, cgorman@rollins.edu, griffin@vanadium.rollins.edu, ballance@vanadium.rollins.edu, dparker@rollins.edu

Abstract: The transient response of a flat circular plate to a sudden impact has been studied experimentally and theoretically. High-speed electronic speckle pattern interferometry reveals the presence of pulses that travel around the edge of the plate ahead of the bending motion initiated by the strike. It is found that the transient motion of the plate is well described by Kirchhoff thin-plate theory over a time approximately equal to the time required for the initial impulse to circumvent the plate; however, a more sophisticated model is required to describe the motion after this time has elapsed.

PACS numbers: 43.40.Kd, 43.40.At, 43.40.Dx [JM]
Date Received: August 21, 2008 **Date Accepted:** December 4, 2008

1. Introduction

Studies of the motion of flat plates have been ongoing for many years. Typically, the interest has been in determining the frequencies and deflection shapes of the normal modes of plates, and there currently exists a firm theoretical and experimental basis for predicting the resonances of plates of various common shapes.[1] However, the transient response of impacted plates, even those with common geometries such as rectangular and circular, is still not well understood.

The struck flat plate is a common occurrence in a large number of situations, including those found in industry, the military, and music. There currently exists a small body of theoretical work on the transient response of impacted plates, and a few researchers have reported simulations of infinite, rectangular, and circular plates.[2–4] Reports of experimental work in this area are even more limited and are generally confined to acoustic measurements.[5,6] To our knowledge there are only two reports on the agreement between the predicted and actual deflection shapes of an impacted plate as a function of time,[7,8] and no reports of such a comparison where the effects of the shape of the plate or boundary conditions at the edges have been considered.

Here we report on an investigation of the transient response of a struck flat circular plate using high-speed electronic speckle pattern interferometry. We compare the results with the predictions of the Kirchhoff model of a thin plate, which was numerically solved using a finite-difference scheme. We show that there is initially excellent agreement between the model and the experiment; however, once the impulse has had sufficient time to traverse the plate the predictions of the model significantly diverge from the observed deflection of the plate.

2. Experiments

To investigate the deflection of a plate as it is being struck, a flat circular plate was held firmly at the center and left free at the edges. The plate was placed in the object plane of an electronic speckle pattern interferometer with the capability of high-speed acquisition. The light source for the interferometer was a 5 W solid-state laser with a wavelength of 532 nm. Configured in this manner, the interferometer produced an image containing fringes of equal displacement, where each fringe on the interferogram represents an increase in the out of plane deflection of 266 nm. The camera used for imaging had a resolution of 320×240 pixels and an acquisition rate of 33 057 frames per second.

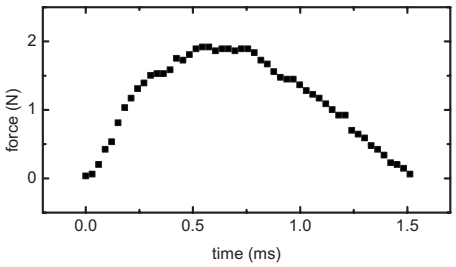


Fig. 1. Plot of the force exerted on the plate as a function of time.

The plate under consideration was a circular brass plate of diameter 134.0±0.1 mm and thickness 4.7±0.1 mm. It was held rigidly in the center by a screw that was secured to a 1.5 in. diameter aluminum post, which was in turn secured to an optical table. The plate was secured to the post to provide a rigid support, but it was offset from the post by a washer with a radius of 0.25 in. The edges of the plate were free to vibrate.

The plate was struck from behind at a point approximately 90% of the radius from the center by a 2.1 mm diameter screw attached to a load sensor. The strike lasted approximately 1.5 ms and had a maximum force of approximately 1.9 N. The output from the force sensor is shown in Fig. 1.

Figure 2 contains the series of interferograms produced during the strike, each representing the deflection of the plate approximately 30 μs after the previous one. The plate was struck from behind in the upper right corner. It is important to note that the fringe traversing the center of the plate represents the line of zero deflection; the point of maximum deflection is at the edge nearest point of impact and is displaced approximately 1.73 μm toward the observer in the final image.

Interferograms such as those shown in Fig. 2 are somewhat difficult to interpret since motion out of the plane of the interferogram creates the same image as motion into the plane. This ambiguity can usually be resolved by following the evolution of the lines of zero deflection. Figure 3 shows one image taken from Fig. 2 with several of the fringes annotated with the magnitude of the deflection at that point. Deflection toward the observer is labeled as being positive while deflection away from the observer is designated as being negative. To assist in interpreting the interferograms, fringes representing zero deflection are annotated by white arrows in Fig. 2.

An inspection of Fig. 2 shows that while the strike is occurring the deflection of the plate increases at the point of the strike as one would expect; however, the pulse due to the

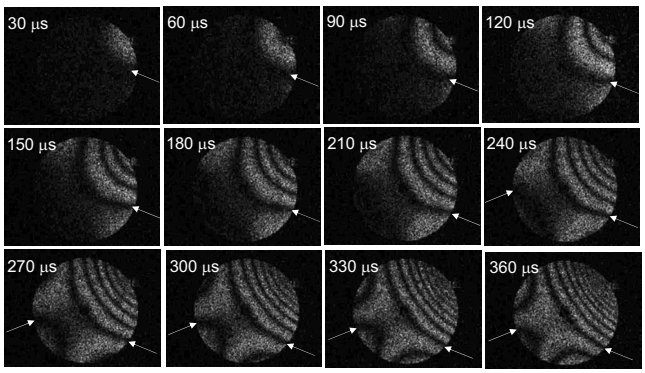


Fig. 2. Interferograms of the plate during the strike. The numbers in the upper left indicate the time since the beginning of the impact. The white arrows designate fringes of zero displacement.

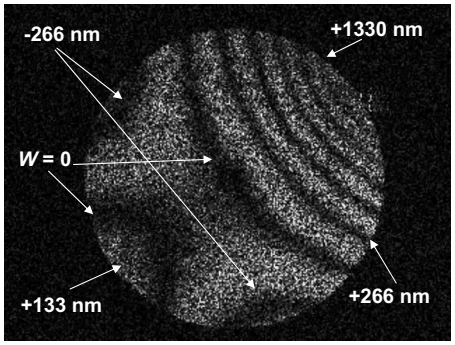


Fig. 3. One of the interferograms shown in Fig. 2 with several of the fringes labeled with the magnitude of the deflection. Motion toward the observer is noted as being positive while motion away from the observer is negative.

impact is transferred around the circumference of the plate without traversing the fringe of zero displacement. That is, the portion of the plate being driven directly by the impulse is being displaced toward the observer, but this displacement gives rise to symmetric pulses that are preceding the outward bending motion. These precursor pulses are characterized by displacement into the plane of the image and are being transferred around the circumference of the plate ahead of the outward displacement. The result is that the diametrical fringe representing zero displacement never traverses past the center of the plate.

By 90 μs after the initiation of the strike, the two symmetric pulses propagating around the circumference of the plate that have displacements into the plane of the interferogram are clearly visible. Surprisingly, these pulses lead to a third pulse, displaced out of the plane of the interferogram, which appears at a position diametrically opposite to the striking point. As will be shown later, eventually the two pulses that are displaced away from the observer coalesce, forcing the entire lower left half of the plate to be displaced into the plane of the interferogram. When this happens the two semicircular portions of the plate are displaced out of phase with one another, with the nodal line being through the center of the plate.

3. Modeling

It is reasonable to ask if the complex dynamics of the struck plate described above can be accurately modeled by a simple theory. Given the dimensions of the plate, one would expect that it can be considered to be thin and therefore Kirchhoff thin-plate theory should be adequate for the task. To test this assertion we modeled the plate using Kirchhoff thin-plate theory and solved the resulting differential equation explicitly using the method of finite differences.

Kirchhoff thin-plate theory describes the displacement w of a plate by the fourth-order differential equation[1]

$$-D\nabla^4 w(r,\phi,t) + p(r_0,\phi_0,t) - R\frac{\partial w(r,\phi,t)}{\partial t} = \sigma\frac{\partial^2 w(r,\phi,t)}{\partial t^2}, \tag{1}$$

where σ is the mass per unit area of the plate, R is a damping coefficient that is proportional to the velocity (and is negligible in this calculation), and $p(r_0,\phi_0,t)$ is the applied force per unit area at the point of impact. The factor D is the flexural rigidity and is given by

$$D = \frac{Eh^3}{12(1-\nu^2)}, \tag{2}$$

where E is the elastic modulus of the plate, h is the thickness and ν is Poisson's ratio.

While it is logical to use the method of finite differences to solve Eq. (1), the circular symmetry and the complex boundary conditions on the free edge of the plate make this a difficult problem to solve accurately. A Cartesian mesh does not fit the symmetry of the problem;

therefore, to properly apply the boundary conditions one must employ cylindrical coordinates. However, if one employs a uniform radial mesh, the area enclosed between adjacent radial and angular mesh points increases with radius; this tends to decrease the accuracy of the solution, especially on the outer edge. To overcome this difficulty, we employed a nonlinear radial mesh that varies inversely with the radial coordinate:

$$dr = \frac{d\rho}{r}. \tag{3}$$

With this change of coordinates, the term involving ∇^4 can be written as

$$\nabla^4 w = 8\frac{\partial^2 w}{\partial \rho^2} + 16\rho\frac{\partial^3 w}{\partial \rho^3} + 4\rho^2\frac{\partial^4 w}{\partial \rho^4} + \frac{1}{4\rho^2}\frac{\partial^4 w}{\partial \phi^4} + \frac{1}{\rho^2}\frac{\partial^2 w}{\partial \phi^2} + 2\frac{\partial^4 w}{\partial \rho^2 \partial \phi^2}. \tag{4}$$

The boundary conditions at the center of the plate are merely that the displacement and its first derivative must be identically zero. The boundary conditions at the free edge, however, are significantly more complicated since there the bending, shear, and twisting moments must all be zero. Additionally, to implement the finite difference scheme for any point on the mesh it is necessary to know the displacements of each of the nearest neighbors as well as the displacements of the points twice removed from that point. If the mesh ends at the edge of the plate, it is impossible to implement the finite difference scheme. Therefore, two virtual rings were added to the mesh and the displacements of the points on those rings were determined so that the boundary conditions at the actual edge of the plate were satisfied. The two virtual rings have no physical significance and are not included in the results, but they do allow one to determine finite difference solutions to Eq. (1) and accurately enforce the free-edge boundary conditions. The mesh used in all of the work reported here had 200 radial and 180 angular points. To achieve a stable solution with this spatial mesh it was necessary to use a time step of 2×10^{-9} s. The driving force $p(r_0, \phi_0, t)$ was modeled with a portion of a sine function scaled to approximate the time-dependent force shown in Fig. 1.

To test the validity of the model the predicted normal-mode frequencies for an annular plate with an inner radius equal to 30% of the outer radius were compared to the analytically calculated values found using Ref. 1. The plate was assumed to be clamped on the inner radius and free on the outer edge. The variation between the analytical and numerical calculations ranged from 0.2% to 5.3%, with the average variation being approximately 2%. These differences are similar to those found when comparing the results using the different analytic solutions presented in Ref. 1.

Once the model had been validated, it was used to determine the time-resolved displacement predicted by the Kirchhoff theory for the plate used in the experiment described above. After the computation was complete, the calculated displacement at each point of the plate was converted into equivalent quarter-wavelengths of the illuminating light and plotted as a gray-scale from black (even integer multiples of quarter-wavelengths) to white (odd integer multiples). Using this method of display, the results of the simulation appeared similar to an interferogram of the deformed plate. The results of the simulation are found in Mm. 1, where the actual interferogram appears on the left and the results of the simulation appear on the right. The time elapsed since the strike is shown in the lower right corner.

[Mm. 1 Video of the interferograms beside the results of the simulation of the struck plate. The time elapsed since the strike appears in the lower right corner.] This is a file of type "avi" (19.8 Mb).

4. Discussion

The evolution of the transient response of the plate shown in Fig. 2 and Mm. 1 is surprising in many ways, as is the fact that it can be accurately modeled with such a simple theory. One important result of the modeling effort is that any ambiguity in the interferograms as to the direction of motion of the plate can be unequivocally determined due to the excellent agreement

between the theory and experiment. The results confirm the assertion above that the first precursor pulses are indeed out of phase with the driving motion, and the nodal lines do not represent a simple minimum in the displacement of the plate. That is, the impact is pushing the top of the plate toward the observer but the displacement changes direction across nodal lines. Therefore, the first precursor pulses are actually displacements away from the observer while the one that appears diametrically opposite to the point of impact is displaced toward the observer.

There are a number of striking features in the results presented above that deserve comment, but clearly the most surprising feature is the evolution and propagation of the precursor pulses. The eventual rocking motion of the plate about a nodal diameter that occurs after approximately one millisecond is expected, but the manner in which this motion evolves is not. The presence of a precursor to the displacement directly attributable to an impact has been both predicted and observed in plates before the wave reaches the boundaries;[2,4,7] however, the nature of the dynamics of this pulse as it propagates around the edge of a circular plate is fundamentally different. In this case it is the edge of the plate that primarily determines how the pulses propagate. Furthermore, the fact that the precursor pulses lead to an initial displacement of the plate in the same direction as the impact, but at a position diametrically opposite to the point of impact, is counterintuitive. This motion is quickly counteracted by the arrival of the first two precursor pulses that are directed into the plane of the interferogram (in the opposite direction of the impulse), which cause the plate to then bend into the plane. The result is that eventually the plate flexes about the center line in a rocking motion.

The presence of a precursor with a displacement opposite to the motion directly created by the impact has previously been posited to be attributable to the dispersion of the flexural waves in a solid; however, the time evolution of the pulse shown in Fig. 2 and Mm. 1 demonstrates that this is not the case. The higher frequency waves do indeed have a higher speed in the plate, but these precursor pulses cannot be attributed to this phenomenon. The fact that the pulses are seen to precede the deflection that can be directly attributed to the motion caused by the strike, without traversing the line of zero-displacement, indicates that they are not attributable simply to the dispersive nature of the bending waves. Furthermore, a comparison of these results with Fig. 1 shows that these pulses propagate completely around the plate before the driving force has reached its maximum. Thus, the top of the plate is still moving toward the observer as these pulses propagate around the plate.

Instead of being attributable to dispersion, these results indicate that the precursor pulses are attributable to the stiffness of the plate and are a consequence of the fact that a positive displacement at one point in a stiff medium induces a displacement in the opposite direction at points close to it. The bending eventually produces a second-order effect that creates a third pulse; however, this also occurs while the initial impact is still in progress and the displacement does not traverse the nodal line. Therefore, this motion cannot be attributed to the dispersive nature of the material either.

The evolution of the motion of the plate also shows clearly that the impulse is not traveling around the plate, nor is the whole-body motion immediately induced about the central nodal line. Instead, a displacement is being induced in the plate at regions far from the point of the strike, which eventually evolves into the rocking whole-body motion characteristic of the (1,0) mode. One may safely assume that a larger plate would result in additional nodal lines and hence the number of induced pulses is not as important as the mere fact that they exist and propagate around the edge of the plate.

The excellent agreement between the model and the experiment is also of considerable interest. It has been recently reported by others that explicitly solving the Kirchhoff equations numerically for the case of a circular thin plate can result in unrealistic solutions.[9] Specifically, they cite low-frequency noise and a high-frequency cut-off in the solution. A comparison of the actual and predicted results of the struck plate presented here reveals that indeed there is poor agreement between this model and the experimentally observed motion in the long-time limit ($\geqslant$1 ms). When the simulation is allowed to progress beyond approximately 1 ms the deflection shapes predicted by the model are strikingly different than those observed. (The beginning of this lack of agreement can be observed in the last few images of Mm. 1.) Furthermore, the

predicted average power spectrum is quite different from the measured power spectrum; the predicted frequencies of the lower resonances do not agree with the actual result and the relative power in each mode is poorly predicted.

It is possible to approximately calculate the eigenfrequencies of the plate analytically by assuming that the ratio of the outer to inner radii is 0.100 rather than the actual ratio of 0.103. The necessary parameters are given in Ref. 1. When the analytically calculated frequencies of the first five modes are compared with those predicted by the numerical simulation it is found that they agree well, indicating that the differences between the model and the experiment are not due to numerical error. However, when these values are compared with the actual resonance frequencies, the lowest frequencies are found to be significantly different, with the lowest predicted frquency larger than the measured value by more than a factor of 3. Despite the differences in the predicted resonance frequencies, the excellent agreement between the model and the experimental results reported here leaves little doubt that Eq. (1) describes the initial dynamics of the plate quite well.

To resolve the discrepancies between the predictions of the model and the physical response of the plate in the long-time limit it will probably be necessary to include a frequency-dependent damping term in Eq. (1). Chainge and Lambourg have implemented such a method for a suspended square plate,[6] but the assumptions made in their work do not appear to apply to the case of a circular plate clamped in the center. Measurements of free rectangular plates reported in Ref. 6 showed that the damping constants are small and of a similar order of magnitude below some cutoff frequency and are much larger with little variation above this frequency. To determine if there is a similar relationship between frequency and damping in a clamped circular plate we experimentally determined the exponential damping constant for five of the modes of the circular plate used in these experiments. It was found that this constant can vary from less than unity to over 30, with no obvious systematic relationship between the frequency and the damping.

5. Conclusion

The work reported here demonstrates that the transient response of a circular flat plate to sudden impact has some interesting and surprising characteristics. The impact creates precursor pulses that propagate around the outer edge of the plate, eventually meeting at the point diametrically opposite to the point of impact. These pulses are evidently not due to the dispersive nature of the bending waves because they occur before the wave motion actually begins and they do not traverse the nodal line. Rather, they appear spontaneously ahead of the motion directly attributable to the impact and well before the impact point has time to complete a half-cycle of the motion.

This unusual transient motion can be accurately predicted by explicitly solving the Kirchhoff plate equations using the method of finite differences. A nonlinear mesh assists in the numerical solution and overcomes some of the difficulties cited in the past. However, the Kirchhoff thin-plate model does not accurately predict the long-term motion of the plate. Indeed, after the initial precursor pulses have reached the opposite side of the plate and coalesced, the motion is not well described by the model. It is probable that producing an accurate time-resolved model of a struck plate will require the inclusion of frequency dependent damping terms; however, the simple model of Eq. (1) appears to be sufficient to predict the initial transient response of the system to an off-center impact.

Acknowledgments

This work was supported by the National Science Foundation Grant No. 0551310, the John Hauck Foundation, and the Rollins College Student-Faculty Collaborative Scholarship Program.

References and links

[1]A. Leissa, *Vibrations of Plates* (Acoustical Society of America, Melville, NY, 1993).
[2]M. El-Raheb, "Flexural waves in a disk soon after impact," J. Acoust. Soc. Am. **96**, 221–234 (1994).
[3]M. El-Raheb and P. Wagner, "Transient flexural waves in a disk and square plate from off-center impact," J.

Acoust. Soc. Am. **110**, 2991–3002 (2001).
[4]C. Lambourg, A. Chaigne, and D. Matignon, "Time-domain simulation of damped impacted plates II. Numerical model and results," J. Acoust. Soc. Am. **109**, 1433–1447 (2001).
[5]A. Wahlin, P. Gren, and N.-E. Molin, "On structure-borne sound: Experiments showing the initial transient acoustic wave field generated by an impacted plate," J. Acoust. Soc. Am. **96**, 2791–2797 (1994).
[6]A. Chaigne and C. Lambourg, "Time-domain simulation of damped impacted plates I. Theory and experiments," J. Acoust. Soc. Am. **109**, 1433–1447 (2001).
[7]K. Fallstrom, H. Gustavsson, N. Molin, and A. Wahlin, "Transient bending waves in plates studied by hologram interferometry," Exp. Mech. **29**, 378–387 (1989).
[8]T. Mazuch and J. Trnka, "Bending waves in an elastic plate generated by ultra short impulse," Acta Tech. CSAV **47**, 293–303 (2002).
[9]K. Arcas, A. Chaigne, and S. Bilbao, "Sound synthesis of circular plates by finite differences," J. Acoust. Soc. Am. **123**, 3522 (2008).

Lanczos iterated time-reversal

Assad A. Oberai
Mechanical Aerospace and Nuclear Engineering, Rensselaer Polytechnic Institute, Troy, New York 12180
oberaa@rpi.edu

Gonzalo R. Feijóo
Applied Ocean Physics & Engineering, Woods Hole Oceanographic Institution, Woods Hole, Massachusetts 02543
gfeijoo@whoi.edu

Paul E. Barbone
Department of Mechanical Engineering, Boston University, Boston, Massachusetts 02215
barbone@bu.edu

Abstract: A new iterative time-reversal algorithm capable of identifying and focusing on multiple scatterers in a relatively small number of iterations is developed. It is recognized that the traditional iterated time-reversal method is based on utilizing power iterations to determine the dominant eigenpairs of the time-reversal operator. The convergence properties of these iterations are known to be suboptimal. Motivated by this, a new method based on Lanczos iterations is developed. In several illustrative examples it is demonstrated that for the same number of transmitted and received signals, the Lanczos iterations based approach is substantially more accurate.

PACS numbers: 43.60.Tj, 43.20.−f, 43.60.Lq [JC]
Date Received: August 29, 2008 **Date Accepted:** December 2, 2008

1. Introduction

Time-reversal methods are often used to target inhomogeneities or image scatterers in an acoustic medium.[1,2] It is being actively pursued in medical ultrasonic imaging, in medical treatment via ultrasound, in underwater communication, in nondestructive evaluation of materials and structures using both linear and nonlinear elastic waves, in imaging the earth's crust, and imaging its oceans.[3–7,2]

The basic idea in time-reversal may be explained in the context of an array of transmitters and receivers embedded in an inhomogeneous medium. A source emits a pulse which is multiply scattered, dispersed, and eventually detected by the receivers. The received signal is then reversed in time and transmitted back into the medium. The reversed signal retraces its multiple paths, and medium dispersion recompresses the signal, so that the pulse later focuses at the original source point. In *iterated time-reversal* this process of transmission, time-reversal, and retransmission is repeated and with each iteration the transmitted signal selectively focuses on the strongest scatterer in the medium.

Iterated time-reversal can be interpreted as the power iteration method applied to the time-reversal operator H, which is related to the scattering operator G through $H=GG^T$, where G^T is the transpose of G in the time domain (see below for precise definitions). An element G_{ij} of the discrete scattering operator represents the scattered signal measured at the ith element of an array due to a unit impulse transmitted by the jth element. Each iteration in the time-reversal method corresponds to one iteration of the power method; that is the evaluation of the product Hv for a signal v. The power iterations converge to the eigenvector corresponding to the largest eigenvalue of H (which is the same as the singular vector corresponding to the largest singular value of G). This vector is the time-reversed scattered field corresponding to the strongest scatterer, restricted to the measurement array. When this field is transmitted, it focuses on the strongest scatterer and this property can be used to selectively transmit energy or information to the scatterer. To locate the scatterer one only needs to search for the maximum of this transmitted

field. When the time signal is time-harmonic with frequency ω, the operation of time-reversal of a signal reduces to taking the conjugate of its complex representation and time-reversal operator is related to scattering operator via $H=GG^*$, where G^* is the complex conjugate of G.

The time-reversal iterations described above are effective when there is one dominant scatterer in the medium. Their usefulness is limited when multiple scatterers of comparable strength are present. In the case of multiple scatterers the iterations converge to the strongest scatterer at a rate that is proportional to the ratio of the strength of the second-strongest scatterer to the strongest scatterer. Clearly, in the case of multiple scatterers of similar strength, this convergence can be poor. Further, the recovery of weaker scatterers is cumbersome and requires a large number of iterations, even when their strengths are very different.[8] The slow convergence of iterated time-reversal can be problematic in applications.[9] In a temporally (slowly) changing environment, iterated time-reversal is particularly useful to adaptively optimize focusing to the current conditions. Such adaptation is impossible, however, if the convergence time is not significantly shorter than the medium correlation time. Therefore, accelerating the convergence of iterated time-reversal is key to realizing its full potential in applications.

The key to accelerating this process is the recognition that the transducer array can be used as a kind of analog computer. Each time-reversal operation can be interpreted as a matrix-vector product. Thus advanced iterative methods from linear algebra can be implemented in a time-reversal context.

When compared with power iterations the convergence properties of methods based on Lanczos iterations are superior, especially when multiple eigenvalues and eigenvectors are desired.[10] As we explain below the estimate from Lanczos is always at least as good as that from the power method and is typically far superior. In addition, for multiple eigenvalues the power method requires a complicated multi-step procedure.[8]

Motivated by these observations we describe a new iterated time-reversal method based on Lanczos iterations that inherits their advantages over power iterations. Through numerical examples we demonstrate here that when compared with the power iterations based method it requires fewer iterations to converge to an accurate answer. This observation directly translates to fewer transmit and receive cycles in an experimental setting. We also note that the experimental realization of the proposed method requires only a slight modification of the current protocol for time-reversal and no modification to the actual hardware. It is worth reemphasizing that the method proposed here and described below offers rapid convergence to multiple eigenvectors with few transmission cycles. The method *does not* depend upon measuring the entire scattering operator.

2. Iterated time-reversal

Time-reversal concepts are conveniently described in terms of scattering and time-reversal operators. Consider a surface in $\mathbb{R}^3$ denoted by Γ that is continuously embedded with transmitters and receivers (discrete transmitters and receivers may be treated identically). In the following treatment, we consider the standard development in time-reversal using single-frequency harmonic signals. Thus the transmitted and received signals are functions of $\boldsymbol{x} \in \Gamma$ and frequency ω. The dependence of signals and operators on the frequency is implicit, so ω is dropped in the notation. We require that signals be square-integrable, that is they are functions $v \in \mathcal{V} \equiv L_2(\Omega)$. We denote the L_2 inner product on Γ by $(u,v)=\int_\Gamma u^* v \, d\boldsymbol{x}$ and the corresponding norm by $\|\cdot\|$. The kernel of the scattering operator of the problem is denoted by $g(\boldsymbol{x};\boldsymbol{x}')$, which is the Green's function corresponding to the appropriate wave propagation problem. It represents the response at a location $\boldsymbol{x}$ due to a unit impulse at $\boldsymbol{x}'$. The received signal, corresponding to a transmitted signal $v(\boldsymbol{x})$, is given by

$$r(\boldsymbol{x}) = \int_\Gamma g(\boldsymbol{x};\boldsymbol{x}')v(\boldsymbol{x}')d\boldsymbol{x}' . \tag{1}$$

The relation above may be written succinctly as $r(\boldsymbol{x})=G[v(\boldsymbol{x})]$, by defining the scattering operator $G:\mathcal{V}\to\mathcal{V}$. The kernel of the time-reversal operator is

$$h(\boldsymbol{x};\boldsymbol{x}') = \int_{\Gamma} g(\boldsymbol{x};\boldsymbol{x}'')g^*(\boldsymbol{x}'';\boldsymbol{x}')d\boldsymbol{x}'', \tag{2}$$

where the asterisk denotes the operation of conjugation. The time-reversed signal corresponding to a transmitted signal v is given by Eq. (1) with g replaced by h. This can be written succinctly as $r(\boldsymbol{x})=H[v(\boldsymbol{x})]$, by defining the time-reversal operator $H:\mathcal{V}\rightarrow\mathcal{V}$.

Let $\{s^{(i)},\phi^{(i)}(\boldsymbol{x})\}$ be the ith singular value and (right) singular vector of the scattering operator, G, arranged and ordered so that $(\phi^{(i)},\phi^{(j)})=\delta_{ij}$ and $|s^{(1)}|\geq|s^{(2)}|\geq\cdots$. Then from Eq. (2) and the time-reversibility of the wave equation, we conclude that the eigenpairs of H are $\{\lambda^{(i)},\phi^{(i)}(\boldsymbol{x})\}$, where $\lambda^{(i)}=|s^{(i)}|^2$. We note that the singular vectors of G are also eigenvectors of H. Under certain conditions it can be shown that if an eigenvector of the time-reversal operator, H, is used as the transmitted signal, the wave field selectively focuses on the scatterer associated with the corresponding eigenvalue. Thus determining the eigenvectors of H, or, equivalently, the singular vectors of G, allows targeting specific scatterers in the propagation domain. Next we describe how this is accomplished by the traditional iterative time-reversal method via power iterations. Thereafter we describe the use of Lanczos iterations.

Power iterations: We define the following sequence as a single time-reversal iteration: given $v_{(0)}(\boldsymbol{x})$; transmit $v^*_{(0)}(\boldsymbol{x})$; receive $v_{(1/2)}(\boldsymbol{x})=G[v^*_{(0)}(\boldsymbol{x})]$; transmit $v^*_{(1/2)}(\boldsymbol{x})$: and receive $v_{(1)}(\boldsymbol{x})=G[v^*_{(1/2)}(\boldsymbol{x})]$. Using the definition of the scattering and the time-reversal operators, this sequence can be written as $v_{(1)}(\boldsymbol{x})=H[v_{(0)}(\boldsymbol{x})]$, where $H=GG^*$. In the time-reversal method based on power iterations the sequence described above is repeated n times to yield $v_{(n)}=H^n[v_{(0)}]$. Using the spectral decomposition of H it is easy to show that as n increases $v_{(n)}$ converges to the eigenvector corresponding to the largest eigenvalue of H. That is $v_{(n)}/\|v_{(n)}\|\approx\phi^{(1)}$, for n large.

The eigenvector corresponding to the second-largest eigenvalue can be determined once the first eigenvector has been determined by performing power iterations with $v_{(0)}-(v_{(0)},\phi^{(1)})\phi^{(1)}$, where $\phi^{(1)}$ is the estimate of the first eigenvector. That is, for every iterate, the component in the direction of $\phi^{(1)}$ is annihilated. This or a similar process can be repeated to locate weaker scatterers.[8]

Lanczos iterations: Lanczos iterations may be used to determine the eigenvectors of the time-reversal operator.[10]

(1) Select arbitrary $v_{(1)}$ such that $(v_{(1)},v_{(1)})=1$.
(2) Set $v_{(0)}=0$ and $\beta_1=0$.
(3) For $j=1,\ldots,n$
(a) $w_{(j)}=H[v_{(j)}]-\beta_j v_{(j-1)}$
(b) $\alpha_j=(w_{(j)},v_{(j)})$
(c) $w_{(j)}=w_{(j)}-\alpha_j v_{(j)}$
(d) $\beta_{j+1}=\sqrt{(w_{(j)},w_{(j)})}$
(e) $v_{(j+1)}=w_{(j)}/\beta_{j+1}$

After n iterations we will have created the $n\times n$ tridiagonal matrix, $\boldsymbol{T}=\text{tridiag}(\beta_{i+1},\alpha_i,\beta_{i+1})$, $i=1,\ldots,n$. Let $\{\gamma^{(i)},\psi^{(i)}\}$ be the ith eigenpair for $\boldsymbol{T}$. *Then the eigenvalues of* $\boldsymbol{T}$ *approximate the eigenvalues of* H, *that is* $\lambda^{(i)}\approx\gamma^{(i)}$ *and the eigenvectors of* H *can be determined from the eigenvectors of* $\boldsymbol{T}$ *and the Lanczos vectors* $v_{(i)}$, *that is* $\phi^{(i)}\approx\Sigma_{j=1}^{n}\psi_j^{(i)}v_{(j)}$.

There are several comments to be made here. First, in implementing this algorithm in an experiment the Lanczos vectors $v_{(i)}(\boldsymbol{x})$ would have to be stored. Though this is an added memory cost over power iterations, it is not a big penalty as they may be stored on the hard disk of a computer. Second, like the power iterations this algorithm also involves a time-reversal step (step 3a). Thus its implementation in an experiment is similar to that of power iterations, except

that every time-reversal step is followed by some additional signal processing. Finally, at the completion of n iterations, approximations to n eigenpairs are available. Out of these a small fraction ($\approx 20\%$) can be expected to be accurate. The accuracy may be increased by increasing the number of iterations

Theoretical advantages of Lanczos: The Lanczos method finds the best approximation of the n largest and smallest eigenvectors with an $n+1$ dimensional Krylov space spanned by the vectors $\{v_{(0)}, Hv_{(0)}, H^2v_{(0)}, \ldots, H^nv_{(0)}\}$. The power method's estimate for the largest eigenvector is $H^nv_{(0)}$, which is contained within the Krylov space. Clearly then, *for the same number of iterations, the Lanczos iterated time-reversal method always provides a more accurate estimate of the eigenvector than does standard iterated-time reversal*. Furthermore, from those same iterations, we have very accurate estimates of eigenvectors for smaller eigenvalues; no further iterations are required.

3. Numerical examples

A simplified problem: We consider a three-dimensional domain with N point scatterers located at $\boldsymbol{x}_j$, $j=1,\ldots,N$, clustered about the origin within a sphere of radius δ. The transmitters/receivers are continuously embedded in the $y-z$ plane at $x=D$ (denoted by Γ), with $D \gg \delta$. We consider the time-harmonic case with frequency ω and wavenumber $k=\omega/c$, where c is the sound speed in the medium. Further, we neglect multiple scattering and assume that scatterers are ideally separated.[11] That is, it is possible to focus on one scatterer without directing any energy to another scatterer. For this case $\{s^{(j)}, \phi^{(j)}(y,z)\}$ is the jth eigenpair of the scattering operator, where $s^{(j)}$ is proportional to the strength of the jth scatterer and $\phi^{(j)}(y,z) = g^{(j)}(D,y,z)/\|g^{(j)}(D,y,z)\|$. In this expression $\|\cdot\|$ denotes the L_2 norm on Γ and $g^{(j)}(\boldsymbol{x}) = e^{ik|\boldsymbol{x}-\boldsymbol{x}_j|}/4\pi|\boldsymbol{x}-\boldsymbol{x}_j|$. The time-reversal operator is given by $H=GG^*$, where $*$ denotes the complex conjugate. The eigenpairs of H are given by $\{|s^{(j)}|^2, \phi^{(j)}(y,z)\}$. Our goal is to use time-reversal methods to determine the $\phi^{(j)}$.

First we consider two scatterers ($N=2$). We start with an initial arbitrary signal $v^{(0)}(y,z)$ and let $\rho_j=(\phi^{(j)}, v^{(0)})$ be the components of this signal along $\phi^{(j)}$. Then after n steps of time-reversal iterations using the power method, the signal $v^{(n)}$, which approximates $\phi^{(1)}$, is given by

$$\frac{v^{(n)}(y,z)}{\rho_1|s^{(1)}|^{2n}} = \phi^{(1)}(y,z) + \frac{\rho_2}{\rho_1}\left|\frac{s^{(2)}}{s^{(1)}}\right|^{2n}\phi^{(2)}(y,z). \tag{3}$$

From Eq. (3) it is clear that the error is $(\rho_2/\rho_1)|s^{(2)}/s^{(1)}|^{2n}$. Assuming that the initial guess is arbitrary and thus not aligned with either $\phi^{(1)}$ or $\phi^{(2)}$ we expect $\rho_2/\rho_1=O(1)$ and conclude that the error is determined by the ratio of the strength of the scatterers. For scatterers of similar strength, say $s^{(2)}=0.98s^{(1)}$, after ten time-reversal iterations the error is $\approx 0.98^{20}=0.66$. For scatterers with disparate strengths, say $s^{(2)}=0.70s^{(1)}$, this error is much smaller ($\approx 0.08\%$). Because the Lanczos method recovers in m iterations the eigenpairs of an operator of rank m,[10] we recover $\phi^{(1)}$ and $\phi^{(2)}$ in exactly two iterations.

Next we consider scatterers with $s^{(j)}$ selected from a normal distribution of mean 0.5 and standard deviation 1. We use time-reversal methods to determine the eigenfunctions corresponding to the two strongest scatterers. For power iterations we use the first ten iterations to determine a guess for $\phi^{(1)}$, then use another ten iterations to determine a guess for $\phi^{(2)}$. This corresponds to a total of 20 time-reversal steps. In the Lanczos iterations based method also we use a total of 20 iterations. In order to obtain reliable indicators of performance we use 10 000 realizations. For each realization we determine the normalized error in estimating the first two eigenfunctions and plot a histogram of the log of this error (see Fig. 1).

Figure 1(a) shows histograms of the error in estimating the first eigenfunction by both the power method and the Lanczos method. We observe that the typical error in the Lanczos

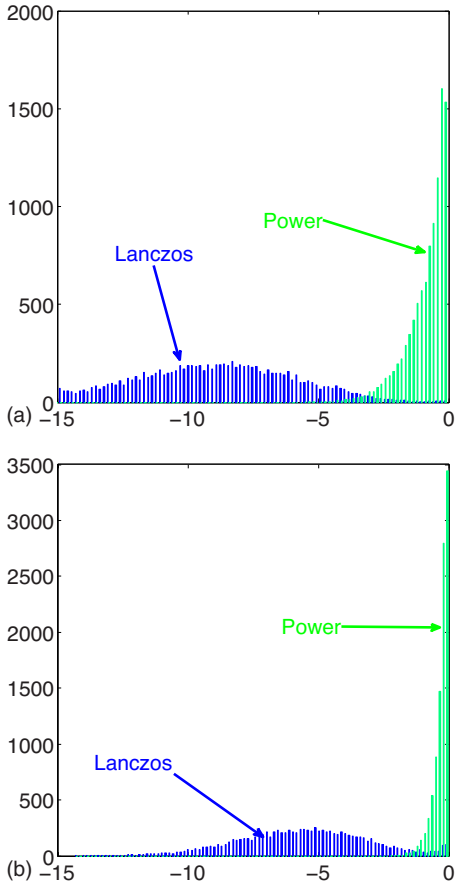


Fig. 1. (Color online) Number of realizations versus log of normalized error in the eigenfunction for power and Lanczos iterations: (a) first eigenfunction and (b) second eigenfunction.

estimate is roughly six orders of magnitude smaller than that for the power method. A similar trend is observed in the histograms shown in Fig. 1(b) for the second eigenfunction error, though there the power method estimate is frequently unusable.

Two-dimensional problems: Here we consider multiple scattering in two dimensions with 40 transmitters and receivers that produce waves with a wavelength of 0.25 units and are equally spaced in the interval (0, 6) along the x-axis. The medium contains 28 (weak) point scatterers located within a square of edge length 6 that is centered at $(x,y)=(4.5,4.5)$. The strength of these scatterers is randomly distributed in the interval]0, 0.2[. There are two other scatterers that are significantly stronger (as described below). Eight time-reversal iterations based on power and Lanczos iterations are used to locate the position of the two strongest scatterers. For this purpose, the computed value of the eigenvector of the time-reversal operator is back-propagated from the transmitters, once the iterations are complete. It is expected that the back-propagated field will focus on the scatterer we wish to target.

In the first case the strength of strongest scatterer is 2.1 and it is placed at $(x,y)=(4,3)$, while the strength of the second-strongest scatterer is 2 and it is placed at $(x,y)=(2,3)$. The back-propagated fields are plotted in Fig. 2. We observe that for both eigenvalues the Lanczos iterations focus tightly on the corresponding scatterer, while the power iterations based method is unable to distinguish between the two scatterers.

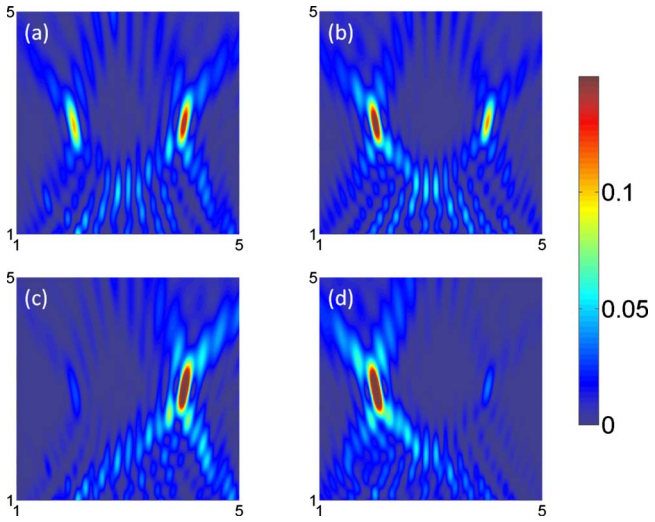


Fig. 2. (Color online) Back-propagated fields for estimates of two largest eigenvalues: (a) and (b) power iterations; (c) and (d) Lanczos.

We next place the strongest scatterer further from the array [strength=3; located at $(x,y)=(2,4.5)$], and the weaker scatterer closer to the array [strength=2; located at $(x,y)=(4,3)$], so that the weaker scatterer produces a stronger scattered field. In Fig. 3 we plot the back-propagated field for the two largest eigenvalues. We observe that the largest eigenvalue corresponds to the second-strongest scatterer and that in this case also Lanczos iterations are better at selectively focusing on scatterers.

4. Conclusions

We have presented a new iterated time-reversal method for efficiently identifying the eigenvectors of a scattering operator, which can be used (e.g., by DORT,[12] MUSIC[13]) to focus on or locate strong scatterers in a propagating medium. Our method makes use of Lanczos iterations

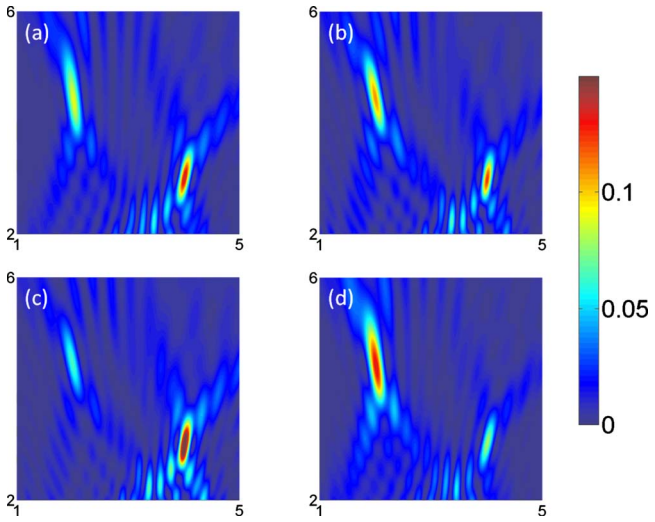


Fig. 3. (Color online) Back-propagated fields for estimates of two largest eigenvalues: (a) and (b) power iterations; (c) and (d) Lanczos.

in order to converge to the eigenpairs of the time-reversal operator. In numerical tests this method is shown to be remarkably superior over traditional iterated time-reversal. The new approach requires relatively little additional hardware and processing; it requires the storage of several (as many as time-reversal iterations) time signals for each transmitter on a hard drive, and some extra numerical processing. Its rapid convergence makes it well suited to applications.

References and links

[1]M. Fink, D. Cassereau, A. Derode, C. Prada, P. Roux, M. Tanter, J. Thomas, and F. Wu, "Time-reversed acoustics," Rep. Prog. Phys. **63**, 1933–1995 (2000).
[2]B. E. Anderson, M. Griffa, C. Larmat, T. J. Ulrich, and P. A. Johnson, "Time Reversal," Acoust. Today **4**, 5–15 (2008).
[3]M. Fink, G. Montaldo, and M. Tanter, "Time-reversal acoustics in biomedical engineering," Annu. Rev. Biomed. Eng. **5**, 465–497 (2003).
[4]M. Griffa, B. E. Anderson, R. A. Guyer, T. J. Ulrich, and P. A. Johnson, "Investigation of the robustness of time reversal acoustics in solid media through the reconstruction of temporally symmetric sources," J. Phys. D **41**, 85415–85428 (2008).
[5]C. Gaumond, D. Fromm, J. Lingevitch, R. Menis, G. Edelmann, D. Calvo, and E. Kim, "Demonstration at sea of the decomposition-of-the-time-reversal-operator technique," J. Acoust. Soc. Am. **119**, 976–990 (2006).
[6]S. Kim, W. Kuperman, W. Hodgkiss, H. Song, G. Edelmann, and T. Akal, "Robust time reversal focusing in the ocean," J. Acoust. Soc. Am. **114**, 145–157 (2003).
[7]T. Ulrich, P. Johnson, and R. Guyer, "Interaction Dynamics of Elastic Waves with a Complex Nonlinear Scatterer through the Use of a Time Reversal Mirror," Phys. Rev. Lett. **98**, 104301 (2007).
[8]G. Montaldo, M. Tanter, and M. Fink, "Revisiting iterative time reversal processing: Application to detection of multiple targets," J. Acoust. Soc. Am. **115**, 776–784 (2004).
[9]A. Song, M. Badiey, H. C. Song, W. S. Hodgkiss, M. B. Porter, and KauaiEx Grp. "Impact of ocean variability on coherent underwater acoustic communications during the Kaual experiment (KauaiEx)," J. Acoust. Soc. Am. **123**, 856–865 (2008).
[10]Y. Saad, *Iterative Methods for Sparse Linear Systems*, 1st ed. (PWS, Boston, MA, 1995).
[11]C. Prada, J. Thomas, and M. Fink, "The iterative time reversal process: Analysis of the convergence," J. Acoust. Soc. Am. **97**, 62–71 (1995).
[12]N. Mordant, C. Prada, and M. Fink, "Highly resolved detection and selective focusing in a waveguide using the DORT method," J. Acoust. Soc. Am. **105**, 2634–2642 (1999).
[13]A. Devaney, E. Marengo, and F. Gruber, "Time-reversal-based imaging and inverse scattering of multiply scattering point targets," J. Acoust. Soc. Am. **118**, 3129–3138 (2005).

 [DOI: 10.1121/1.3062145] Published Online 21 January 2009

Effect of bandwidth extension to telephone speech recognition in cochlear implant users

Chuping Liu
Department of Electrical Engineering, University of Southern California, Los Angeles, California 90089
chupingl@usc.edu

Qian-Jie Fu
Department of Biomedical Engineering, University of Southern California, Los Angeles, California 90089
and Department of Auditory Implants and Perception, House Ear Institute,
2100 West Third Street, Los Angeles, California 90057
qfu@hei.org

Shrikanth S. Narayanan
Department of Electrical Engineering, University of Southern California, Los Angeles, California 90089
shri@sipi.usc.edu

Abstract: The present study investigated a bandwidth extension method to enhance telephone speech understanding for cochlear implant (CI) users. The acoustic information above telephone speech transmission range (i.e., 3400 Hz) was estimated based on trained models describing the relation between narrow-band and wide-band speech. The effect of the bandwidth extension method was evaluated with IEEE sentence recognition tests in seven CI users. Results showed a relatively modest but significant improvement in the speech recognition with the proposed method. The effect of bandwidth extension method was also observed to be highly dependent on individual CI users.

PACS numbers: 43.71.Ky, 43.64.Me, 43.60.Dh [DS]
Date Received: July 31, 2008 **Date Accepted:** December 8, 2008

1. Introduction

Telephone use is still challenging for many deaf or hearing-impaired individuals including cochlear implant (CI) users. According to a previous study (Kepler *et al.*, 1992), there are three major contributors to the difficulties in telephone communication: the limited frequency range, the elimination of visual cues, and the reduced audibility of telephone signal. For example, the telephone bandwidth in use today is limited to 300–3400 Hz. Compared to speech in face-to-face conversational settings, telephone speech does not convey information above 3400 Hz, which is useful in the identification of many speech sounds, notably certain consonants such as fricatives. Since CI users generally receive frequency information up to approximate 8 kHz or even higher, the narrow-band telephone speech may present an obstacle even when they can achieve a fairly good wide-band speech perception.

Previous studies have assessed the capability of CI patients to communicate over telephones. While many CI patients were capable of certain degree of communication over the phones, speech understanding was significantly worse than with broad-band speech (Milchard and Cullington, 2004; Ito *et al.*, 1999; Fu and Galvin, 2006). For example, word discrimination score obtained from telephone speech was decreased by 17.7% than those with wide-band speech. Analysis of the word errors revealed that the place of articulation was the predominant type of error (Milchard and Cullington, 2004). On the other hand, investigation of telephone use among CI recipients reported that 70% of the respondents communicated via the telephone, of which 30% used cellular phones (Cray *et al.*, 2004). Hence, improved capability to understand telephone speech using just auditory cues will increase the opportunities for the use of the

telephone and will promote independent living, employment, socialization, and self-esteem in CI users.

To improve the telephone communication ability of hearing-impaired people, one solution, albeit expensive, is to change the current public switched telephone network to transmit wide-band speech and to enrich the spoken information with videos. This is, however, difficult to accomplish in the near future. A more economical and near term approach is to add external equipment to enhance the audibility of telephone speech. For example, the telephone adapter, which was used to reduce noise level in the telephone and to record telephone speech into a tape recorder, was found to boost speech-tracking scores in CI users (Ito *et al.*, 1999). Yet, such auxiliary instruments may not be easy to obtain, especially in mobile communication. Another potential approach is to improve speech processing and transmission technique. A previous study (Terry *et al.*, 1992) investigated frequency-selective amplification and compression via digital signal processing techniques to compensate for high-frequency hearing loss in hearing-impaired people. Nevertheless, the approach required audiometric data from individual users to achieve the best performance.

On the other hand, to overcome the deficit of telephone speech in terms of narrow bandwidth, bandwidth extension as a front end processing was studied (e.g., Nilsson and Kleijn, 2001; Jax and Vary, 2003). For example, Jax and Vary (2003) proposed an approach to extend telephone bandwidth to 7 kHz based on hidden Markov model. Nilsson and Kleijn (2001) studied a bandwidth extension approach to avoid overestimation of high-band energy. Through listening tests, the method was shown to reduce the degree of artifacts. Yet, it is not clear how much gain the bandwidth-extension method can actually bring to speech recognition with listeners, especially for CI users.

In this study, we propose a bandwidth-extension method to enhance telephone speech. Gaussian mixture model (GMM) was used to model the spectrum distribution of narrow-band speech. The relationship between wide-band and narrow-band speech was learned a priori in a data driven fashion and was used to recover the missing information based on the available telephone band speech. Such an approach does not require auxiliary instruments and patient data for its implementation. We then studied the effect of the proposed bandwidth-extension method on speech recognition performance in CI users.

2. Methods

The step to expanding narrow-band speech to wide-band speech basically consists of two parts: spectral envelope extension and excitation spectrum extension, which are introduced in Secs. 2.1 and 2.2, respectively.

2.1 GMM-based spectral envelope extension

A GMM represents the distribution of the observed parameters by m mixture Gaussian components in the form of

$$p(\mathbf{x}) = \sum_{i=1}^{m} \alpha_i N(\mathbf{x}, \mu_i, \Sigma_i), \tag{1}$$

where α_i denotes the prior probability of component i ($\Sigma_{i=1}^{m} \alpha_i = 1$ and $\alpha_i \geqslant 0$) and $N(\mathbf{x}, \mu_i, \Sigma_i)$ denotes the normal distribution of the ith component with mean vector μ_i and covariance matrix Σ_i in the form of

$$N(\mathbf{x}, \mu_i, \Sigma_i) = \frac{1}{(2\pi)^{p/2} |\Sigma_i|^{1/2}} \exp\left[-\frac{1}{2} (\mathbf{x} - \mu_i)^T \Sigma_i^{-1} (\mathbf{x} - \mu_i) \right], \tag{2}$$

where p is the vector dimension. The parameters of the model (α, μ, Σ) can be estimated using the well-known expectation maximization algorithm.

Let $\mathbf{x}=[\mathbf{x}_1\mathbf{x}_2\cdots\mathbf{x}_n]$ be the sequence of n spectral vectors produced by the narrow-band telephone speech, and let $\mathbf{y}=[\mathbf{y}_1\mathbf{y}_2\cdots\mathbf{y}_n]$ be the time-aligned spectral vectors produced by the wide-band speech. The objective of the bandwidth-extension method was to define a conversion function $F(x_t)$ such that the total conversion error of spectral vectors

$$\varepsilon = \sum_{t=1}^{n} (\mathbf{y}_t - F(\mathbf{x}_t))^2 \tag{3}$$

was minimized over the entire training spectral feature set, using the trained GMM that represents the feature distribution of the telephone speech. A minimum mean square error method was used to estimate the conversion function. The conversion function was (Stylianou *et al.*, 1998; Kain and Macon, 1998)

$$F(\mathbf{x}_t) = \sum_{i=1}^{m} P(C_i|\mathbf{x}_t)[\mathbf{v}_i + \mathbf{T}_i\Sigma_i^{-1}(\mathbf{x}_t - \mu_i)], \tag{4}$$

where $P(C_i|\mathbf{x}_t)$ is the posterior probability that the ith Gaussian component generates $\mathbf{x}_t$; $\mathbf{v}_i$ and $\mathbf{T}_i$ are the mean wide-band spectral vector and the cross-covariance matrix of the wide-band and narrow-band spectral vectors, respectively. When a diagonal conversion is used (i.e., $\mathbf{T}_i$ and Σ_i are diagonal), the above optimization problem simplifies into a scalar optimization problem and the computation cost is greatly decreased.

2.2 Excitation spectrum extension

Two methods are considered for excitation spectrum extension in this study (Makhoul and Berouti, 1979): spectral folding and spectral translation. Spectral folding simply generates a mirror image of the narrow-band spectrum for high-band spectrum. The implementation of spectral mirroring was equivalent to upsampling the excitation signal in the time domain by zero padding. This almost added no extra cost in the processing. Yet, the energy in the reconstructed high band is typically overestimated with this approach; the harmonic pattern of the restored high band is a flipped version of the original narrow-band spectrum, centered around the highest frequency of the narrow-band speech. Spectral translation, on the other hand, did not have these problems, but involves more expensive computation. The excitation spectrum of the narrow-band speech, obtained from Fourier transformation of the time domain signal, is translated to the high-frequency part and padded to fill the desired whole band. A low pass filter is applied to do spectral whitening, such that the discontinuities between the translations are smoothed. The extended wide-band excitation in the time domain is then obtained from inverse Fourier transformation.

2.3 Speech analysis and synthesis

In this study, Mel-scaled line spectral frequency (LSF) features (18th order) and energy were extracted to model the spectral characteristics of speech in a 19 dimensional space. The spectral features between narrow-band and wide-band speech were aligned with dynamic time warping computation. The spectral mapping function between narrow-band and wide-band speech was trained with 200 randomly selected sentences from the IEEE database (100 sentences from a female talker and the other 100 sentences from a male talker). The excitation component between 1 and 3 kHz was used to construct the high-band excitation component because the spectrum in this range was relatively white. A low pass Butterworth filter (first order with cutoff frequency 3000 Hz) was used to do spectral whitening. The synthesized high-band speech (i.e., frequency information above 3400 Hz) was obtained from high pass filtering the convolution result of the extended excitation and extended spectrum. It was then appended to the original telephone speech to render the reconstructed wide-band speech that covered the frequency band from 300 to 8000 Hz.

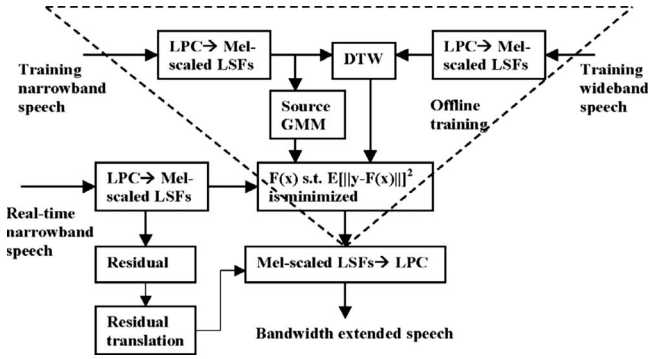


Fig. 1. Implementation framework of the GMM-based bandwidth-extension method.

2.4 Implementation framework of the bandwidth-extension method

Figure 1 illustrates the GMM-based bandwidth-extension method. The three major components of the model (i.e., GMM-based spectral envelope extension, excitation spectrum extension, and speech analysis/synthesis) are as detailed in Secs. 2.1 and 2.3.

2.5 Test materials and procedures

The test materials in this study were IEEE (1969) sentences, recorded from one male talker and one female talker at the House Ear Institute with a sampling rate of 22 050 Hz. The narrowband telephone speech was obtained by bandpass filtering the above wide-band speech (ninth order Butterworth filter, bandpass between 300 and 3400 Hz) and was downsampled to 8 kHz. Three conditions were tested: restored wide-band speech (carrying information up to 8 kHz), telephone speech (carrying information up to 3.4 kHz), and originally recorded wide-band speech (carrying information up to 11 kHz). All sentences were normalized to have the same long-term root mean square value. Note that the GMM training sentences (i.e., 200 randomly selected sentences) were also bandwidth extended and included in the listening test to increase the available speech materials for the experiment.

Seven CI subjects (two women and five men) participated in this study. Table 1 lists relevant demographics for the CI subjects. All subjects were native speakers of American English and had extensive experience in speech, recognition experiments. For all the listening conditions including restored wide-band speech, telephone speech, and originally recorded wide-band speech, subjects were tested using their clinically assigned speech processor and

Table 1. Subject demographics for the CI patients who participated in the present study.

Subject	Age	Gender	Etiology	Implant type	Strategy	Duration of implant use (years)
S1	55	M	Hereditary	Freedom	ACE	1
S2	62	F	Genetic	Nucleus-24	ACE	2
S3	48	M	Trauma	Nucleus-22	SPEAK	13
S4	67	M	Hereditary	Nucleus-22	SPEAK	14
S5	64	M	Trauma/unknown	Nucleus-22	SPEAK	15
S6	75	M	Noise induced	Nucleus-22	SPEAK	9
S7	72	F	Unknown	Nucleus-24	ACE	5

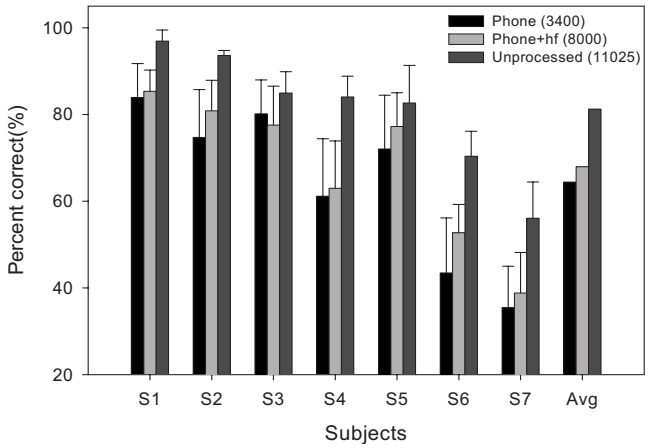


Fig. 2. Sentence recognition performance for individual CI subjects with and without the bandwidth-extension method, and with the unprocessed wide-band speech. The error bars indicate one standard deviation.

comfortable volume/sensitivity settings. As shown in Table 1, the subjects used ACE (Skinner *et al.*, 2002) or SPEAK strategy (Seligman and McDermott, 1995). The maximum number of activated electrodes is typically 6 for SPEAK strategy and 8 for ACE strategy, respectively. While the number of activated electrodes is the same for both telephone speech and broad-band speech, the number of total usable electrodes is different. In general, all 20 electrodes will be used when listening to broad-band speech while only 13 electrodes will be used for telephone speech. Once testing began, these settings were not changed. Subjects were tested while seated in a double-walled sound-treated booth (IAC). Stimuli were presented via a single loud speaker at 65 dBA. The test order of different conditions was randomized for each subject. No feedback was provided during the test.

3. Results and discussion

The sentence recognition performance with and without the restored high-band components is shown in Fig. 2, together with the performance with the naturally recorded wide-band speech. Note that the subjects are ordered according to their performance with wide-band speech. On average, compared to the performance with the naturally recorded wide-band speech, the performance with the narrow-band telephone speech was about 16.8% lower, which was significant (paired t-test: $p \leqslant 0.001$). The recognition score with the bandwidth-extension method was about 3.5% higher than without the bandwidth-extension method. The improvement was small but significant (paired t-test, $p = 0.050$). Yet, the performance with the bandwidth-extension method was still significantly lower than with the unprocessed wide-band speech (paired t-test, $p \leqslant 0.001$).

Figure 2 demonstrates substantial cross subject variability in performance. First, the cross subject variability was observed in terms of the performance for the same test materials. For example, subject S1 obtained over 80% correct under with and without the restored high-band component conditions. In contrast, subject S7 obtained only about 40% in average. Second, the cross subject variability was observed in terms of the effect of the bandwidth-extension method. For example, subject S6 achieved about 10% improvement with the restored high-band information; while subject S3 had even about 3% deficit in performance with the restored high-band information.

4. Discussion

The present study showed a 16.8% performance drop in CI users' listening to narrow-band telephone speech than listening to the originally recorded wide-band speech. This percentage drop was similar to the performance drop reported in Milchard and Cullington, 2004, although

the testing materials and testing procedures were different between these two studies. In the current study, seven CI subjects were tested with IEEE (1969) sentences. In Milchard and Cullington's (2004) study, ten CI subjects were tested with 80 consonant-vowel-consonant type stimuli (e.g., BAD BAG BAT BACK) using the four alternative auditory feature test procedure. The present study confirmed the findings in previous studies that the bandwidth effect was substantial in CI listeners.

The observed cross subject performance difference may be due to different CI device settings and different electropsychoacoustic listening patterns across subjects. For example, for those CI users whose speech processor encoded more information on the high-band speech, the potential benefit of the bandwidth-extension method may be relatively larger than the other CI users.

In the present study, a bandwidth-extension method was proposed to improve the telephone speech recognition performance in CI listeners. Although speech recognition was improved with the proposed bandwidth-extension method in a significant manner, the improvement was relatively small compared to the observed 16.8% performance drop from wide-band speech to telephone speech. There are four possible reasons for this marginal improvement. First, the proposed bandwidth-extension method only recovered information up to 8 kHz, while the 16.8% performance drop was the performance difference between wide-band speech (11 kHz) and narrow-band telephone speech (3.4 kHz). It was not clear how much the recognition benefit might be for the acoustic information between 8 and 11 kHz. Second, in this study, Mel-scaled LSF features were used, which placed lower resolution on the high-frequency components. The feature order used for speech analysis was the same (18th order) for both wide-band and narrow-band speech, although their frequency ranges were different. Such signal processing procedures may not result in high accuracy in parameter estimation. Third, due to the nature of speech synthesis, it was difficult to accomplish a synthesis without perceptual distortion. The introduced artifacts may be very detrimental for CI listeners, who typically receive degraded spectrotemporal information. Finally, performance with the bandwidth-extended speech was acutely measured in CI listeners in free field; the potential benefit with the bandwidth extended method might be underestimated since the training effect was not taken into account.

5. Conclusions

This paper studied a bandwidth-extension method to enhance telephone speech understanding in CI users. The lost high-band acoustic information was estimated based on the available narrow-band telephone speech and a pretrained relation between narrow-band and wide-band speech. The narrow-band excitation was extended to wide-band excitation by spectral translation. A source filter model was used to synthesize estimated wide-band speech, whose high-band frequency information was filtered out and appended to the original telephone speech. The effect of bandwidth-extension method was evaluated with IEEE (1969) sentence recognition tests in seven CI users. Results showed that CI speech recognition was significantly improved with the bandwidth-extension method, although it was relatively small compared to the performance drop seen from the wide-band speech to telephone speech. The benefit of the bandwidth-extension method was also highly dependent on individual CI users.

Acknowledgments

We acknowledge all the subjects that participated in this study. Research was supported in part by NIH-NIDCD.

References and links

Cray, J. W., Allen, R. L., Stuart, A., Hudson, S., Layman, E., and Givens, G. D. (**2004**). "An investigation of telephone use among cochlear implant recipients," Am. J. of Audiology **13**, 200–212.

Fu, Q. J., and Galvin, J. J. (**2006**). "Recognition of simulated telephone speech by cochlear implant users," Am J. Audiol. **15**, 127–32.

IEEE (**1969**). *IEEE Recommended Practice for Speech Quality Measurements* (Institute of Electrical and Electronic Engineers, New York).

Ito, J., Nakatake, M., and Fujita, S. (**1999**). "Hearing ability by telephone of patients with cochlear implants," Otolaryngol.-Head Neck Surg. **121**, 802–804.
Jax, P., and Vary, P. (**2003**). "On artificial bandwidth extension of telephone speech," Signal Process. **83**, 1707–1719.
Kain, A., and Macon, M. W. (**1998**). "Spectral voice conversion for text-to-speech synthesis," IEEE ICASSP, pp. 285–288.
Kepler, L. J., Terry, M., and Sweetman, R. H. (**1992**). "Telephone usage in the hearing-impaired population," Ear Hear., **13**, 311–319.
Makhoul, J., and Berouti, M. (**1979**). "*High-frequency regeneration in speech coding systems*," IEEE ICASSP, pp. 428–431.
Milchard, A. J., and Cullington, H. E. (**2004**). "An investigation into the effect of limiting the frequency bandwidth of speech on speech recognition in adult cochlear implant users." Int. J. Audiol., **43**, 356–362.
Nilsson, M., and Kleijn, W. B. (**2001**). "Avoiding over-estimation in bandwidth extension of telephony speech," IEEE ICASSP, pp. 869–872.
Seligman, P. M., and McDermott, H. J. (**1995**). "Architecture of the spectra-22 speech processor," Ann. Otol. Rhinol. Laryngol. Suppl. **166**, 139–141.
Skinner, M. W., Arndt, P. L., and Staller, S. J. (**2002**). "Nucleus 24 advanced encoder conversion study: Performance versus preference," Ear Hear. **23**, 2S–17S.
Stylianou, Y., Cappe, O., and Moulines, E. (**1998**). "Continuous probabilistic transform for voice conversion," IEEE Trans. Commun. **6**, 131–142.
Terry, M., Bright, K., Durian, M., Kepler, L., Sweetman, R., and Grim, M. (**1992**). "Processing the telephone speech signal for the hearing impaired," Ear Hear. **13**, 70–79.

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.

Extensional edge modes in elastic plates and shells (L)

J. Kaplunov[a] and A. V. Pichugin[b]
Department of Mathematical Sciences, Brunel University, Uxbridge UB8 3PH, United Kingdom

V. Zernov[c]
Waves and Fields Research Group, Department of Electrical, Computer and Communication Engineering, Faculty of Engineering, Science and Built Environment, South Bank University, London SE1 0AA, United Kingdom

(Received 26 November 2008; accepted 3 December 2008)

The recently discovered undamped localized mode at the end of an elastic strip is demonstrated to be particularly relevant in the plane stress setting, where it exists for the Poisson ratio 0.29. This paper also emphasizes the difference between low-frequency edge modes, typically characterized by low variation across the plate (or shell) thickness, and high-frequency edge modes, whose natural frequencies are of the order of thickness resonance frequencies.
 [DOI: 10.1121/1.3056558]

PACS number(s): 43.40.Dx, 43.40.Ey, 43.20.Ks [AJMD] Pages: 621–623

I. EDGE MODES IN PLATES

It is well understood that the free end of a semi-infinite elastic waveguide, such as a cylindrical rod, a plate in plane strain, or a cylindrical shell, can support localized modes. Generally such modes correspond to complex-valued natural frequencies, i.e., they radiate a certain part of their vibration energy to infinity. Nevertheless, it was only recently noticed that for certain special values of the Poisson ratio natural frequencies of these localized modes can become real, thus justifying more traditional term "edge resonance" (see Refs. 1–3 and references therein). Specifically, for the extensional localized edge mode in an isotropic elastic layer with traction-free faces in plane strain this happens when the Poisson ratio $\nu=0$ or $\nu\approx 0.2248$. The real part of the natural frequency associated with this mode may be approximated by the simple expression

$$\Re(\omega_h)=\frac{151+68\nu+50\nu^2}{76h}c_2, \tag{1}$$

in which h is the layer half-thickness and $c_2=\sqrt{\mu/\rho}$ is the shear wave speed (where μ is the shear modulus and ρ is the volume density). The imaginary part of the normalized natural frequency $\mathcal{I}(\omega_h h/c_2)$ is shown as dotted line in Fig. 1.

While interesting solutions for theoreticians, edge modes are often perceived as a curiosity from the practical point of view. First, the Poisson ratios of most engineering materials are relatively far from the two aforementioned values that enable undamped edge resonance. Second, due to its high frequency, the edge resonance itself remains an exotic phenomenon from the point of view of an engineer. For example, Eq. (1) predicts that in a typical 10 mm steel plate with $\nu=0.29$ and $c_2=3200$ m/s one may expect to observe a localized edge mode at approximately 234.4 kHz (see also the left hand diagram in Fig. 2).

It was Filon who first noted that a thin plate in the state of *generalized plane stress* may be approximately described by the system of two-dimensional equations that is formally equivalent to the conventional plane strain theory.[4] The relevant governing equations only differ from their plane strain counterpart by a new definition of the Lamé parameter λ_* given by

$$\lambda_*=\frac{2\lambda\mu}{\lambda+2\mu}, \tag{2}$$

where λ and μ are the standard Lamé parameters in the corresponding exact three-dimensional problem of elasticity. This implies that the Poisson ratio of the plane stress problem ν_* is also distinct from the three-dimensional definition; straightforward derivation shows that $\nu_*=\nu/(\nu+1)$.

The plane stress approximation applies to thin plates with free faces that are loaded along their perimeter by systems of in-plane forces. In particular, it is valid for a thin semi-infinite strip (see the left hand diagram in Fig. 2). We shall assume in the sequel that all edges of the strip are stress-free and refer to it as the *thin strip*. An undamped edge

[a]Electronic mail: julius.kaplunov@brunel.ac.uk
[b]Electronic mail: aleksey.pichugin@brunel.ac.uk
[c]Electronic mail: zernovv@lsbu.ac.uk

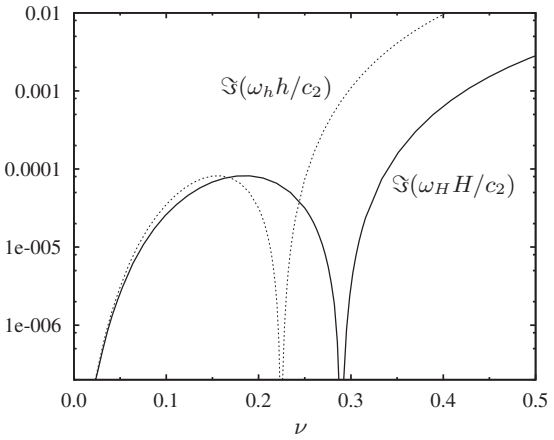


FIG. 1. Imaginary part of natural frequency of the edge localized mode in plane strain (dotted) and plane stress (solid).

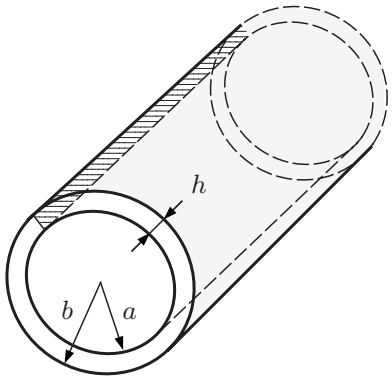


FIG. 3. A cylindrical shell.

mode is known to exist in a semi-infinite strip when $\nu_* \approx 0.2248$. The corresponding three-dimensional Poisson ratio may be found as

$$\nu = \frac{\nu_*}{1-\nu_*} \approx 0.2900, \tag{3}$$

which means that an edge resonance is going to be particularly well pronounced in thin strips made of a wide range of iron-based alloys, including steel.

More generally, approximate formula (1) may be rewritten to represent the real part of the natural frequency of symmetric (extensional) localized edge mode in a thin strip:

$$\Re(\omega_H) = \frac{151 - 234\nu + 133\nu^2}{76(1-\nu)^2 H} c_2. \tag{4}$$

The imaginary part of the normalized natural frequency, which characterizes the damping of the edge mode, is presented in Fig. 1 (solid line) along with the corresponding plane strain results. In order to ensure the validity of the plane stress assumption, the half-width of the strip must satisfy $H \gg h$. In other words, thickness-to-width ratio $\epsilon \equiv h/H$ must be kept small.

Direct comparison of estimates (1) and (4) indicates that $\omega_H \sim \epsilon\omega_h$. Therefore, it is clear that Eq. (4) results in a significantly lower prediction for the natural frequency of a thin strip. For example, a 20 cm wide strip made of steel with $\nu=0.29$ and $c_2=3200$ m/s will exhibit an edge mode at a frequency around 12.54 kHz. It is remarkable that the imaginary part of natural frequency $\mathcal{I}(\omega_H)$ is also reduced by a factor of $O(\epsilon)$ in the related plane stress problem (see Fig. 1). This indicates that the corresponding edge mode is significantly less damped for all values of the Poisson ratio.

The described relationship between edge modes in plane strain and plane stress suggests a natural classification of localized modes in thin plates into high- and low-frequency ones, which falls in line with general nomenclature developed in the dynamics of thin-walled structures.[5] The high-frequency modes are characterized by a wavelength of the same order of magnitude as the plate thickness, so that their natural frequencies are of the same order of magnitude as frequencies of thickness resonances. Contrary to this, the low-frequency modes are characterized by a wavelength that is $O(h/\epsilon)$, so that the associated natural frequencies are $O(\epsilon)$. The presence of a natural asymptotic parameter implies that the low-frequency edge modes will typically be amenable to various types of asymptotic treatment. This is not in general true for high-frequency edge modes.

II. EDGE MODES IN THIN SHELLS

Let us now consider a cylindrical shell with an inner radius a and an outer radius b and assume it is thin, i.e., $h \ll R$, where $h=b-a$ and $R=(a+b)/2$ (see Fig. 3). The edge modes in such shell can also be classified into low- and high frequency with respect to the thickness resonance frequencies. The low-frequency edge modes can be analyzed on the basis of the classical two-dimensional Kirchhoff-Love theory of shells.[6] In the particular case of high-order extensional edge modes, whose circumferential variation makes them less sensitive to the curvature, a shell can be modeled by a

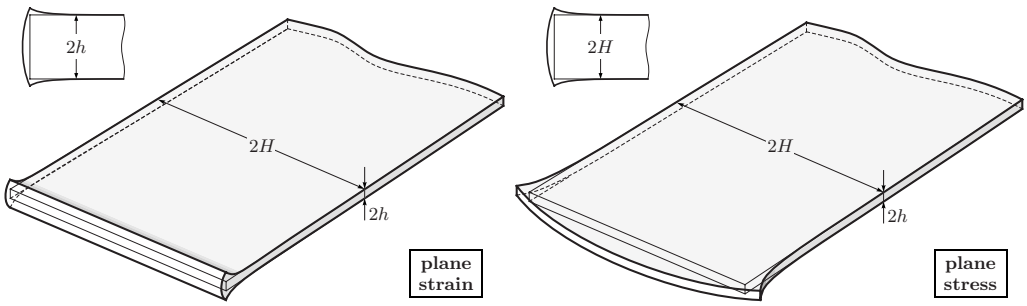


FIG. 2. Diagram to illustrate the difference between plane strain and plane stress edge modes in a thin elastic strip.

flat plate in a generalized plane stress state, as in the right hand diagram of Fig. 2. The associated boundary conditions require vanishing of surface traction along the free edge of length $H=2\pi R$ and specify mixed boundary data along semi-infinite sides. The use of mixed boundary conditions enables separation of variables[6] and leads to the following estimate for the real part of edge modes' natural frequencies:

$$\Re(\omega_H) \sim \frac{\pi n c_R^*}{H}, \tag{5}$$

where $1 \ll n \ll R/h$. Constant c_R^* in Eq. (5) is the Rayleigh wave speed in the theory of generalized plane stress, i.e., the Rayleigh wave speed calculated for a half-space with Lamé parameter λ_* defined by Eq. (2). The corresponding estimates for the imaginary parts of natural frequencies were previously obtained[6] and are not quoted here for the sake of brevity.

It may also be expected from the general asymptotic theory of shells[5] that the axisymmetric high-frequency extensional mode of a thin cylindrical shell must be very similar to the already discussed edge mode in a layer subjected to plane strain. The presence of curvature does not significantly distort the plane strain field for transverse cross-sections of the shell so, as a result, the associated correction to the natural frequency of the edge mode is only $O(h^2/R^2)$.[5] This was recently confirmed both numerically and experimentally by Ratassepp *et al.*[7] For example, they observe edge resonance in an aluminium pipe with $c_2=3113$ m/s, $\nu=0.348$, $R=8.8$ mm, and $h=1.1$ mm at frequency 1.074 MHz, whereas approximate formula (1) predicts the edge mode natural frequency to be 1.071 MHz. It is also instructive to compare this with the frequency of e.g., the tenth extensional "plane stress" edge mode in the same pipe. Formula (5) predicts that it occurs at 256.7 kHz.

ACKNOWLEDGMENTS

This paper benefited from several fruitful discussions with Dr. E. V. Nolde (Brunel University), which are very gratefully acknowledged.

[1] I. Roitberg, D. Vassiliev, and T. Weidl, "Edge resonance in an elastic semi-strip," Q. J. Mech. Appl. Math. **51**(1), 1–13 (1998).
[2] V. Pagneux, "Revisiting the edge resonance for Lamb waves in a semi-infinite plate," J. Acoust. Soc. Am. **120**(2), 649–656 (2006).
[3] V. Zernov, A. V. Pichugin, and J. Kaplunov, "Eigenvalue of a semi-infinite elastic strip," Proc. R. Soc. London, Ser. A **462**(2068), 1255–1270 (2006).
[4] A. E. H. Love, *A Treatise on the Mathematical Theory of Elasticity*, 4th ed. (Cambridge U.P., Cambridge, 1927).
[5] J. D. Kaplunov, L. Yu. Kossovich, and E. V. Nolde, *Dynamics of Thin Walled Elastic Bodies* (Academic, New York, 1998).
[6] J. D. Kaplunov, L. Yu. Kossovich, and M. V. Wilde, "Free localized vibrations of a semi-infinite cylindrical shell," J. Acoust. Soc. Am. **107**(3), 1383–1393 (2000).
[7] M. Ratassepp, A. Klauson, F. Chati, F. Léon, and G. Maze, "Edge resonance in semi-infinite thick pipe: Numerical predictions and measurements," J. Acoust. Soc. Am. **124**(2), 875–885 (2008).

An acoustic survey of beaked whales at Cross Seamount near Hawaii (L)

Mark A. McDonald[a)]
WhaleAcoustics, 11430 Rist Canyon Road, Bellvue, Colorado 80512

John A. Hildebrand[b)] and Sean M. Wiggins[c)]
Scripps Institution of Oceanography, University of California San Diego, La Jolla, California 92093-0205

David W. Johnston[d)]
Division of Marine Science and Conservation, Nicholas School of the Environment, Duke University Marine Laboratory, 135 Duke Marine Laboratory Road Beaufort, North Carolina 28516

Jeffrey J. Polovina[e)]
Pacific Islands Fisheries Science Center, National Marine Fisheries Service, 2570 Dole Street, Honolulu, Hawaii 96822

(Received 2 July 2008; revised 24 November 2008; accepted 24 November 2008)

An acoustic record from Cross Seamount, southwest of Hawaii, revealed sounds characteristic of beaked whale echolocation at the same relative abundance year-around (270 of 356 days), occurring almost entirely at night. The most common sound had a linear frequency upsweep from 35 to 100 kHz (the bandwidth of recording), an interpulse interval of 0.11 s, and duration of at least 932 μs. A less common upsweep sound with shorter interpulse interval and slower sweep rate was also present. Sounds matching Cuvier's beaked whale were not detected, and Blainville's beaked whale sounds were detected on only one occasion.
 [DOI: 10.1121/1.3050317]

PACS number(s): 43.30.Sf, 43.80.Ka [WWA] Pages: 624–627

I. INTRODUCTION

Two species of beaked whales, Cuvier's (*Ziphius cavirostris*) and Blainville's (*Mesoplodon densirostris*), are known to use frequency upswept echolocation sounds, in contrast to the short duration clicks of most echolocating cetaceans. Cuvier's beaked whales echolocation sounds are 200 μs duration linear upsweeps with a center frequency near 42 kHz, interpulse interval (IPI) of 0.38 s, source level up to 214 dBp.p. 1 μPa at 1 m, and bandwidth of 23 kHz (Zimmer *et al.*, 2005). The characteristic sounds of Blainville's beaked whales are only subtly different from Cuvier's beaked whales, with a sharper cutoff below 25 kHz (Johnson *et al.*, 2006). We describe a one-year-long acoustic study of the most common type of whale recorded at Cross Seamount; a preliminary analysis was described by Johnston *et al.* (2008). It is thought that these sounds were produced by a species of beaked whale other than Cuvier's or Blainville's.

II. METHODS

A high-frequency Acoustic Recording Package or HARP (Wiggins and Hildebrand, 2007) was placed on top of Cross Seamount (18° 43.325′ N, 158° 15.230 W) at 395 m depth, 290 km south of Oahu. The HARP frequency response is 2 dB more sensitive at 40 kHz than at 25 kHz and 12 dB more sensitive at 80 kHz than at 40 kHz, rolling off above 80 kHz. The electronic noise floor of the HARP is equivalent to the ambient ocean noise in sea state 5 at frequencies above 3 kHz. The HARP sampled at 200 ksamples/s for five of every 25 min from 26 April to 28 October 2005 and from 11 November 2005 to 11 May 2006.

Automated detection of beaked whale sweeps was performed using spectrogram correlation with frequency bounds of 40–85 kHz and a sweep rate of 0.075 kHz/μs. The detector provided a low false alarm rate ($<10\%$) at the cost of missing a larger percentage ($>75\%$) of sweeps. All detections were reviewed to eliminate false detections. Call sequences occurring less than 0.25 s between sweeps were counted as a single detection; thus a continuous train of sweeps was counted as one detection.

The highest amplitude pulse within each detection was selected for a detailed analysis to minimize range and orientation bias. Pulse modulation was measured by least squares fitting of the instantaneous frequency with linear and second order equations before applying the HARP response function. When signal to noise ratio was poor, the pulse was discarded, leaving about 15 000 pulses to compute start frequency, modulation rate, curvature of modulation rate, and duration.

III. RESULTS

The most common type of pulse (Fig. 1) was often truncated by the 100 kHz bandwidth limit, with some leakage through the antialias filter. The pulse bandwidth [Fig. 1(c)]

a)Electronic mail: mark@whaleacoustics.com
b)Electronic mail: jhildebrand@ucsd.edu
c)Electronic mail: swiggins@ucsd.edu
d)Electronic mail: david.johnston@duke.edu
e)Electronic mail: jeffrey.polovina@noaa.gov

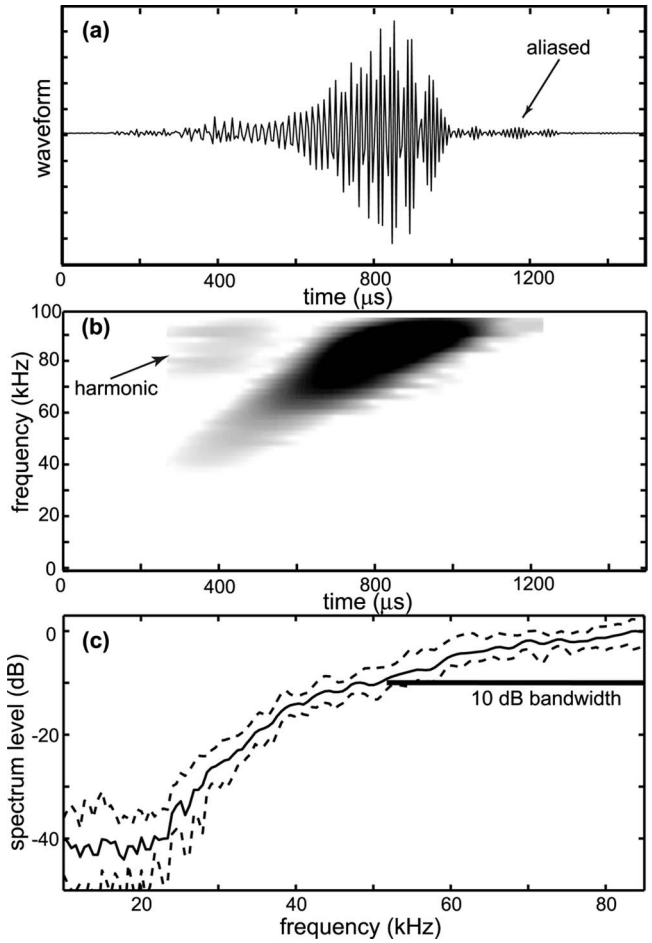


FIG. 1. (a) Waveform and (b) spectrogram (Hann window, 60 sample fast Fourier transform, 59 sample overlap) of the echolocation sweep. (c) Mean received spectrum level for 20 of the highest amplitude echolocation sweeps, with 10% and 90% shown as dashed lines and 10 dB bandwidth after applying the instrument frequency response function. The waveform shown in (a) and the spectrogram (b) are not corrected for instrument response and thus correspond more or less to the signal to noise ratio of the signal, given the instrument response approximately corresponds to the change in ocean ambient noise with frequency. The spectra (c) has been corrected for instrument frequency response.

may shift to higher frequency with higher bandwidth recordings. The 20 highest amplitude signals had an average duration of 987 μs [standard deviation (SD)=82], with a shortening bias because of bandwidth limitations. Received level for these sweeps was 145 dBp.p. re 1 μPa over 50–85 kHz. Other sounds from 50 kHz echo-sounders, sperm whales, and probable pilot whales commonly had higher received levels. Linear and quadratic fitting was applied to 15 211 of the highest amplitude sweeps. The quadratic fits randomly distributed about the linear sweep rate. For the linear fits, the mean start frequency was 35.1 kHz (SD=3.6), the sweep rate was 0.069 kHz/μs (SD=0.019), and the duration (downward biased from bandwidth limitation) was 932 μs (SD=186).

The detector found 25 612 beaked whale echolocation sounds in the first deployment (185 days), and 16 451 in the next (181 days). Manual inspection revealed that about 80% of the sweeps were missed by the detector. The high percentage of missed sweeps was due to seafloor reflections, resulting in smeared arrivals when the animals were near the seafloor. This caused a bias for better detection when arrival angles were more than a few degrees above the horizontal. A bias in seasonal call detection would result if the whales have a seasonal change in feeding depth relative to the seafloor.

Three signal categories were evident: (1) a single echolocation sweep, (2) long duration sweep trains, and (3) short intersweep interval bursts. Approximately 60% of the detections had only a single sweep present, while 40% consisted of sweep trains. The short IPI burst category was less than 0.5% of the total detections. When more than one sweep occurred in sequence (a sweep train), the durations of these sequences had a mean of 0.62 s and a median of 0.35 s. These data were best fit with a lognormal distribution (σ =0.95, μ=−0.95, $K-Sp<0.001$). A subset of sweeps with the highest signal levels had a mean IPI of 110 ms (SD =35).

Short IPI bursts make up a third temporal pattern, their 0.5% occurrence probably being an underestimate. These bursts had a longer IPI between the first two and last two sweeps and are clustered in time. The frequency modulation of the sweeps in the short IPI bursts is nonlinear with decreasing sweep rate toward higher frequencies. The highest amplitude sweeps, selected from each of 227 short IPI bursts, were fitted with quadratic equations. The mean start frequency was 37.2 kHz (SD=7.7), the slope was 0.070 kHz/μs, the curvature was 0.000 018 (SD=0.000 010), and the end frequency was 89.1 kHz (SD=5.7). Unusual examples were found, some of which truncate abruptly near 60 kHz, inconsistent with frequency dependent attenuation, and few of these sweeps exceeded the 100 kHz recording limit. The mean IPI was 14.7 ms with SD of 3.8 ms (n =25) and mean duration of 1145 μs with SD of 282 μs (n =25). The first and last sweep of each sequence had a longer interval and was ignored for the IPI calculation. Multiple animals were producing sweeps 10% of the time when 200 randomly selected sequences were examined. This suggests a highly correlated occurrence, as total detection durations represent only about 0.4% of the total nighttime hours when the recorder was on.

Only 12 feeding buzzes were discovered when examining the 25 612 echolocation detections from the first deployment. These lacked the decreasing IPI of buzzes from Blainville's beaked whales and had no apparent relationship with the echolocation sweeps (Johnson *et al.*, 2008). Buzzes not associated with beaked whale echolocation sweeps were common throughout the recordings; thus the buzzes recorded adjacent to beaked whale echolocation sweeps may be coincidental recordings from another species, such as false killer whales (*Psuedorca crassidens*). During the second deployment more buzzes were coincident with the beaked whale signals, but the overall presence of buzzes was greater, suggesting a seasonal increase in the other species believed to be producing the buzzes.

Detections occurred in about 50% of the nighttime hours monitored, over the duration of the two deployments (Fig. 2). Detections had a strong diel pattern (Fig. 3), with a peak near sunset and nearly all sounds occurring during darkness.

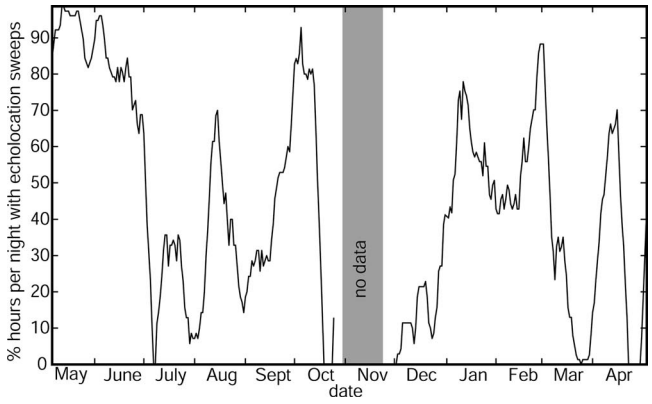


FIG. 2. Daily presence of frequency swept sounds (42 063 verified detections) plotted as the percentage of 1 h intervals in the night which contains one or more detections using a 7 day smoothing filter. Percentage is calculated starting at sunset, ignoring fractional hours near sunrise. Each 1 h time window containing a beaked whale sound was counted, the integer total of these being divided by the integer number of hours in the night.

Sounds nearly stop about 1 h before sunrise. Sounds occur during the day but were too rare to be visible in Fig. 3. Sounds matching Cuvier's beaked whale were not detected and sounds matching Blainville's beaked whales were detected only once [February 11, 2006 at 15:18 Greenwich Mean Time (GMT)].

IV. DISCUSSION

The sweeps reported here have longer durations, higher peak frequencies, shorter IPIs, and greater variability than either Cuvier's or Blainville's beaked whale sounds (Zimmer *et al.*, 2005; Johnson *et al.*, 2006). The Cross Seamount sounds may be from either a geographic variant of Cuvier's or Blainville's beaked whales, Longman's beaked whale (*Indopacetus pacificus*) or another beaked whale species not yet known to occur in this region. The difference in interpulse interval, the relatively shallow water depths, and the strong diel pattern observed here argue against these signals being a geographic variant of Cuvier's or Blainville's echolocation. An additional distinction of the echolocation described here is the association with the short IPI burst sounds.

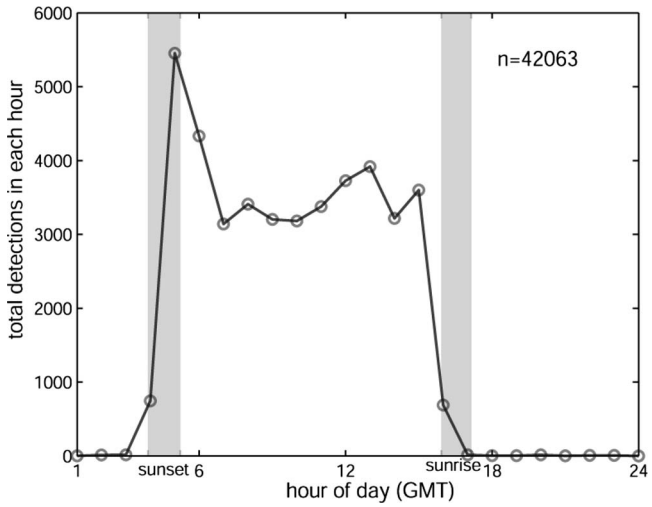


FIG. 3. The diel pattern is shown for all beaked whale sounds. The gray shaded regions show the seasonal range of sunrise and sunset times. The hour of day is in GMT.

The near absence of other beaked whale echolocation sounds at Cross Seamount provides evidence of niche differentiation. Given the relatively short detection ranges for the Cross Seamount beaked whale sounds (<3 km), and their frequent occurrence, it appears that a visual sighting effort may identify the species during calm weather.

The whale species associated with these sounds is present year-around since gaps in detection are brief. Mesoscale oceanographic patterns that might contribute to the observed seasonal variations were considered, but none were found. Previous beaked whale studies have not found evidence of diel foraging patterns (Baird *et al.*, 2008). It is possible that the whales recorded in this study continue to produce sounds during the day but migrate horizontally off the edge of the seamount beyond the detection range of the recorder. This is unlikely because of the abrupt transitions near sunrise and sunset and the near absence of echolocation during the day.

Either or both horizontal and vertical diel migrations of prey species may cause the observed diel foraging activity if the whales feed within the scattering layer. An asymmetry has been observed in scattering layer movements off Hawaii where the layer moves down well prior to sunrise, mirroring the pattern seen in Fig. 3 (Benoit-Bird and Au, 2006) where foraging stops well before sunrise, providing support for this hypothesis. Preliminary results from active acoustic surveys at Cross Seamount show deep scattering layer migration asymmetry (Domokos PIFSC pers. comm. 2008).

Feeding buzzes are of lower amplitude than normal echolocation signals, so it is possible that the beaked whales at Cross Seamount are producing buzzes that were not detected, although this is unlikely. Otherwise we are left with the mystery of how these animals navigate during the terminal phase of prey capture. The short IPI bursts do not show either the decreasing interval or the very short intervals typical of prey capture attempts by Cuvier's beaked whales (Johnson *et al.*, 2008). These short IPI bursts are more reminiscent of codas in sperm whales, which are believed to serve a social function (Rendell and Whitehead, 2004).

Daylight would reach the top of Cross Seamount at sufficient levels for vision. Foraging effectiveness may be compromised during the day if feeding on bioluminescent species may be more difficult to detect or if prey can see their potential predator. The bioluminescent prey argument gains strength from the lack of feeding buzzes in these data. Wood and Evans (1980) found that a blindfolded dolphin could track live fish without echolocating, arguably using passive listening for fish swimming sounds. Perhaps these beaked whales use echolocation to get within passive listening range of prey and then switch to passive acoustics for a capture attempt. Work by Gannon *et al.* (2005) shows dolphins using passive listening to detect the presence of soniferous fish and then switching to echolocation to capture the potential prey. Different species of beaked whales may use both passive and active acoustics to forage, depending on their ecological niche.

Some signals had nearly vertical incidence angles on the recorder, as evidenced by the seafloor echo time delay. Since these whales were located directly above the hydrophone, and the hydrophone was 385 m below the sea surface, the range to the whale was less than 385 m. Assuming a source level the same as Cuvier's beaked whales, the range to the highest amplitude sweeps would be about 385 m. Since the Cross Seamount beaked whales were undoubtedly closer than 385 m, their source levels cannot be higher than those of Cuvier's beaked whale and are probably considerably lower. Harmonics are present with some signals [Fig. 1(b)] but are not always associated with high amplitude signals, suggesting that variability at the source controls the presence of harmonics.

Baird, R. W., Webster, D. L., Schorr, G. S., McSweeney, D. J., and Barlow, J. (**2008**). "Diel variation in beaked whale diving behavior," Marine Mammal Sci. **24**, 630–642.

Benoit-Bird, K. J., and Au, W. W. L. (**2006**). "Extreme diel horizontal migrations by a tropical nearshore resident micronekton community," Mar. Ecol.: Prog. Ser. **319**, 1–14.

Gannon, D. P., Barros, N. B., Nowacek, D. P., Read, A. J., Waples, D. M., and Wells, R. S. (**2005**). "Prey detection by bottlenose dolphins (*Tursiops truncatus*): An experimental test of the passive listening hypothesis," Anim. Behav. **69**, 709–720.

Johnson, M., Hickmott, L. S., Aguilar Soto, N., and Madsen, P. T. (**2008**). "Echolocation behaviour adapted to prey in foraging Blainville's beaked whale (*Mesoplodon densirostris*)," Proc. R. Soc. London, Ser. B **275**, 133–139.

Johnson, M., Madsen, P., Zimmer, W., Aguilar de Soto, N., and Tyack, P. (**2006**). "Foraging Blainville's beaked whales (*Mesoplodon densirostris*) produce distinct click types matched to different phases of echolocation," J. Exp. Biol. **209**, 5038–5050.

Johnston, D. W., McDonald, M., Polovina, J., Domokos, R., Wiggins, S., and Hildebrand, J. (**2008**). "Temporal patterns in the acoustic signals of beaked whales at Cross Seamount," Biol. Lett. **4**, 208–211.

Rendell, L., and Whitehead, H. (**2004**) "Do sperm whales share coda vocalizations? Insights into coda usage from acoustic size measurement," Anim. Behav. **67**, 865–874.

Wiggins, S. M., and Hildebrand, J. A. (**2007**). "High-frequency Acoustic Recording Package (HARP) for broad-band, long-term marine mammal monitoring," Underwater Technology Symposium (UT07)/Workshop on Scientific Use of Submarine Cables and Related Technologies (SSC07), Tokyo, Japan, April 2007, pp 551–557.

Wood, F. G., and Evans., W. E. (**1980**). "Adaptiveness and ecology of echolocation in toothed whales," in *Animal Sonar Systems*, edited by R. Busnel and J. Fish (Plenum, New York), pp. 381–426.

Zimmer, W., Johnson, M., Madsen, P., and Tyack, P. (**2005**). "Echolocation clicks of free-ranging Cuvier's beaked whales (*Ziphius cavirostris*)," J. Acoust. Soc. Am. **117**, 3919–3927.

Trapping of shear acoustic waves by a near-surface distribution of cavities (L)

C. Aristégui, A. L. Shuvalov, O. Poncelet, and M. Caleap
Université de Bordeaux, UMR 5469, Laboratoire de Mécanique Physique, Talence, F-33405, France and CNRS, UMR 5469, Talence, F-33405, France

(Received 5 September 2008; revised 27 November 2008; accepted 4 December 2008)

For a halfspace containing random and uniform distribution of empty cylindrical cavities within finite depth beneath the surface, the dispersion spectrum of coherent shear horizontal waves is calculated and analyzed based on the effective-medium approach. The scattering-induced dispersion and attenuation are coupled with the effect of a surface waveguide filled with scatterers. As a result, the obtained spectrum bears certain essential particularities in comparison with the standard Love-wave pattern. Simple analytical estimates enable a direct evaluation of the concentration of scatterers from the dispersion data. © *2009 Acoustical Society of America.*
[DOI: 10.1121/1.3056565]

PACS number(s): 43.35.Cg [PEB] Pages: 628–631

I. INTRODUCTION

The paper is concerned with the shear horizontal (SH) wave propagation in an isotropic halfspace containing multiple cylindrical cavities near its free surface. The cavities reduce the velocity and thus create a surface waveguide. An intrinsic frequency dispersion due to the scattering is superposed with that due to the wave trapping beneath the surface. An evident analogy with the Love waves suggests using a dynamic-homogenization approach developed in Ref. 1 and based on the classical Waterman-Truell theory.[2] It replaces the actual material with cavities by an appropriate "effective" one, whose elastic parameters with respect to the coherent SH wave motion are spatially constant but frequency dispersive and complex valued. Based on this approach, we calculate the dispersion spectrum of SH waves in a given halfspace and examine its particular properties, which either do not arise or have not been considered in the physically similar cases of a viscous[3,4] or a porous[5] layer on a substrate.

II. BACKGROUND

Suppose that mutually parallel cylindrical cavities of the radius a are distributed randomly and uniformly under the free surface of an isotropic halfspace up to a certain depth h ($\gg a$) (see Fig. 1). The concentration ϕ of cavities over the layer of thickness h is assumed small. In view of the coherent SH wave propagation with frequency ω in the direction orthogonal to the axes of cavities, the given medium may be seen as a transversely isotropic homogeneous layer of the effective material bonded to a substrate of the matrix material. Denote the density and the shear modulus of the matrix by ρ_2 and μ_2. The density ρ_1 and the shear modulus μ_1 of the layer depend on the concentration ϕ and on the scattering dispersion parameter

$$\tilde{\omega} = \omega s_2 a, \tag{1}$$

where $s_2 = \sqrt{\rho_2/\mu_2}$ is the slowness of the bulk shear wave in the matrix material. According to Ref. 1, this dependence is as follows:

$$\rho_1(\tilde{\omega}) = \rho_2(1-\phi)[1+\phi r(\tilde{\omega})],$$
$$\mu_1(\tilde{\omega}) = \frac{\mu_2}{1+2\phi}[1+\phi m(\tilde{\omega})], \tag{2}$$

where $r(\tilde{\omega})$ and $m(\tilde{\omega})$ are certain complex-valued functions, which turn to zero along with their first derivatives at $\tilde{\omega}=0$. They are defined in full in Ref. 1. To the leading order in low frequency such that $\tilde{\omega}^2 \ln \tilde{\omega} \ll 1$,

$$r(\tilde{\omega}) = -\frac{1}{2(1-\phi)}\tilde{\omega}^2\left\{[\ln \tilde{\omega} + O(1)] + i\left[-\frac{\pi}{2} + O(\tilde{\omega}^2 \ln \tilde{\omega})\right]\right\},$$
$$m(\tilde{\omega}) = \frac{1}{1+2\phi}\tilde{\omega}^2\left\{[\ln \tilde{\omega} + O(1)] + i\left[-\frac{\pi}{2} + O(\tilde{\omega}^2 \ln \tilde{\omega})\right]\right\}, \tag{3}$$

where the next-order terms $O(\cdot)$ in $r(\tilde{\omega})$ and $m(\tilde{\omega})$ are different. Note that the signs $\mathrm{Im}\, r(\tilde{\omega}) > 0$ and $\mathrm{Im}\, m(\tilde{\omega}) < 0$ are in agreement with the sign of dissipation. Evidently the layer and substrate (matrix) parameters differ in the measure of concentration of scatterers ϕ, so the overall effect in question is gauged by a small ϕ.

By combining Eqs. (2) and (3), the complex-valued slowness of shear bulk wave in the layer $s_1(\tilde{\omega}) = \sqrt{\rho_1(\tilde{\omega})/\mu_1(\tilde{\omega})}$ is

$$s_1(\tilde{\omega}) = s_2\sqrt{(1-\phi)(1+2\phi)\frac{1+\phi r(\tilde{\omega})}{1+\phi m(\tilde{\omega})}}. \tag{4}$$

For small $\tilde{\omega}$, the phase velocity and the scattering-induced attenuation

$$c_1(\tilde{\omega}) = \mathrm{Re}[s_1^{-1}(\tilde{\omega})], \quad \alpha_1(\omega) = \omega\, \mathrm{Im}\, s_1(\tilde{\omega}) \tag{5}$$

of shear bulk waves in the layer are approximated as follows:

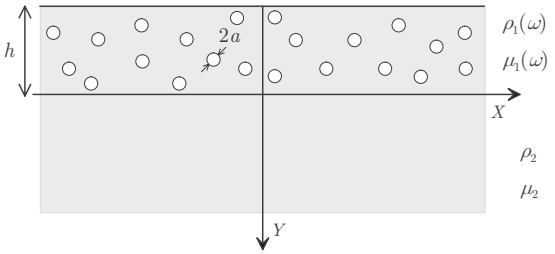


FIG. 1. Geometry of the problem and the material parameters.

$$c_1(\tilde{\omega}) = c_2 \frac{1}{\sqrt{(1-\phi)(1+2\phi)}} \times \left\{ 1 + \frac{3\phi}{4(1-\phi)(1+2\phi)} \tilde{\omega}^2 [\ln \tilde{\omega} + O(1)] \right\}, \tag{6a}$$

$$\alpha_1(\omega) = \frac{3\pi\phi}{8a\sqrt{(1-\phi)(1+2\phi)}} \tilde{\omega}^3 [1 + O(\tilde{\omega}^2 \ln \tilde{\omega})], \tag{6b}$$

where $c_2 = s_2^{-1}$. The high-frequency asymptotic trend of the layer parameters is

$$\rho_1(\tilde{\omega}) = \rho_2 \left(1 + i \frac{2\phi}{\pi\tilde{\omega}} \right), \quad \mu_1(\tilde{\omega}) = \mu_2 \left(1 + i \frac{2\phi}{\pi\tilde{\omega}} \right)^{-1}, \tag{7}$$

i.e., they tend to the matrix values ρ_2 and μ_2, whence $c_1(\tilde{\omega})$ tends to c_2 while the attenuation $\alpha_1(\omega)$ approaches a constant value $\alpha_1(\infty) = 2\phi/\pi a$.[1]

III. THE DISPERSION SPECTRUM

The guided waves are sought in the form $u_z(x,y) = U(y)\exp[i\omega(s_x x - t)]$, where the frequency ω is kept real. Taking into account the traction-free condition at the upper surface and the continuity at the bonded interface yields the dispersion equation for s_x^2,

$$i\sqrt{s_1^2(\omega) - s_x^2} \tan(\omega h \sqrt{s_1^2(\omega) - s_x^2}) = \frac{\mu_2 s_{y2}}{\mu_1(\omega)}. \tag{8}$$

Unlike the standard case of Love waves,[6] the layer parameters s_1 and μ_1 are dispersive and also complex valued, hence all the solutions s_x of Eq. (8) are generally complex valued. Equation (8) admits different families of formal solutions related to two Riemann sheets of $s_{y2} = \sqrt{s_2^2 - s_x^2}$. We will be concerned only with the Love-type dispersion branches $s_x^{(n)}(\omega)$, $n = 0, 1, \ldots$, which describe waves decaying into the depth of the substrate. Imposing the inequality condition

$$\mathrm{Im}\, s_{y2}^{(n)} = \mathrm{Im}\, \sqrt{s_2^2 - s_x^{(n)2}} \geq 0, \tag{9}$$

we observe that the found solutions satisfy $\mathrm{Im}\, s_x^{(n)} \geqslant 0$ (for $\mathrm{Re}\, s_x^{(n)} > 0$ and the axes X, Y as in Fig. 1), i.e., they also decrease along the propagation direction due to the layer attenuation.

Figure 2 shows the Love-type branches calculated for the concentration $\phi = 8\%$ of cavities with the radius $a = 0.06$ mm, which are distributed within the layer of thickness $h = 1$ mm in the aluminium matrix with $\rho_2 = 2.7$ g/cm^3 and $\mu_2 = 26.45$ GPa. The results are displayed in terms of the phase velocity and the attenuation

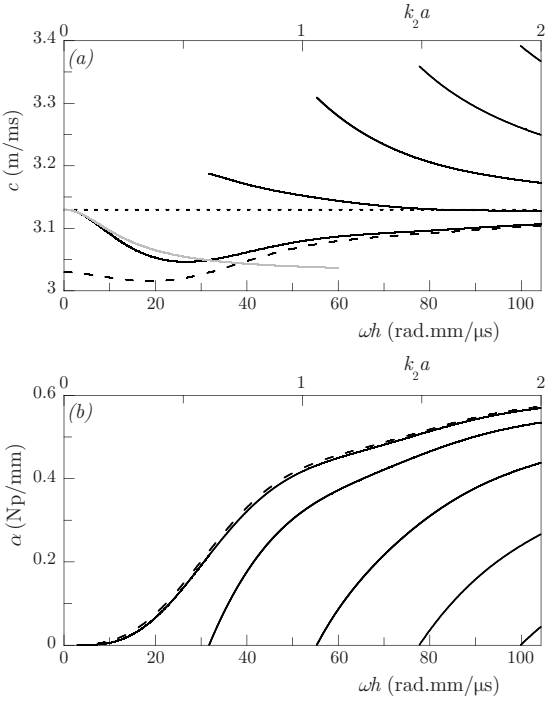


FIG. 2. Dispersion branches of (a) the phase velocity $c^{(n)}$ and (b) the attenuation $\alpha^{(n)}$ for the Love-type coherent waves guided by the near-surface distribution of cavities. The input data are specified in the text. Solid curves are dispersion branches, Eq. (10); dashed curves are the velocity c_1 and the attenuation α_1 of bulk waves in the effective layer material; a dotted horizontal line is the velocity c_2 for the matrix material (substrate). The gray curve is an approximation for the fundamental branch calculated from Eq. (8) with statically averaged constants.

$$c^{(n)} = \mathrm{Re}(1/s_x^{(n)}), \quad \alpha^{(n)} = \omega\, \mathrm{Im}\, s_x^{(n)}, \tag{10}$$

which are plotted as functions of $\tilde{\omega} = k_2 a$ [$k_2 = \omega s_2$, see Eq. (1)] and of ωh. The frequency band shown in Fig. 2 ($\tilde{\omega} \leqslant 2$) is within typical range of application of homogenization modeling, see, e.g., Refs. 7 and 8. Figure 3 presents the wave-vector vertical component k_{2y} in the substrate. Its imaginary part $\mathrm{Im}\, k_{2y}$ governs the amplitude decay $\exp(-\mathrm{Im}\, k_{2y} y)$ into the substrate depth.

IV. DISCUSSION

A. Fundamental branch

It is noted that the "waveguiding" dispersion governed by $\omega s_0 h$ is h/a times stronger than that due to the scattering dispersion parameter $\tilde{\omega}$. Given $h/a \gg 1$, the fundamental phase-velocity curve $c^{(0)}(\omega)$ for small $\tilde{\omega}$ can be approximately defined via confining in Eq. (8) the effective parameters $s_1^2(\omega)$ and $\mu_1(\omega)$ by their statically averaged values $s_1^2(0) = s_2^2(1-\phi)(1+2\phi)$ and $\mu_1(0) = \mu_2/(1+2\phi)$ following from Eq. (2) (see the gray curve in Fig. 2). For another long-wave scale which is $(\omega h s_2)^2 \ll 1$, the leading order explicit estimate is

$$c^{(0)}(\omega) \approx c_2 \left[1 - \frac{1}{2} (\omega h s_2)^2 \frac{\phi^2 (1-2\phi)^2}{(1+2\phi)^2} \right]. \tag{11}$$

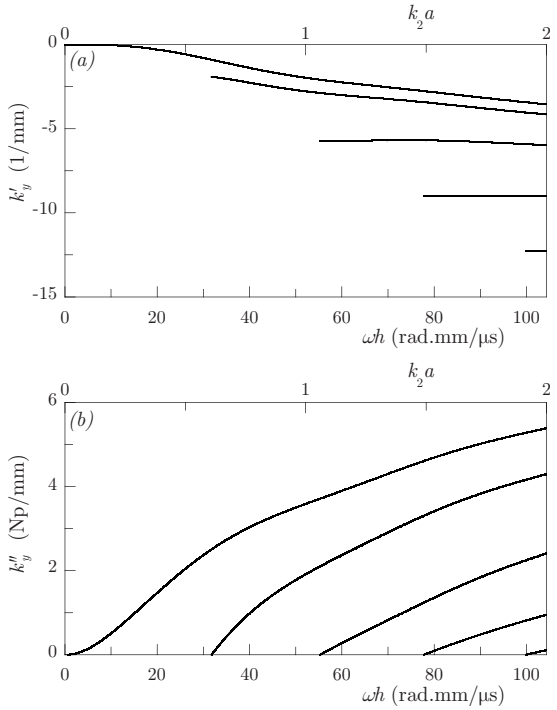


FIG. 3. The normal component k_{2y} of wave vector in the substrate for the Love-type branches presented in Fig. 2: (a) the real and (b) the imaginary parts.

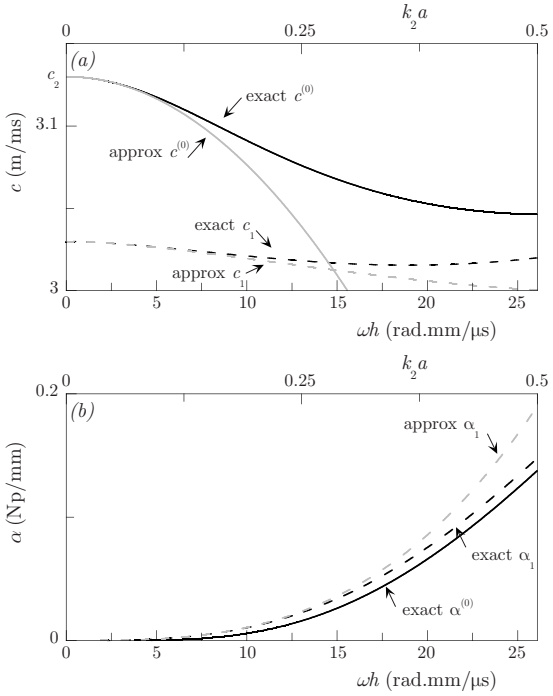


FIG. 4. Long-wave exact and approximate curves for (a) the velocity and (b) the attenuation of the bulk waves in the layer (c_1 and α_1) and of the fundamental Love-type waves ($c^{(0)}$ and $\alpha^{(0)}$). See Eqs. (6), (11), and (12).

Starting from high enough frequency, the fundamental branch $c^{(0)}(\omega)$ trails above the bulk-wave velocity in the layer $c_1(\omega)$. This is as usual. However, for the case in hand $c_1(\omega)$ is dispersive. According to Eq. (6a), it reaches minimum at about the minimum point $\tilde{\omega}=e^{-1/2}$ of the function $\tilde{\omega}^2 \ln \tilde{\omega}$ and then curves upwards. Hence so does the branch $c^{(0)}(\omega)$.

The attenuation curve $\alpha^{(0)}(\omega)$ corresponding to the fundamental branch is close from below to the attenuation $\alpha_1(\omega)=\omega \operatorname{Im} s_1(\omega)$ of shear bulk waves in the layer. For $\tilde{\omega}^2 \ln \tilde{\omega} \ll 1$, the latter is approximated through Eq. (6b), so that

$$\alpha^{(0)}(\omega) \lesssim \alpha_1(\omega) \approx \frac{3\pi\phi\tilde{\omega}^3}{8a\sqrt{(1-\phi)(1+2\phi)}}. \tag{12}$$

The long-wave approximations (11) and (12) are shown in Fig. 4.

B. Nonfundamental branches

Scattering-induced attenuation underlies an unusual layout of the origin points (cutoffs) of the dispersion curves with $n>0$. Denote the frequency and velocity at these points by $\omega_c^{(n)}$ and $c_c^{(n)} \equiv c^{(n)}(\omega_c^{(n)})$. In the absence of absorption, all the cutoffs lie on the constant line c_2 corresponding to the grazing propagation $k_{2y}=0$ in the substrate.[6] This is no longer the case due to the layer attenuation.

The Love-type branches in hand satisfy the conditions $\operatorname{Im} s_{y2}^{(n)} \geq 0$ and $\operatorname{Im} s_x^{(n)} \geq 0$. For a purely elastic substrate (with real s_2), $\operatorname{Im} s_x^{(n)}$ and $\operatorname{Im} s_{y2}^{(n)}$ vanish simultaneously. Thus the cutoff points $(\omega_c^{(n)}, c_c^{(n)})$ occur when

$$\operatorname{Im} s_x^{(n)} = 0 \Leftrightarrow \operatorname{Im} s_{y2}^{(n)} = 0, \tag{13}$$

i.e., the imaginary part of the slowness vector $\mathbf{s}$ in the substrate turns to zero. Its real part has positive components $\operatorname{Re} s_x^{(n)} = 1/c_c^{(n)}$ and $\operatorname{Re} s_{y2}^{(n)} = -\sqrt{s_2^2 - s_x^{(n)2}}$, whence $c_c^{(n)} > c_2$. That is why the cutoff points (13) lie above the substrate velocity c_2. Beyond the cutoffs, the Love-type velocity and attenuation branches shown in Fig. 2 are continued by the branches of nonphysical solutions with $\operatorname{Im} s_x^{(n)} < 0$ and $\operatorname{Im} s_{y2}^{(n)} < 0$, which increase both into the depth and along the propagation direction.

Figure 2 shows that the cutoff velocity $c_c^{(n)}$ increases for the first few branches. It is also seen that all the attenuation branches $\alpha^{(n)}(\omega)$ with $n>0$, which start from zero value at the successive cutoff frequencies $\omega_c^{(n)}$, increase with growing ω in a similar manner. The high-frequency extent of all the branches $c^{(n)}(\omega)$ and $\alpha^{(n)}(\omega)$ approach the limits c_2 and $\alpha_1(\infty)$ [see Eq. (7)].

The cutoff frequency and velocity values $\omega_c^{(n)}$ and $c_c^{(n)}$ imply a small imaginary part of Eq. (8) and hence they may be approximately expressed through the layer and substrate parameters as follows:

$$\omega_c^{(n)} h \sqrt{[s_1'(\omega_c^{(n)})]^2 - \frac{1}{c_c^{(n)2}}} \approx \pi n \tag{14}$$

and

$$\frac{\pi n}{\omega_c^{(n)} h} \tanh\left[\frac{\omega_c^{(n)} h^2}{\pi n} \alpha_1(\omega_c^{(n)}) \mathrm{Re}\, s_1(\omega_c^{(n)})\right] \approx \frac{\mu_2}{\mathrm{Re}\, \mu_1(\omega_c^{(n)})} \sqrt{s_2^2 - \frac{1}{c_c^{(n)2}}}. \quad (15)$$

The cutoff velocity $c_c^{(n)}$, which first increases (see Fig. 2), is then decreasing very slowly with further growing n. Thus the high-order cutoffs occur at an almost constant velocity $c_c^{(n)} \approx c_c^{(\infty)}$ satisfying Eq. (14) with s_2 in place of $s_1(\omega)$, and at an almost equidistant frequency step:

$$\omega_c^{(n+1)} - \omega_c^{(n)} \approx \frac{\pi}{h\sqrt{s_2^2 - \frac{1}{c_c^{(\infty)2}}}}, \quad \frac{\omega_c^{(n+1)}}{\omega_c^{(n)}} \approx \frac{n+1}{n}. \quad (16)$$

Equation (15) is basically equivalent to Eq. (14) for the high-order cutoffs, when $\mu_1(\omega)$ approaches μ_2 and the hyperbolic tangent in Eq. (15) becomes close to 1. It therefore varies too slow for an accurate evaluation of its argument from its value inferred from Eq. (15). In this regard, Eq. (15) is more useful for the first one or several cutoffs where the hyperbolic tangent is yet small enough.

V. CONCLUSIONS

The effective-medium approach has been used for calculating the spectrum of coherent SH waves in a halfspace which contains random and uniform distribution of cylindrical cavities within finite depth beneath the surface. It is shown that the effect of the dispersion and attenuation caused by the scattering leads to some unusual spectral features such as a curved high-frequency asymptotic of the fundamental branch and supersonic cutoffs of the higher-order branches starting above the substrate velocity.

[1] C. Aristégui, Y. C. Angel, and Z. E. A. Fellah, "Using coherent waves to evaluate dynamic material properties," in *Proceedings of the 5th World Congress on Ultrasonics* Paris, France, 7–10 September 2003, p. 463; C. Aristégui and Y. C. Angel, "Effective material properties for shear-horizontal acoustic waves in fiber composites," Phys. Rev. E **75**, 056607 (2007).

[2] P. C. Waterman and R. Truell, "Multiple scattering of waves," J. Math. Phys. **2**, 512–537 (1961).

[3] P. Kielczyński, "Attenuation of Love waves in low-loss media," J. Appl. Phys. **82**, 5932–5937 (1997).

[4] G. McHale, M. I. Newton, and F. Martin, "Theoretical mass, liquid, and polymer sensitivity of acoustic wave sensors with viscoelastic guiding layers," J. Appl. Phys. **93**, 675–690 (2003).

[5] L.-L. Ke, Y.-S. Wang, and Z.-M. Zhang, "Propagation of Love waves in an inhomogeneous fluid saturated porous layered half-space with properties varying exponentially," J. Eng. Mech. **131**, 1322 (2005); "Love waves in an inhomogeneous fluid saturated porous layered half-space with linearly varying properties," Soil Dyn. Earthquake Eng. **26**, 574 (2006).

[6] J. L. Rose, *Ultrasonic Waves in Solid Media* (Cambridge U.P., Cambridge, 1999), Chap. 10.

[7] J. Mobley, K. R. Waters, C. S. Hall, J. N. Marsh, M. S. Hughes, G. H. Brandenburger, and J. M. Miller, "Measurements and predictions of the phase velocity and attenuation coefficient in suspensions of elastic microspheres," J. Acoust. Soc. Am. **106**, 652–659 (1999).

[8] D. G. Aggelis, D. Polyzos, and T. P. Philippidis, "Wave dispersion and attenuation in fresh mortar: theoretical predictions vs. experimental results," J. Mech. Phys. Solids **53**, 857–883 (2005).

Validation of theoretical models of phonation threshold pressure with data from a vocal fold mechanical replica (L)

Jorge C. Lucero[a)]
Department of Mathematics, University of Brasilia, Brasilia DF 70910-900, Brazil

Annemie Van Hirtum,[b)] Nicolas Ruty,[c)] Julien Cisonni,[d)] and Xavier Pelorson[e)]
GIPSA-lab, UMR CNRS 5216, Grenoble Universities, 961 rue de la Houille Blanche, BP 46, 38402 Saint-Martin d'Heres, France

(Received 20 May 2008; revised 26 November 2008; accepted 1 December 2008)

This paper analyzes the capability of a mucosal wave model of the vocal fold to predict values of phonation threshold lung pressure. Equations derived from the model are fitted to pressure data collected from a mechanical replica of the vocal folds. The results show that a recent extension of the model to include an arbitrary delay of the mucosal wave in its travel along the glottal channel provides a better approximation to the data than the original version of the model, which assumed a small delay. They also show that modeling the vocal tract as a simple inertive load, as has been proposed in recent analytical studies of phonation, fails to capture the effect of the vocal tract on the phonation threshold pressure with reasonable accuracy.
 [DOI: 10.1121/1.3056468]

PACS number(s): 43.70.Aj, 43.70.Bk, 43.70.Jt [CHS] Pages: 632–635

I. INTRODUCTION

The phonation threshold of lung pressure is defined as the minimum value required to initiate vocal fold oscillation. It is an important factor for building empirical laws of laryngeal aerodynamics (Titze, 1992) and represents the pressure level at which the energy transferred from the airflow to the vocal folds is large enough to overcome the energy dissipated in the tissues, so that an oscillatory movement of growing amplitude may take place (Lucero, 1999). The phonation threshold pressure value has also been interpreted as a measure of ease of phonation and proposed as a diagnostic tool for vocal health (Titze *et al.*, 1995).

Two decades ago, Titze (1988) derived an equation for the phonation threshold pressure by modeling the vocal fold oscillatory movement as a superficial mucosal wave propagating in the direction of the airflow. The equation related the threshold pressure to biomechanical parameters, namely, glottal geometry, tissue damping coefficient, and mucosal wave velocity. However, it lacked the oscillation frequency as an explicit parameter. It is well known that phonation threshold pressure increases with frequency, as demonstrated by experimental measures (e.g., Titze, 1992). In his works, Titze (1988, 1992) pointed out the missing parameter and offered a possible solution by relating the vocal fold thickness and mucosal wave velocity to the oscillation frequency.

In a recent paper (Lucero and Koenig, 2007), it was shown that the lack of the frequency factor is a consequence of one of the simplifications made in the vocal fold model: the assumption of a small time delay for the mucosal wave to travel along the vertical dimension of the vocal folds. A more general analysis for an arbitrary time delay results in an extended equation for the phonation threshold pressure, which includes the oscillation frequency explicitly.

Because a direct validation, using *in vivo* measurements on human speakers, of these theoretical predictions cannot be achieved easily, we propose to test them against *in vitro* experiments using a mechanical replica of the vocal folds. Mechanical replicas of the voice production system, such as the one introduced by Ruty *et al.* (2007), allow us to test theoretical models against experimental data quantitatively and to extract conclusions about the range of validity of those models.

II. EXTENSION OF THE MUCOSAL WAVE MODEL

Figure 1 shows a schematic of the mucosal wave model. Complete right-left symmetry of the folds is assumed, and motion of tissues is allowed only in the horizontal direction. A surface wave propagates through the superficial tissues, in the direction of the airflow (upward).

The equation of motion of the vocal fold tissues is obtained by lumping their biomechanical properties at the midpoint of the glottis and assuming that they are forced by the mean glottal pressure P_g, which yields

$$M\ddot{\xi} + B\dot{\xi} + K\xi = P_g, \tag{1}$$

where ξ is the tissue displacement at the midpoint, and M, B, K, are the mass, damping, and stiffness, respectively, per unit area of the medial surface of the vocal folds.

The glottal aerodynamics is modeled by assuming that the flow is frictionless, stationary, and incompressible. Further, we assume that the subglottal pressure is equal to a constant lung pressure P_L, the vocal tract input area is much

[a)]Electronic mail: lucero@unb.br
[b)]Electronic mail: annemie.vanhirtum@gipsa-lab.inpg.fr
[c)]Electronic mail: nicolas.ruty@gipsa-lab.inpg.fr
[d)]Electronic mail: julien.cisonni@gipsa-lab.inpg.fr
[e)]Electronic mail: pelorson@icp.inpg.fr

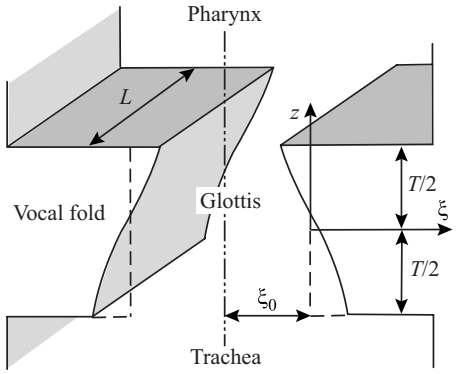


FIG. 1. Vocal fold model (after Titze, 1988).

larger than the glottal area, and the prephonatory glottal channel is rectangular. Under such conditions, the mean glottal air pressure P_g may be expressed as

$$P_g = P_i + (P_L - P_i)(1 - a_2/a_1)/k_t, \tag{2}$$

where P_i is the supraglottal pressure (at the entry of the vocal tract), k_t is a transglottal pressure coefficient, and a_1, a_2 are the glottal areas at the lower and upper edges of the glottal channel, respectively, given by

$$a_1(t) = 2L[\xi_0 + \xi(t+\tau)], \tag{3}$$

$$a_2(t) = 2L[\xi_0 + \xi(t-\tau)], \tag{4}$$

where ξ_0 is the prephonatory glottal half-width, τ is the time delay for the mucosal wave to travel half the glottal height ($T/2$ in Fig. 1), and L is the vocal fold length.

Following Chan and Titze (2006), the input pressure to the vocal tract is modeled as $P_i \approx I\dot{u}$, where I is the vocal tract inertance and $\dot{u}$ is the time derivative of the airflow. This approximation is valid when the oscillation frequency of the vocal folds (F_0) is below the first formant (F_1) of the vocal tract (Titze, 1988). For a quasisteady flow condition and small amplitude oscillations around an abducted (open) glottis, the flow derivative may be approximated by $\dot{u} \approx v_2 \dot{a}_2$, where $v_2 = \sqrt{2P_L/(k_t\rho)}$ is the air particle velocity at the glottal exit, ρ is the air density, and a_2 is the glottal area at the upper edge of the vocal folds, given by Eq. (4). Therefore, we have the approximation

$$P_i = 2LIv_2\xi'(t-\tau), \tag{5}$$

where $\xi'(t-\tau) = d\xi/d(t-\tau)$.

With the above assumptions, the mean glottal air pressure is then

$$P_g = 2LIv_2\xi'(t-\tau) + \left[\frac{P_L - 2LIv_2\xi'(t-\tau)}{k_t}\right] \times \left[\frac{\xi(t+\tau) - \xi(t-\tau)}{\xi_0 + \xi(t+\tau)}\right]. \tag{6}$$

The equation of motion for the vocal fold oscillation is then given by Eqs. (1) and (6). More details on the assumptions of the model and the derivation of the equations may be easily found in the cited references.

III. OSCILLATION THRESHOLD PRESSURE

Equations (1) and (6) constitute a functional differential equation with advance and delay arguments ($t+\tau$ and $t-\tau$, respectively). It has a unique fixed point at $\xi=0$, which corresponds to the prephonatory position.

Linearization around that position produces

$$M\ddot{\xi} + B\dot{\xi} + K\xi = 2LIv_2\xi'(t-\tau) + \frac{P_L}{\xi_0 k_t}[\xi(t+\tau) - \xi(t-\tau)]. \tag{7}$$

Proposing a solution of the form $\xi(t) = Ce^{\lambda t}$, where C and λ are complex constants, and seeking nonzero solutions produces the associated characteristic equation

$$M\lambda^2 + B\lambda + K - 2LIv_2\lambda e^{-\lambda\tau} - \frac{2P_L}{k_t\xi_0}\sinh(\lambda\tau) = 0. \tag{8}$$

Let P_{th} denote the phonation threshold value of the lung pressure P_L, at which the vocal fold oscillation starts. At the threshold, a pair of complex roots of the above equation cross the imaginary axis from left to right. Next, letting $\lambda = i\omega$, $P_L = P_{\text{th}}$, and separating real and imaginary parts, we obtain the conditions

$$-\omega^2 M + K - 2LIv_2\omega\sin(\omega\tau) = 0, \tag{9}$$

$$\omega B - 2LIv_2\omega\cos(\omega\tau) - \frac{2P_{\text{th}}}{k_t\xi_0}\sin(\omega\tau) = 0, \tag{10}$$

and, from Eq. (10), we obtain

$$P_{\text{th}} = \frac{k_t\xi_0 B\omega}{2\sin(\omega\tau)} - k_t\xi_0 lIv_2\omega\cot(\omega\tau), \tag{11}$$

where $0 < (\omega\tau) < \pi$, and v_2 is computed at the threshold condition, i.e., $v_2 = \sqrt{2P_{\text{th}}/(k_t\rho)}$.

If we ignore the effect of the vocal tract by setting $I = 0$ (no vocal tract load), we obtain

$$P_{\text{th}} = \frac{k_t\xi_0 B\omega}{2\sin(\omega\tau)}, \tag{12}$$

which is the equation found by Lucero and Koenig (2007). For $\tau \to 0$, $\sin(\omega\tau) \to \omega\tau$. Eq. (12) simplifies further to Titze's (1988) result

$$P_{\text{th}} = \frac{k_t\xi_0 B}{2\tau}. \tag{13}$$

Note also that a Taylor expansion of Eq. (12) around $\omega=0$ produces

$$P_{\text{th}} = \frac{k_t\xi_0 B}{2}\left(\frac{1}{\tau} + \frac{\tau}{6}\omega^2 + \mathcal{O}(\omega^4)\right). \tag{14}$$

Keeping only the first two terms, we obtain a quadratic approximation to P_{th} in terms of ω, as proposed by Titze (1992).

Considering now $\tau \to 0$ in Eq. (11), and therefore $\sin(\omega\tau) \to \omega\tau$ and $\cot(\omega\tau) \to 1/(\omega\tau)$, we obtain

$$P_{\text{th}} = \frac{k_t \xi_0 B}{2\tau} - \frac{k_t \xi_0 L I v_2}{\tau}, \tag{15}$$

which is the equation found by Chan and Titze (2006).

IV. DATA

To test the above results, we used data collected from a mechanical replica for a previous study by Ruty *et al.* (2007, Figs. 8, 9 and 10 of their paper). The replica consists of two metal half-cylinders covered with latex, which mimic the vocal fold structure in a 3:1 scale, with a similar aspect ratio. Geometrical dimensions and other parameters of the replica were chosen in order to match as closely as possible the glottal aerodynamics (see Table I of Ruty *et al.*, 2007). The cylinders are filled with water, at a controlled internal pressure P_c. The initial separation between the latex tubes decreases when P_c is increased, and becomes zero for $P_c > 5000$ Pa. The vocal tract is simulated with a downstream cylindrical resonator. Two different tubes were used, with a diameter of 25 mm, and lengths of 250 mm and 500 mm, respectively. Their dimensions were chosen in order to present a weak and a strong acoustical coupling. The first acoustical resonances of the tubes are 340 Hz, for the 250 mm tube, and 170 Hz, for the 500 mm tube. Those resonance frequencies are, respectively, higher than and comparable to the oscillation frequency of the latex structure, which is in the range of 110–170 Hz.

Measures of oscillation threshold pressure were obtained by increasing the air pressure upstream of the vocal fold replica until an oscillation of the latex structures was detected. The oscillation frequency at the oscillation onset was then computed by spectral analysis on the acoustic output signal. A threshold pressure for the oscillation offset was also measured, by decreasing the upstream pressure until the oscillation was interrupted, but those values are not used here. This procedure was repeated for various values of the water pressure P_c, and for the two cylindrical resonators.

For our analysis, we ignored all data for $P_c > 5000$ Pa, because in that range the latex tubes are in contact ($\xi_0 = 0$), and consequently the above equations produce $P_{\text{th}} = 0$. Let us also recall that the mucosal wave model assumes an open prephonatory glottis, wide enough so that the effect of air viscosity may be neglected (Titze, 1988).

V. NUMERICAL RESULTS

We fitted the above theoretical equations to Ruty *et al.*'s (2007) data by a standard least squares procedure implemented in Matlab, with the oscillation threshold pressure P_{th} as the target.

In a first numerical experiment, we fitted Eq. (12) to each resonator's data, with $(k_t B)$ and τ as parameters; the results are shown in Fig. 2. The computed optimal values were $(k_t B) = 350.81$ Pa s/m, $\tau = 2.66$ ms, and $(k_t B) = 1864.0$ Pa s/m, $\tau = 2.90$ ms for the 250 and 500 mm resonators, respectively. For comparison, we also fitted Eq. (13), obtaining $(k_t B) = 3436.4$ Pa s/m, $\tau = 7.12$ ms, and $(k_t B) = 248.68$ Pa s/m, $\tau = 0.0962$ ms for the 250 and 500 mm resonators, respectively. As shown by the plots, our extended

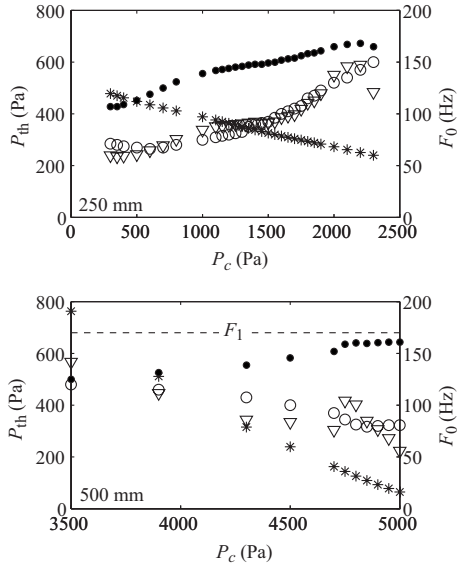


FIG. 2. Oscillation threshold pressure P_{th} and frequency F_0 vs internal pressure P_c for a 250 mm resonator (upper panel) and 500 mm resonator (lower panel). Circles: measured pressure values; triangles: theoretical pressure values given by Eq. (12), stars: theoretical pressure values given by Eq. (13); filled circles: measured oscillation frequency. The broken line in the lower panel indicates the first acoustical resonance (F_1) of the resonator.

equation (12) provides a reasonably good approximation for both resonators, better than Eq. (13). In case of the 250 mm resonator, Eq. (13) produces a decreasing P_{th} pattern, instead of the measured increasing pattern, because ξ_0 decreases when P_c increases (Ruty *et al.*, 2007, Fig. 8). The extended Eq. (12), on the other hand, is able to compensate for the decrease in ξ_0 by the increase of oscillation frequency F_0 at larger values of P_c.

In the case of the 500 mm resonator, the plot also shows the location of the first acoustical resonance F_1, at 170 Hz (for the 250 mm resonator, $F_1 = 340$ Hz falls outside the frequency range of the plot). Note that the oscillation frequency F_0 is close to F_1, particularly at large values of P_c, and therefore the pure inertance approximation for the vocal tract load does not hold.

In a second numerical experiment, we fitted Eqs. (11) and (15) to both 250 and 500 mm resonator data sets simultaneously, with k_t, B, and τ as parameters, to see how well they capture the vocal tract effect (Fig. 3). We set $L = 45$ mm (from Ruty *et al.*, 2007) and $\rho = 1.14$ kg/m^3 (from Chan and Titze, 2006). Also, the range of possible values for the transglottal coefficient k_t was limited to [1.0, 1.4] (Titze, 1988). The vocal tract inertance was computed as $I = \rho l / A$, where l is the length and A is the cross sectional area. For the 250 and 500 mm resonators, we have $I = 580.60$ kg/m^4 and $I = 1161.2$ kg/m^4, respectively. The computed optimal parameters were $k_t = 1.40$, $B = 783.95$ Pa s/m, $\tau = 1.37$ ms, for Eq. (11), and $k_t = 1.04$, $B = 1363.3$ Pa s/m, $\tau = 4.83 \times 10^{-7}$ ms, for Eq. (15).

In this experiment, the results for the 250 mm resonator are similar to those in Fig. 2: the extended Eq. (11) provides a good approximation, better than Eq. (15). Equation (15) does not predict the observed increase of P_{th} with P_c. The best approximation it can produce is by setting a very small

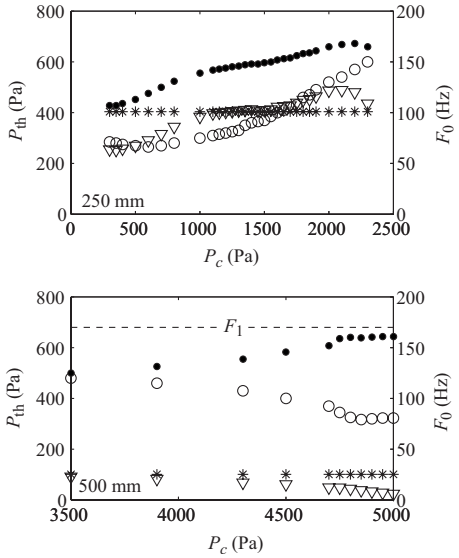


FIG. 3. Oscillation threshold pressure P_{th} and frequency F_0 vs internal pressure P_c for a 250 mm resonator (upper panel) and 500 mm resonator (lower panel). Circles: measured pressure values; triangles: theoretical pressure values given by Eq. (11), stars: theoretical pressure values given by Eq. (15); filled circles: measured oscillation frequency. The broken line in the lower panel indicates the first acoustical resonance (F_1) of the resonator.

value of τ, which results in an almost constant P_{th}. The results for the 500 mm resonator, on the other hand, are much poorer than those in Fig. 2: both Eqs. (11) and (15) produce values of threshold pressure much lower than the measured values.

VI. CONCLUSIONS

The above results show that the extended equation for phonation threshold pressure, given by Eq. (12), provides a better theoretical characterization than Eq. (13) previously derived by Titze (1988). In particular, the extended model contains the oscillation frequency as an explicit parameter, which was missing in the previous model, and therefore is able to capture phonation threshold versus frequency relations.

The results also show that modeling the vocal tract input pressure with the simple inertive load of Eq. (5) seems a crude approximation, which fails to model the effect of the vocal tract on the phonation threshold pressure with reasonable accuracy. However, two issues must be considered here: First, the inertive model is based on the assumption of an oscillation frequency much lower than the first vocal tract formant. This assumption does not hold well for the 500 mm resonator, for which the theoretical results are poor compared to the data. Second, a lumped impedance representation for the vocal tract may still be too simple to fit the experimental data, and a more sophisticated frequency-dependent model might be required.

Finally, note that the theoretical flow model relies on many simplifying, and thus questionable, assumptions by considering that the glottal flow is frictionless, quasi-steady, and incompressible. Of all these assumptions, the work of Ruty *et al.* (2007) tends to show that viscous effects are the most critical.

All of the above issues are currently being considered for extensions of this work.

ACKNOWLEDGMENTS

This work was supported by CAPES—Brazil under Program STIC-AmSud, MCT/CNPq, and by a Ph.D. grant from the French Ministry of Education and Research.

Chan, R. W., and Titze, I. R. (**2006**). "Dependence of phonation threshold pressure on vocal tract acoustics and vocal fold tissue mechanics," J. Acoust. Soc. Am. **119**, 2351–2362.

Lucero, J. C. (**1999**). "Theoretical study of the hysteresis phenomenon at vocal fold oscillation onset-offset," J. Acoust. Soc. Am. **105**, 423–431.

Lucero, J. C., and Koenig, L. L. (**2007**). "On the relation between the phonation threshold lung pressure and the oscillation frequency of the vocal folds," J. Acoust. Soc. Am. **121**, 3280–3283.

Ruty, N., Pelorson, X., Hirtum, A. V., Lopez-Arteaga, I., and Hirschberg, A. (**2007**). "An in vitro setup to test the relevance and the accuracy of low-order vocal folds models," J. Acoust. Soc. Am. **121**, 479–490.

Titze, I. R. (**1988**). "The physics of small-amplitude oscillation of the vocal folds," J. Acoust. Soc. Am. **83**, 1536–1552.

Titze, I. R. (**1992**). "Phonation threshold pressure: a missing link in glottal aerodynamics," J. Acoust. Soc. Am. **91**, 2926–2935.

Titze, I. R., Schmidt, S. S., and Titze, M. R. (**1995**). "Phonation threshold pressure in a physical model of the vocal fold mucosa," J. Acoust. Soc. Am. **97**, 3080–3084, part 1.

Vowel-to-vowel coarticulation in Japanese: The effect of consonant duration (L)

Anders Löfqvist[a)]
Haskins Laboratories, New Haven, Connecticut 06511 and Department of Logopedics, Phoniatrics and Audiology, Clinical Sciences, Lund University, Lund, Sweden

(Received 31 January 2008; revised 24 July 2008; accepted 25 July 2008)

This paper examines vowel-to-vowel lingual coarticulation in sequences of vowel-bilabial consonant-vowel, where the duration of the oral closure for the consonant is either long or short. Native speakers of Japanese served as subjects. The linguistic material consisted of Japanese word pairs that only differed in the duration of the labial consonant, which was either long or short. Recordings were made of lip and tongue movements using a magnetometer system. It was hypothesized that there would be greater vowel-to-vowel coarticulation in the context of a short consonant, since a long consonant would allow the tongue more time to move. The overall results do not show any strong support for this hypothesis, however. Subjects modulate the speed of the tongue movement between the two vowels, making it slower during the long than during the short consonant. © *2009 Acoustical Society of America.* [DOI: 10.1121/1.2973234]

PACS number(s): 43.70.Bk, 43.70.Aj [BHS] Pages: 636–639

I. INTRODUCTION

This paper examines vowel-to-vowel coarticulation in sequences of vowel-bilabial consonant-vowel, where the duration of the oral closure for the consonant is varied for linguistic purposes, using speakers of Japanese. In Japanese, the ratio of closure duration between long and short consonants is about 2:1 (Beckman, 1982; Han, 1994; Hirata and Whiton, 2005; Löfqvist, 2005, 2006, 2007). One might thus hypothesize that there is a greater influence from an upcoming vowel on the preceding one when the intervening labial consonant is short than when it is long. In the short consonant context, there is less time for the tongue to make the transition from the first to the second vowel. A similar argument can be made for an influence of the preceding vowel on the following one, although variations in tongue movement kinematics could neutralize the effect of such a reduced temporal window for the movement. That is, the results of Löfqvist (2006) show that the duration of the tongue movement between the two vowels is longer when the consonant is long. In addition, the average speed of the tongue movement between the vowels is slower in the long consonant context, although there is no systematic difference in the magnitude of the tongue movement as a function of consonant length.

It is well known that successive sounds in speech influence each other. Such influences can occur over quite large temporal intervals (e.g., Magen, 1997) and across segment, syllable, and word boundaries. The nature and extent of such coarticulatory influences seem to depend on several factors, including speaking rate, stress, and the articulatory requirements for different segments (e.g., Modaressi *et al.*, 2004). These requirements have sometimes been indexed in terms of coarticulatory resistance, originally proposed by Bladon and Al-Bamerni (1976), i.e., how resistant a sound is to coarticulatory influencies. For example, a fricative consonant made with the the front part of the tongue in contact with the hard palate has generally been found not to be very much affected by coarticulation (e.g., Fowler and Brancazio, 2000). In the present study, the coarticulation resistance of the consonant in the VCV sequence is not a serious issue, since it is a labial nasal, not produced with the tongue, and the tongue is not rigidly coupled to the jaw. Moreover, Fowler and Brancazio (2000) found little influence of coarticulation resistance on vowel-to-vowel coarticulation.

The present study thus examines vowel-to-vowel coarticulation across a labial nasal consonant with two different durations using articulatory movement tracking with a magnetometer in native speakers of Japanese. The specific hypothesis being addressed is that any such influences will be stronger in the context of a short than of a long labial consonant; this specific issue was not examined by Löfqvist (2006).

II. METHOD

A. Subjects

Five native speakers of Japanese, three male and two female, served as subjects. They reported no speech, language, or hearing problems. They were naive as to the purpose of the study. Before participating in the recording, they read and signed a consent form. (The experimental protocol was approved by the IRB at the Yale University School of Medicine.)

B. Linguistic material

The linguistic material consisted of Japanese words with a sequence of vowel-labial nasal-vowel. These words formed minimal pairs, where the only difference between the pairs was the duration of the labial consonant. The words were designed to require a substantial amount of tongue move-

[a)]Present address: Haskins Laboratories, 300 George St., New Haven, CT 06511. Electronic mail: lofquist@haskins.yale.edu

ment from the first to the second vowel. The following words were used: /kami, kammi/ ("god," "sweets"), /kamee, kammee/ ("participation," "impression") /kema, kemma/ ("Kema, place name," "polish"). The linguistic material was organized into randomized lists and presented to the subjects in Japanese writing, with the words occurring in a short frame sentence. Fifty repetitions of each word were recorded.

C. Movement recording

The movements of the lips, the tongue, and jaw were recorded using a three-transmitter magnetometer system (Perkell *et al.*, 1992). Receivers were placed on the vermilion border of the upper and lower lip, on three positions of the tongue, referred to as tip, blade, and body, and on the lower incisors at the gum line. Two additional receivers placed on the nose and the upper incisors were used for the correction of head movements. Two receivers attached to a plate were used to record the occlusal plane by having the subject bite on the plate during the recording. All data were subsequently corrected for head movements and rotated to bring the occlusal plane into coincidence with the x-axis. This rotation was performed to obtain a uniform coordinate system for all subjects (cf., Westbury, 1994).

The articulatory movement signals were sampled at 500 Hz after low-pass filtering at 200 Hz. The resolution for all signals was 12 bits. After voltage-to-distance conversion, the movement signals were low-pass filtered using a 25 point triangular window with a 3-dB cutoff at 14 Hz; this was done forwards and backwards to maintain phase. To obtain the instantaneous velocity of the tongue receivers, the first derivative of the position signals was calculated using a three-point central difference algorithm. For each tongue receiver, its speed $[v=\sqrt{(\dot{x}^2+\dot{y}^2)}]$ was also calculated. The velocity and speed signals were smoothed using the same triangular window. The acoustic signal was pre-emphasized, low-pass filtered at 4.5 kHz and sampled at 10 kHz.

The horizontal and vertical positions of the tongue body during the first and second vowels were defined algorithmically in the tongue body speed signal as minima during the first and second vowels, see Fig. 1(a); Fig. 1(b) shows the movements of all receivers during this interval. They correspond to the onset and offset of the tongue movement between the two vowels. We should note that at these points in time, the horizontal and vertical velocities of the tongue are usually not zero. This is partly because the kinematic signals are expressed in a maxilla-based coordinate system, thus the recorded tongue body movement also includes the contribution of the jaw. Such a coordinate system is appropriate when we are interested in the tongue as the end effector

T-tests were used to assess differences between the long and short consonants for each subject. Given the large number of comparisons, an α-level of 0.001 was adopted based on dividing the standard alpha level of 0.05 by the number of comparisons.

III. RESULTS

The duration of the oral closure for the labial consonant showed a robust difference with no overlap between the values for the short and long ones. The range of durations for the short consonants was 54–95 ms, while that for the long ones was 119–165 ms (Löfqvist, 2006).

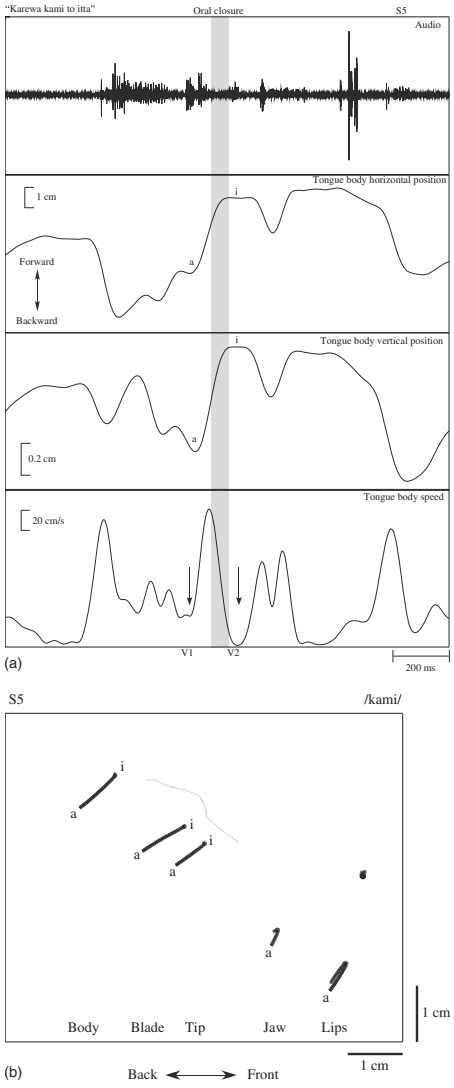


FIG. 1. (a) Audio and tongue body signals for the word /kami/ with arrows showing the points used in the speed signal for defining the onset and offset of the tongue movement between the two vowels. The baseline in the bottom panel with the speed signal represents zero speed. (b) Articulatory movements from the first to the second vowel in "kami." The gray line represents a tracing of the hard palate.

A. Tongue body position during the first vowel

Figure 2 presents the tongue body positions during the first vowel. According to the hypothesis, there should be a difference in the tongue positions between the long and short vowel contexts. In particular, it was expected that the tongue would be in a higher and more front position in the words with a short consonant /kami, kamee/ than in the words with a long consonant /kammi, kammee/ due to the influence of the front second vowel. In the words /kema, kemma/, the hypothesis predicts a lower and more retracted tongue position for the first vowel in /kema/ than in /kemma/ due to the influence of the second, back, vowel. An inspection of Fig. 2 suggests that the prediction for the words /kami, kammi, kame, kammee/ is partly supported: The unfilled circles and

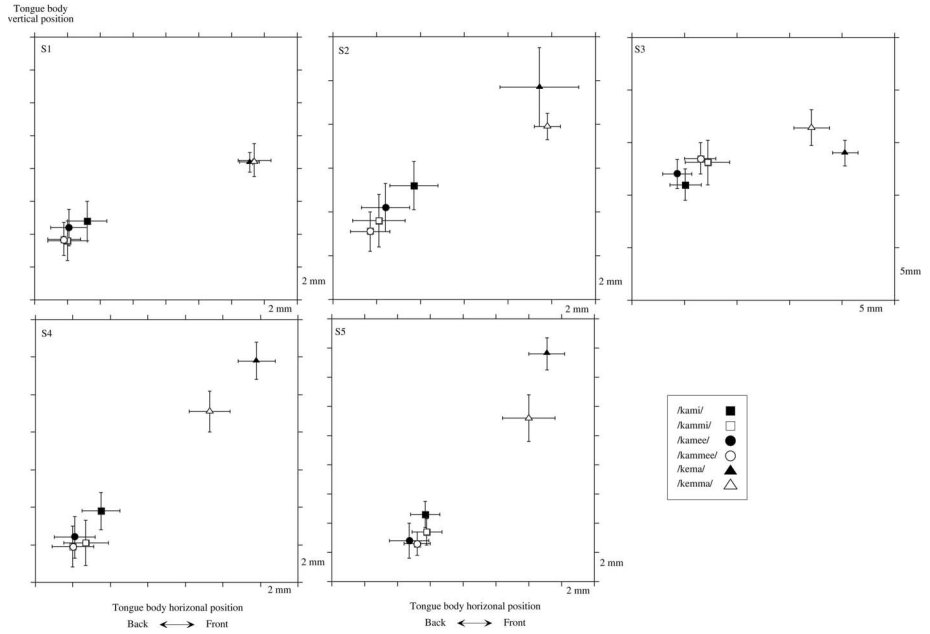


FIG. 2. Tongue body positions during the first vowel (mean and standard deviation).

squares are in a lower and more posterior position than their filled counterparts for four of the subjects 1, 2, 4, and 5. That is, the vowels in the words with a long consonant are less influenced by the second vowel than those in the words with a short consonant. However, for subject 3, the opposite is the case. For the words /kema, kemma/, the prediction is not supported for most subjects, since the filled triangles (/kema/) tend to be in a higher and more anterior position than their unfilled counterparts for subjects 2, 3, 4, and 5. The statistical analysis for /kami, kammi/ showed a significant difference in horizontal tongue position for subjects 1, 2, 3, and 4 ($t=6.4$, 10.89, −7.97, 4.32, with $p<0.001$ in all cases). Note however, that for subject 3, the results are opposite to the predicted ones. Subject 5 showed no statistical difference ($t=-0.62$ ns). For the vertical tongue position, all subjects showed a significant difference ($t=4.81$, 7.3, −5.81, 7.73, 7.28, $p<0.001$), again with the results of subject 3 opposite to the predicted ones.

In the words /kamee, kammee/, only subjects 2 and 3 showed a difference in the horizontal position ($t=4.81$ and −8.71, $p<0.001$), but not subjects 1, 4, and 5 ($t=1.17$, 0.26, and −1.85 ns). Again, the results for subject 3 are opposite to the predicted ones. For the vertical position, only subjects 2 and 3 showed a significant difference ($t=5.94$, and −5.06, $p<0.001$) but not subjects 1, 4, or 5 ($t=3.22$, 2.55, 0.94 ns). Finally, for the words /kema, kemma/, the horizontal position was different for subjects 3, 4, 5 ($t=10.13$, 8.62, and 4.49, $p<0.001$), but not for subjects 1 and 2 ($t=-1.17$, and −2.53 ns). For the vertical position, subjects 2, 3, 4, and 5 showed a difference ($t=6.68$, −8.46, 12.66, 15.87, $p<0.001$), but not subject 1 ($t=-0.34$ ns). Note, however, that this difference was opposite to the predicted one.

The results for the first vowel thus show that for the low back vowel /a/, three of the subjects showed some qualified support for the hypothesis that there would be less vowel-to-vowel coarticulation across a long consonant. However, for the front vowel /e/, the opposite pattern was found in four subjects. Overall, the support for the hypothesis is very weak.

B. Tongue body position during the second vowel

Figure 3 shows the tongue body position during the second vowel. Here, the prediction is that the front vowels /i, e/ would have a lower and more retracted tongue position in the words /kami/ and /kamee/ due to the influence of the low back first vowel. For the low back vowel /a/ in /kema, kemma/ the prediction is that it will have a higher and more advanced tongue position due to the influence of the first vowel /e/.

The overall results suggest that there is no reliable difference in the tongue position during the second vowel as a function of consonant length. Subject 2 showed the horizontal positions in the words /kamee, kammee/ to be significantly different ($t=-3.77$, $p<0.001$). For subject 3, there was a significant difference in the horizontal position for the words /kame, kammee/ ($t=4.46$, $p<0.001$), and also for the horizontal position for the words /kema, kemma/ ($t=4.58$, $p0.001$). Subject 4 had differences in the horizontal position for the words /kame, kamme/ ($t=-3.48$, $p=0.001$), and the vertical position for the words kema, kemma/ ($t=5.15$, $p<0.001$). Finally, subject 5 only showed a difference in the horizontal position for the words /kema, kemma/ ($t=8.89$, $p<0.001$). Of these statistically reliable differences, only three of them were in the predicted direction, subjects 2, 4, and 5, but in no case was the difference found for both the horizontal and vertical positions. For subject 3, the two reliably different results were opposite to the predicted ones. Overall, the results for the second vowel show no consistent differences in tongue position as a function of consonant length.

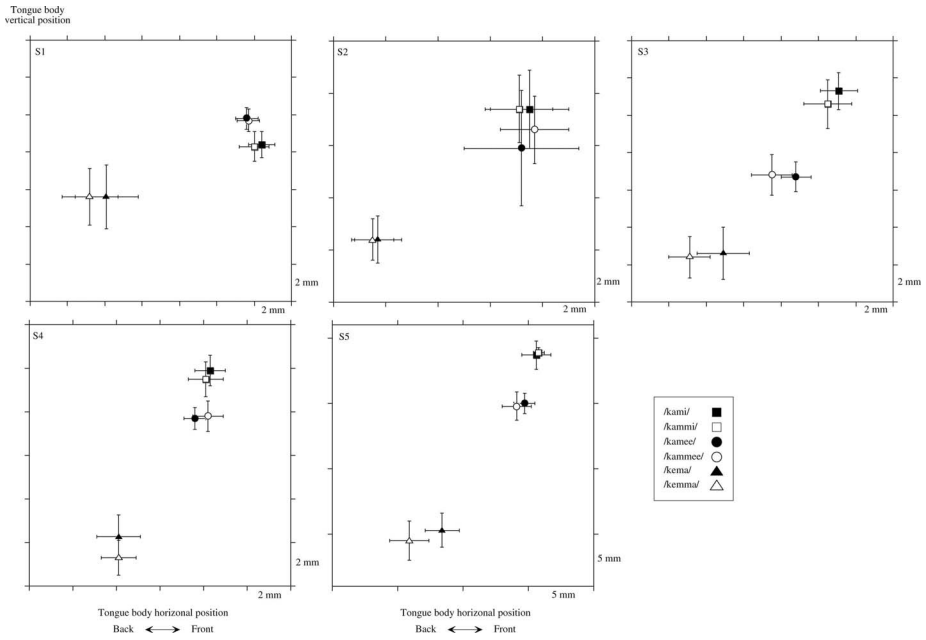


FIG. 3. Tongue body positions during the second vowel (mean and standard deviation).

IV. DISCUSSION

This study examined the influence of consonant duration on vowel-to-vowel coarticulation in Japanese. It was hypothesized that a short intervocalic labial consonant would allow more coarticulation than a long consonant, in particular since the duration of a long consonant is about twice as long as that of a short consonant. The overall results do not show any strong support for this hypothesis, however. Three of the subjects showed the expected influence of a following high vowel on a preceding low back vowel, but one subject showed the opposite results. There were no effects on the second vowel. Thus, there was some limited evidence for more anticipatory influences than carryover effects.

The most likely reason for the small effects is that Japanese speakers adjust the speed of the tongue movement to maintain a similar, but not identical, coordination of lip and tongue movements for long and short consonants (Löfqvist, 2006). That is, the onset of the tongue movement occurs before the oral closure for the consonant, and its offset occurs after the oral release. As a consequence, the tongue positions for the vowels in the context of the long and short consonants are very similar. A further consequence is that the duration of the tongue movement between the two vowels is longer when the intervening consonant is long than when it is short. The magnitude of the movement path between the two vowels did not vary systematically with consonant duration. Not modulating the speed of the tongue movement would result in the tongue reaching the intended target for the second vowel well before the release of the long consonant and it might then have to stop moving. Such a movement pattern would involve successive accelerations and decelerations of the tongue that would involve a higher cost of effort. Thus, speakers avoid excessive accelerations and decelerations of the tongue by keeping it moving.

ACKNOWLEDGMENTS

The author is grateful to Mariko Yanagawa for help with the Japanese material and running the experiments. This work was supported by the National Institute on Deafness and Other Communication Disorders, National Institutes of Health Grant No. DC-00865.

Beckman, M. (**1982**). "Segment duration and the mora in Japanese," Phonetica **39**, 113–135.

Bladon, A., and Al-Bamerni, A. (**1976**). "Coarticulation resistance in English /l/," J. Phonetics **4**, 137–150.

Fowler, C. A., and Brancazio, L. (**2000**). "Coarticulation resistance of American English consonants and its effects on transconsonantal vowel-to-vowel coarticulation," Lang Speech **43**, 1–41.

Han, M. (**1994**). "Acoustic manifestations of mora timing in Japanese," J. Acoust. Soc. Am. **96**, 73–82.

Hirata, Y., and Whiton, J. (**2005**). "Effects of speaking rate on the single/geminate stop distinction in Japanese," J. Acoust. Soc. Am. **118**, 1647–1660.

Löfqvist, A. (**2005**). "Lip kinematics in long and short stop and fricative consonants," J. Acoust. Soc. Am. **117**, 858–878.

Löfqvist, A. (**2006**). "Interarticulator programming: Effects of closure duration on lip and tongue coordination in Japanese," J. Acoust. Soc. Am. **120**, 2872–2883.

Löfqvist, A. (**2007**). "Tongue movement kinematics in long and short Japanese consonants," J. Acoust. Soc. Am. **122**, 512–518.

Magen, H. (**1997**). "The extent of vowel-to-vowel coarticulation in English," J. Phonetics **25**, 187–205.

Modaressi, G., Sussman, H., Lindblom, B., and Burlingame, E. (**2004**). "An acoustic analysis of the bidirectionality of coarticulation in VCV utterances," J. Phonetics **32**, 291–312.

Perkell, J., Cohen, M., Svirsky, M., Matthies, M., Garabieta, I., and Jackson, M. (**1992**). "Electromagnetic midsagittal articulometer (EMMA) systems for transducing speech articulatory movements," J. Acoust. Soc. Am. **92**, 3078–3096.

Westbury, J. (**1994**). "On coordinate systems and the representation of articulatory movements," J. Acoust. Soc. Am. **95**, 2271–2273.

The direct simulation of acoustics on Earth, Mars, and Titan

Amanda D. Hanford[a)]
Applied Research Laboratory, The Pennsylvania State University, University Park, Pennsylvania 16804

Lyle N. Long[b)]
Department of Aerospace Engineering, The Pennsylvania State University, University Park, Pennsylvania 16802

(Received 9 April 2008; revised 5 November 2008; accepted 19 November 2008)

With the recent success of the Huygens lander on Titan, a moon of Saturn, there has been renewed interest in further exploring the acoustic environments of the other planets in the solar system. The direct simulation Monte Carlo (DSMC) method is used here for modeling sound propagation in the atmospheres of Earth, Mars, and Titan at a variety of altitudes above the surface. DSMC is a particle method that describes gas dynamics through direct physical modeling of particle motions and collisions. The validity of DSMC for the entire range of Knudsen numbers (Kn), where Kn is defined as the mean free path divided by the wavelength, allows for the exploration of sound propagation in planetary environments for all values of Kn. DSMC results at a variety of altitudes on Earth, Mars, and Titan including the details of nonlinearity, absorption, dispersion, and molecular relaxation in gas mixtures are given for a wide range of Kn showing agreement with various continuum theories at low Kn and deviation from continuum theory at high Kn. Despite large computation time and memory requirements, DSMC is the method best suited to study high altitude effects or where continuum theory is not valid.
 [DOI: 10.1121/1.3050279]

PACS number(s): 43.28.Bj, 43.28.Js, 43.35.Ae [VEO] Pages: 640–650

I. INTRODUCTION

The Cassini-Huygens spacecraft is the largest interplanetary, international spacecraft ever built. Its mission is to perform an in-depth study of the diverse phenomena of the Saturn system, including the planet itself, its rings, and its largest moon, Titan. The orbiter, Cassini, boasts cutting-edge instruments capable of collecting sophisticated data for 250 scientists in 17 countries. It was also designed to piggyback the Huygens probe all the way to Titan's orbit and to be able to eject the probe properly. The orbiter also was designed to forward to Earth the information it receives from Huygens during the probe's descent through Titan's atmosphere.[1] The Huygens probe descended to the surface of Titan in early 2005, capturing data on the structure, composition, and climate of Titan's atmosphere. The surface science package (SSP) on board the probe included passive acoustic sensors to record ambient sounds in hopes of capturing thunder in the moon's atmosphere as well as active sensors for measuring surface topography, average molecular weight, altitude, wind speed, and surface acoustic impedance.[2,3] Recently, physical properties of the topological and geological makeup of the Huygens' landing site have been derived from the sonar measurements made during the decent[4,5] and speed of sound measurements were used to calculate the methane concentration as a function of altitude.[6]

Other planetary missions have also included acoustic sensors in hopes that they would transmit sound samples back to Earth. In the 1980s, the Russian-built Venera mission lasted 120 min on Venus and survived long enough to record a possible thunder event.[7] But it was not again until 1999 when a low-cost microphone was constructed for the ill-fated Mars Polar Lander mission[8] that an acoustic sensor was created for planetary exploration. While the Mars Polar Lander lost contact with Earth shortly after its descent into the Martian atmosphere and never recovered, there is hope that a future mission will include more acoustic experiments. With the recent success of the Huygens probe, there has been renewed interest in further exploring the acoustic environments of the other planets in the solar system.

The planetary environments of Earth, Mars, and Titan provide an opportunity for exploring and measuring sound propagation in a wide range of atmospheric environments and conditions, which is very sensitive to changes in pressure, temperature, and molecular composition. On these planets (while Titan is a moon, it will be referred to as a planet), atmospheric pressures range from 0.007 atm on Mars to 1.5 atm on Titan and temperatures range from 90 K on Titan to 300 K on Earth. As on Earth, the majority of Titan's atmosphere is made up of nitrogen, while Mars' atmosphere is made up mostly of carbon dioxide. The atmospheric composition on the surface of the three planets is summarized in Table I.[9–11]

Despite the fact that several space missions were implemented with acoustic sensors, little work has been done to quantify sound propagation in planetary atmospheres based on nonempirical models. Even less has been done to study sound waves of finite amplitude where nonlinear effects could be present. Notable contributions include a study of absorption and dispersion near the Martian surface based on semiempirical methods,[12] an overview of the acoustic envi-

a)Electronic mail: ald227@psu.edu
b)Electronic mail: lnl@psu.edu

TABLE I. Atmospheric conditions at the surface on Earth, Mars, and Titan.

Planet	Earth	Mars	Titan
Temperature (K)	300	200	95
Pressure (atm)	1.0	0.007	1.5
Molecular composition	N_2 (70%), O_2 (20%)	CO_2 (95%)	N_2 (95%), CH_4 (5%)

ronment on Mars,[13] and a comparison of atmospheric linear acoustics on Mars, Titan, Venus, and Earth using an effective wavenumber approach.[14]

This paper presents a simulation technique to model sound in extraterrestrial environments that combines all relevant physics to study sound, including nonlinearity, without making assumptions about the fluid dynamics processes. This technique is the direct simulation Monte Carlo (DSMC) which is the stochastic, particle-based method developed by Bird.[15] Particle simulations of sound propagation investigating the effects of amplitude, relaxation, and altitude on Earth, Mars, and Titan will now be presented.

II. DIRECT SIMULATION MONTE CARLO

The DSMC method is a simulation tool that describes the dynamics of a gas through direct physical modeling of particle motions and collisions. DSMC is based on the kinetic theory of gas dynamics modeled on the Boltzmann equation, where representative particles are followed as they move and collide with other particles. The movement of particles is determined by their velocities. While the collisions between particles are determined statistically, they are required to satisfy mass, momentum, and energy conservation. Introductory[16,17] and detailed[15] descriptions of DSMC, as well as formal derivations,[18] can be found in the literature. Due to the particle nature of the method, DSMC offers considerable flexibility with regard to the type of system available for modeling: rarefied gas dynamics,[15,19–21] hypersonic flows,[15,22,23] and acoustics.[16,24,25] DSMC's origins are based on the Boltzmann equation, but the applications of and the extensions to DSMC method have now gone beyond the range of validity of this equation by simulating chemical reactions,[26–28] detonations,[29,30] and volcanic plumes and upper atmospheric winds on Jupiter's moon Io.[31,32]

DSMC is a particle method that describes the state of the gas at the microscopic level, and is valid beyond the continuum assumption. The Knudsen number (Kn) is defined as the mean free path, λ_m, divided by a characteristic length of the system and is a measure of the nonequilibrium effects in the gas. In the case of one-dimensional acoustic wave propagation, the characteristic length is the acoustic wavelength so the Knudsen number is directly proportional to the frequency of oscillation. The mean free path can be calculated using the equation[15]

$$\lambda_m = \frac{RT}{\sqrt{2}\pi d^2 p_0 \mathcal{N}}, \tag{1}$$

where R is the universal gas constant, T is the temperature, d is the molecular diameter, p_0 is the ambient pressure, and $\mathcal{N}$ is Avogadro's number. The Knudsen number is also used to distinguish the regimes where different governing equations of fluid dynamics are applicable. The Navier–Stokes equations are the mathematical model for continuum methods which assumes deviations from thermal equilibrium are small.[15,33] The Navier–Stokes equations are valid for Kn <0.05 and reduce to the Euler equations as Kn approaches zero. The Boltzmann equation is the mathematical model for particle methods and is valid for all Kn. Therefore, particle methods are necessary for, but not limited to, problems where the Knudsen number is greater than about 0.05.

Despite the fact that DSMC is valid for all Kn, DSMC is most efficient for high Kn flows and has, in fact, become the *de facto* tool for high Kn situations. The Knudsen number is large for sound propagation in dilute gases (e.g., high altitude conditions) or at high frequencies, requiring a particle method solution. Experimental work[34] and DSMC simulations[16,25] have shown that sound absorption depends heavily on Kn which deviates significantly from traditional continuum theory at high Kn. Traditionally, DSMC has primarily been used in regimes where continuum methods fail. However, DSMC has many advantages even for low Kn situations. Without modification, DSMC is capable of simulating all physical properties of interest at the molecular level for sound propagation: absorption, dispersion, nonlinearity, and molecular relaxation.

The current DSMC program that was used for simulations in this paper contains several types of energy models to treat molecules in gas mixtures with internal energy. Internal energy is represented by rotational and vibrational modes (electronic energy is ignored) and has been programmed to simulate either classical or quantum behavior. Each case uses a phenomenological approach developed by Borgnakke and Larsen[35] which treats only a fraction of intermolecular collisions as inelastic. This fraction is the reciprocal of the relaxation collision number Z and is specified independently for rotational Z_{rot} and vibrational Z_{vib}, degrees of freedom and is dependent on the molecular species of the colliding pair. If a collision is regarded as inelastic, the total energy of the particles is reassigned between the translational and internal modes by sampling from known equilibrium distributions during the collision but the total energy is conserved.[15,35]

In the case of molecules having vibrational internal degrees of freedom, postcollision energy is assigned through either a classical procedure that assigns a continuously distributed vibrational energy to each molecule or through a quantum approach that assigns a discrete vibrational level to each molecule. The treatment of the vibrational energy modes is treated independently from the rotational modes. Each vibrational mode of a molecule can be modeled as a separate molecular species to model the transitions between levels that occur as a result of intermolecular collisions. While this Borgnakke–Larson phenomenological method is quite unrealistic from a physical point of view, it satisfies detailed balance, produces statistically correct macroscopic behavior and is computationally efficient.[15]

In the case of molecules having vibrational internal degrees of freedom, postcollision energy is assigned through either a classical procedure that assigns a continuously distributed vibrational energy to each molecule or through a

quantum approach that assigns a discrete vibrational level to each molecule. The vibrational energy exchange is treated independently from the energy exchange in rotational modes. With DSMC each discrete vibrational energy level of a molecule can be modeled as a separate molecular species. This phenomenological model satisfies detailed balance and produces statistically accurate macroscopic behavior in addition to being very computationally efficient.[15] Details about the DSMC implementation of these internal energy models and their effect on acoustics have been described in the literature for a wide range of temperatures.[24]

The flexibility of the DSMC algorithm allows for modeling of sound in specific gas mixtures including models for Earth, Mars, and Titan.[36–38] Little modification to the method is needed to change the molecular and ambient atmospheric properties in order to simulate sound on the different planets. This feature of DSMC makes it beneficial for use in planetary acoustics where atmospheric conditions are dependent on planet, time of year, altitude, etc. In addition, the Kn is high in upper atmospheric conditions, thus requiring a particle method solution.

III. SIMULATION APPROACH

DSMC simulations were performed to simulate the sound propagation on Earth, Mars, and Titan. On each planet, acoustic waves were simulated in a one-dimensional simulation domain by creating a pistonlike boundary condition at one end of the domain. The piston was simulated as a rigid wall where particle collisions with the piston face would result in sinusoidally oscillating velocity components. The far end of the simulation domain was terminated with a specular wall boundary condition. In order to minimize computation costs, the domain length was limited to a few wavelengths.

Molecules were simulated as hard sphere particles. The computational cell size Δx was required to be less than a half of a mean free path. Alexander *et al.*[39] showed that the transport coefficients deviate from kinetic theory values proportional to the square of the cell size. Cell sizes on the order of a mean free path can produce an error on the order of 10%. Therefore, cell sizes were restricted to $\Delta x \geq 0.5\lambda_m$ so the error is less than 2.5%. The variation in Kn was obtained by varying the acoustic wavelength from 100 to 2500 cells.

The time step was taken to be at least an order of magnitude smaller than the mean collision time in each case and is on the order of picoseconds. Care was taken to ensure that the time step remained much smaller than the acoustic period of oscillation. Each case was initialized in thermal equilibrium. Details on the implementations for each planet are given below. Atmospheric properties for all planets were derived from public-access general circulation models.[9,11,40–42]

A parallel, object-oriented DSMC solver was developed for this problem. The code was written in C++, Message Passing Interface (MPI) for interprocessor communication, and was run on parallel computers. The object-oriented approach allows the DSMC algorithm to be divided into physical objects that are individually maintained. Cell and particle classes were created to govern fundamental components of the algorithm. With this object-oriented technique it was possible to develop a C++ code that was easy to read, maintain, and modify. Despite excellent parallel efficiency, CPU time and memory requirements were quite large, taking approximately 6 h on 32 processors for each run.

The memory requirements of DSMC imply that domain size and the number of time steps are limited and are machine specific. By limiting domain length, the largest acoustic wavelength that is possible to simulate with the current version of code is approximately 2500 cells with current computational resources. With the cell size restriction of $\Delta x \leqslant 0.5\lambda_m$, DSMC is currently limited to situations where $\mathrm{Kn} \geqslant 0.0008$.

IV. RESULTS AND DISCUSSION

The physical properties that govern the absorption of sound include classical losses from the transfer of acoustic energy into heat and relaxation losses associated with the redistribution of internal energy of molecules. The relaxation losses are associated with the relaxation of the molecule's rotational energy and the losses from the relaxation of the molecule's vibrational energy.

At high Kn where frequencies are well above the vibrational relaxation frequency, vibration does not contribute to the specific heat. In addition, the frequency ranges for rotational and translational relaxations overlap each other and coincide with the breakdown of the continuum assumption. The rotational relaxation frequencies for $Z_{\mathrm{rot}} < 5$ is beyond the continuum breakdown.

The theoretical predictions for the absorption of sound that include relaxation phenomenon can accurately describe the absorption and dispersion in gases at very small Kn.[43,44] However, when the continuum assumption breaks down, continuum theory can no longer describe the transport phenomenon correctly, and comparisons between experimental and computational results show significant deviations at high Kn due to slow or incomplete translational relaxation.[16,25,34] Several molecular-kinetics adjustments have been made to the theory to account for the discrepancy at high Kn with varying degrees of success. Sutherland and Bass[45] used an empirical adjustment to account for the high Kn behavior while Buckner and Ferziger[46] and Sirovich and Thurber[47] used approximations to the Boltzmann equation to describe deviation from the Navier–Stokes prediction for a monatomic gas. Both experimental[48] and DSMC simulation results[24,25] do approach the free-molecular flow limit for extreme Kn values.[46] Nevertheless, large discrepancies between theory and experiment still exist, most notably in the mid-Kn range where translational relaxation effects are most important.

The DSMC approach described here can help investigate these discrepancies in describing acoustic phenomenon on Earth, Mars, and Titan for all Kn.

A. Earth

Earth's atmosphere close to the surface contains roughly 78% nitrogen, 20.95% oxygen, 0.93% argon, 0.038% carbon dioxide, and trace amounts of other gases (most notably, wa-

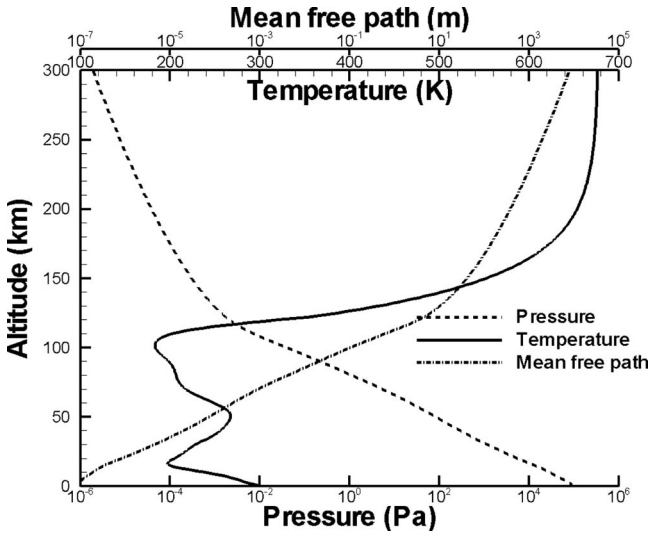


FIG. 1. Temperature (solid line), pressure (dashed line), and mean free path (dash-dotted line) profiles as a function of altitude above Earth's surface (Refs. 9, 40, and 41).

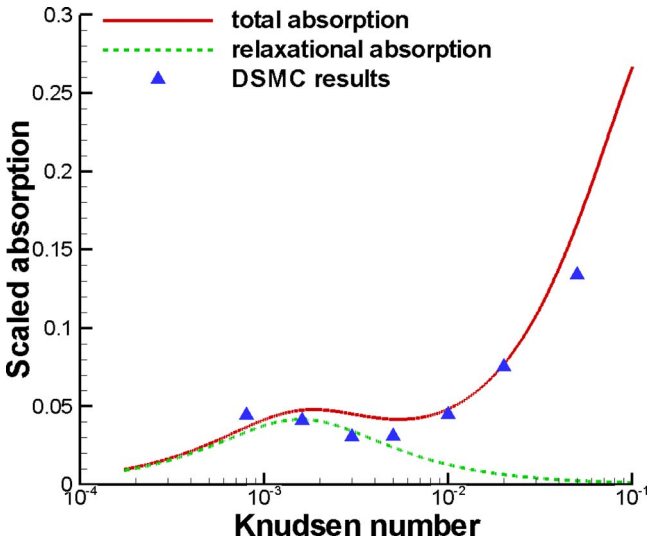


FIG. 2. (Color online) Scaled absorption for dry air on Earth with $Z_{vib}=200$ at standard atmospheric conditions ($p=101\,000$ Pa, $T=273$ K, and $\lambda_m=6.3\times10^{-8}$ m). DSMC simulations (points) are plotted with continuum theory for the vibrational relaxation (dashed line) and total absorption (solid line).

ter vapor). The average atmospheric pressure at sea level is about 101 000 Pa with average temperatures near 300 K.

However, pressure, temperature, and mean free path are also functions of altitude above the surface. Figure 1 gives the temperature, pressure, and mean free path profiles as a function of altitude up to 300 km.

Within the homosphere, at an altitude below 100 km, the molecular composition of the atmosphere is more or less uniform and makes up 99.999 97% of the atmosphere by mass.[49] Above the homosphere, which is often referred to as "outer space," molecular diffusion begins to dominate over turbulent mixing that dominates within the homosphere. Here, Earth's atmosphere has a composition which varies with altitude. This transition into the heterosphere marks the inflection point seen in the temperature profile of Fig. 1 at about 100 km.

It is widely known that vibrational relaxation losses play an important role in sound propagation at audible frequencies in Earth's atmosphere.[43] The introduction of water vapor to a dry air mixture has been shown to increase the vibration relaxation frequencies of oxygen and nitrogen. This happens through complicated vibrational energy transfer pathways during intermolecular collisions. However, there is no simple way to determine a relationship between various vibrational energy transfer rates and relaxation frequencies. Therefore, empirical formulations for the relaxation frequencies as a function of humidity have been developed based on experimental measurements in air.[50]

In an effort to model relaxation effects, DSMC simulations were performed in Earth-like conditions. The absorption of sound due to a simple relaxation process given by a single relaxation time is presented in air at atmospheric conditions ($p=101\,000$ Pa, $T=273$ K, and $\lambda_m=6.3\times10^{-8}$ m). Because the addition of water vapor in the atmosphere decreases the relaxation time, two sets of simulations were run to represent dry and humid air on Earth. The relaxation time is proportional to the vibrational collision number Z_{vib} which is given to be 200 for the "dry air" case and $Z_{vib}=40$ for the "humid air" case.

The rotational relaxation numbers for nitrogen and oxygen should be very close at atmospheric temperatures. Based on experimental values, the rotational relaxation number Z_{rot} as a function of temperature T, in K, for the species of interest is given by the equation[12]

$$Z_{rot}=61.1\exp(-16.8/T^{1/3}). \quad (2)$$

While the choice of vibrational collision numbers is not necessarily physically realistic, the choice of collision numbers was based on the size of the system. In this case, relaxation effects could be investigated within the computational restrictions that DSMC imposes on the domain size.

The scaled absorption coefficient, α/k_0, where k_0 is the acoustic wavenumber, is plotted as a function of Kn for the dry air case in Fig. 2 and the humid air case in Fig. 3. Results are shown with theoretical predictions for vibrational relaxation[51] and the combination of vibrational relaxation, rotational and classical losses[52] given the collision numbers used. The frequency investigated ranges 4–265 MHz which spans the Knudsen number range of $0.0008\leqslant \mathrm{Kn}\leqslant 0.05$. This range is within the continuum assumption; therefore it is expected that DSMC results would follow theory. The relaxation peak is evident in the dry air case, but in the humid air case, classical thermal-viscous losses become more important so the relaxation peak is less defined. The relaxation peak does increase in Knudsen number when decreasing the collision number (increasing humidity), as shown in Figs. 2 and 3.

In addition, dry air earth conditions have also been modeled for frequencies above the vibrational relaxation frequencies where classical losses dominate. A frequency of 230 MHz corresponding to $\mathrm{Kn}=0.02$ was simulated in a gas mixture of nitrogen, oxygen, and argon at atmospheric con-

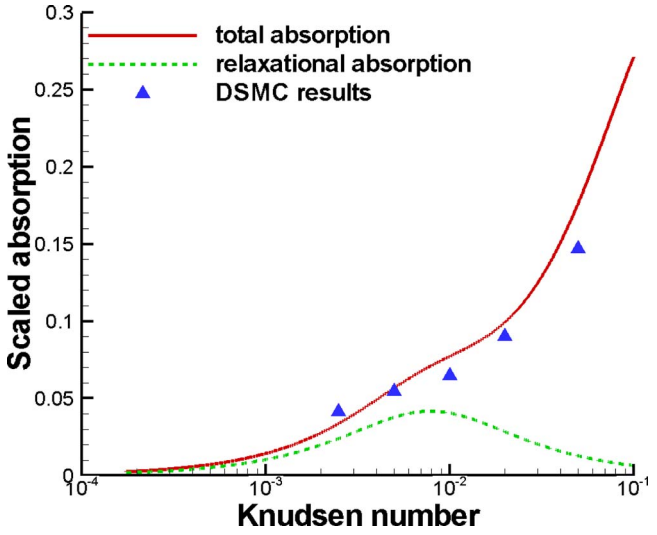


FIG. 3. (Color online) Scaled absorption for humid air on Earth with $Z_{vib}=40$ at standard atmospheric conditions ($p=101\,000$ Pa, $T=273$ K, and $\lambda_m=6.3\times10^{-8}$ m). DSMC simulations (points) are plotted with continuum theory for the vibrational relaxation (dashed line) and total absorption (solid line)

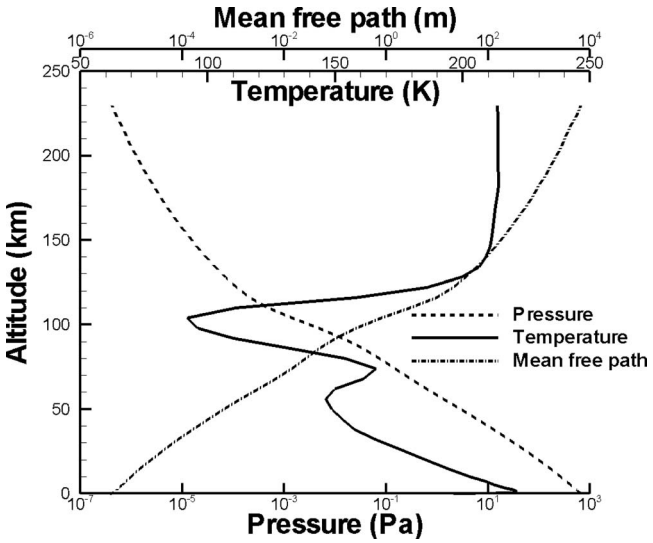


FIG. 5. Temperature (solid line), pressure (dashed line), and mean free path (dash-dotted line) profiles as a function of altitude above the Mars surface (Refs. 10 and 42).

ditions ($p=101\,000$ Pa, $T=273$ K, and $\lambda_m=6.3\times10^{-8}$ m). DSMC results for the acoustic pressure at a point in time are compared to the predicted amplitude dependence determined from the Navier–Stokes derived absorption coefficient in Fig. 4. Comparisons between the DSMC results and theory are quite good as expected.

B. Mars

The molecular composition of the Martian atmosphere is well established by previous missions and ground based observations. The most prominent constituent near the surface is carbon dioxide (95.3%). There are also minor amounts of nitrogen (2.7%) and argon (1.6%), with trace amounts of oxygen and water vapor.[53]

Mars' thin atmosphere has a surface pressure that varies greatly throughout the planet, which supports large winds

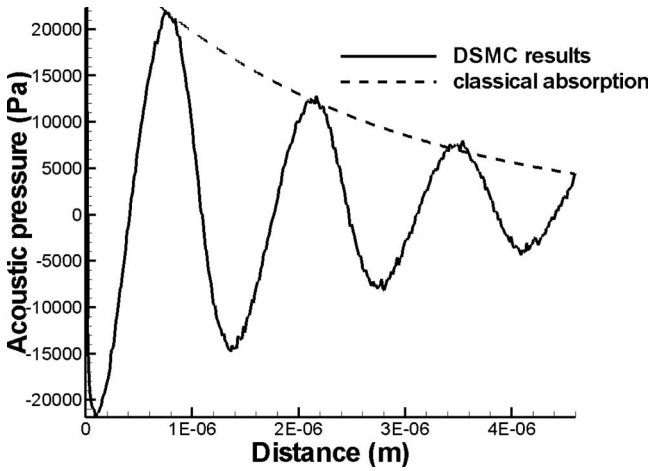


FIG. 4. DSMC results (solid line) for the acoustic pressure on Earth for Kn=0.02 at standard atmospheric conditions ($p=101\,000$ Pa, $T=273$ K, and $\lambda_m=6.3\times10^{-8}$ m) compared to predicted amplitude dependence determined from the Navier–Stokes derived absorption coefficient (dashed line).

that create planetwide dust storms.[54] The global average pressure at the surface ranges 500–700 Pa which is only about 0.7% of that on Earth at sea level. However, the average surface pressure varies significantly based on topography. For instance, the pressure at the Phoenix landing site is about 850 Pa whereas on top of Mars' largest mountain, Olympus Mons, the pressure is around 30 Pa.[55] The atmosphere is dusty with measurements from the Mars Exploration Rovers indicating a significant amount of dust particles roughly 1.5 μm in diameter.[56] Although the average Martian surface atmospheric pressure is less than 1% of that seen on Earth, the much lower gravity on Mars relative to Earth allowing the mean free path on Mars to increase at a slower rate than on Earth.

Also, the temperature on Mars is cooler than Earth, ranging from 160 to 300 K which can result in frozen carbon dioxide on the polar caps during the winter months. Given the large temperature and pressure changes throughout the planet, the mean molecular mass will also vary spatially and temporally.

Data are available for atmospheric conditions extending over 200 km above the Martian surface[10,40,41] with multiple datasets available given the time of year, time of day, geographical location, dust, and solar flux. As an example, temperature, pressure, and mean free path profiles are plotted as a function of altitude in Fig. 5 assuming average dust and solar flux during the first month of the Martian year, at 2:00 Local True Solar Time, and at the equator. DSMC simulation conditions were drawn from this model for a variety of altitudes to simulate "average" Martian conditions.[10,40,41]

Because of the low temperatures on Mars, vibrational energy is not active for frequencies above the relaxation frequency and does not contribute to the specific heat. The relaxation frequency for carbon dioxide, the primary atmospheric constituent, is below 100 Hz under Martian conditions.[12] However, DSMC simulations were performed

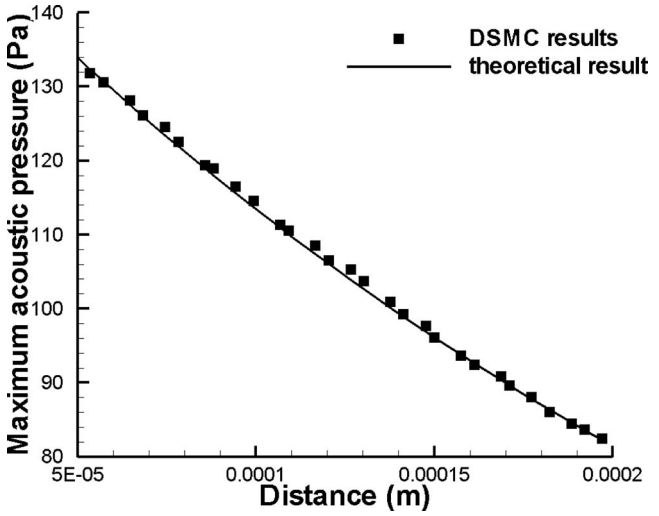


FIG. 6. DSMC results (points) for the acoustic pressure amplitude as a function of distance on Mars for Kn=0.02 at $p=700$ Pa, $T=200$ K, and $\lambda_m=8\times10^{-8}$ m compared to theoretical results (Ref. 12) (solid line).

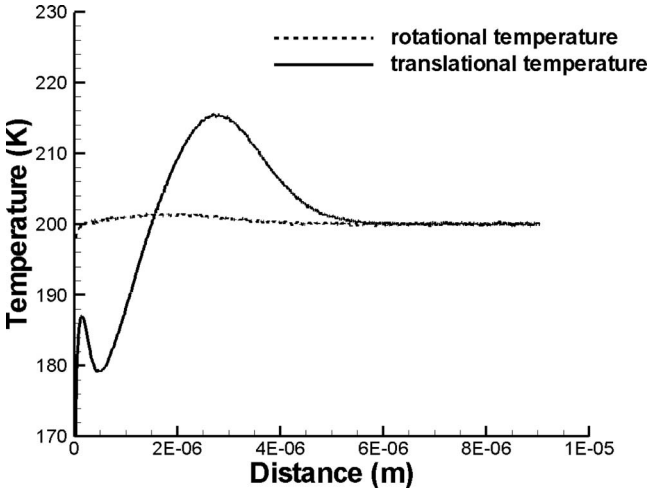


FIG. 7. Nonequilibrium effects on Mars showing rotational (dashed line) and translational (solid line) temperatures for Kn=2 at $p=700$ Pa, $T=200$ K, and $\lambda_m=6\times10^{-8}$ m.

for frequencies far above the relaxation frequency, so only the rotational internal energy of the molecules needs to be considered.

The rotational relaxation numbers for carbon dioxide, nitrogen, and oxygen should be very close at Martian temperatures[12] and can be given by Eq. (2).

The differences in atmospheric conditions between Earth and Mars result in the low amplitude, low frequency speed of sound on Mars being 66% lower than on Earth.[57] This coupled with the lower ambient density results in an acoustic impedance which is two orders of magnitude smaller than on Earth. This two orders of magnitude difference implies that not only would it take 100 times the particle velocity to get the same acoustic pressure as on Earth, but that pressure wave signals will be 20 dB weaker on Mars[14] for the same acoustic pressure ignoring attenuation.

Model predictions by Bass and Chambers[12] and Petculescu and Lueptow[14] give the absorption of sound on the surface of Mars to be at least 100 times larger than on Earth. This is dominated by classical thermal-viscous losses for frequencies within the continuum limit and larger than 10 KHz. For frequencies smaller than 10 KHz, absorption is dominated by vibrational relaxation losses from the relaxation of the doubly degenerate bending mode of carbon dioxide.[12] This allows the absorption and dispersion to be written in terms of a single relaxation time so the theory is not complicated by complex energy transfer pathways as on Earth. In addition, Bass and Chambers used the low frequency approximation to the classical absorption coefficient.

DSMC results for the maximum acoustic pressure amplitude as a function of distance is plotted in Fig. 6 for Kn =0.02. This corresponds to a frequency of 1.3 MHz, which is significantly above the vibrational relaxation frequency of carbon dioxide and within the continuum assumption. At this frequency, classical thermal-viscous absorption dominates. Simulations were performed at 700 Pa and 200 K. In this case, a mixture of carbon dioxide, nitrogen, argon, and oxygen (water vapor and other trace gases were omitted in the simulations) was considered. DSMC results are plotted against theoretical predictions presented by Bass and Chambers. Comparisons between the DSMC results and theoretical predictions are quite good.

Rotational relaxation at Martian temperatures occurs simultaneously with translational relaxation given the low relaxation collision numbers given by Eq. (2). For frequencies corresponding to Kn$>$0.05, the time delay between the exchange in energy between translational and internal modes becomes more noticeable, creating a higher state of nonequilibrium. At high Kn the frequency of oscillation is well above the relaxation frequencies for rotation. DSMC results showing the degree of nonequilibrium is plotted in Fig. 7 where the temperatures associated with the translational and rotational modes are computed for a frequency of 130 MHz corresponding to Kn=2. Despite starting in equilibrium, the temperatures associated with the translational and rotational modes in this case are considerably different. In this case, slow translational and rotational relaxation effects are more evident. Very little energy has relaxed from the translational mode into the rotational mode where the rotation mode of the molecules is in its frozen state.

C. Titan

Scientists have tried for decades to penetrate the thick, hazy atmosphere of Titan with a variety of telescopes, but were only able to obtain vague hints at the surface below. The surface was considered "hidden" after the flyover by the Voyager I mission in 1980, but using the Hubble to observe Titan at specific wavelengths, gross characterization of relative bright and dark regions of Titan's surface could be established.[58] Recently, however, the Cassini-Huygens mission unveiled a great deal about Titan and its atmosphere.[1,59] Titan's landscape resembles a younger, colder Earth and is the only other object in the solar system that has stable bodies of surface liquid.[60] Titan's atmosphere is made up of 90%–97% nitrogen, with at least a dozen of other trace

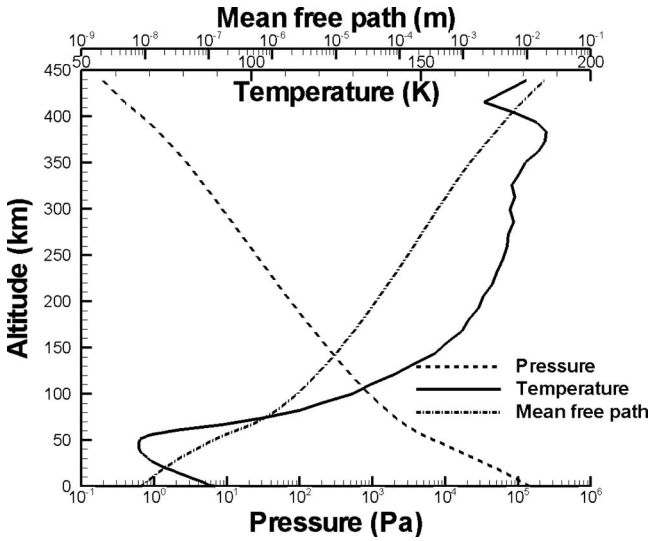


FIG. 8. Temperature (solid line), pressure (dashed line), and mean free path (dash-dotted line) profiles as a function of altitude above Titan's surface (Ref. 11).

gases, including methane, argon, hydrogen, ethane, propane, acetylene, hydrogen cyanide, and helium, which implies that Titan's atmosphere is chemically active.

At Titan's surface, the atmospheric pressure is 1.6 atm and its temperature of 90 K is much colder than that on Earth's surface. The atmosphere extends 500 km above the surface, and the temperature and pressure are a complex function of altitude.[42,59] Temperature, pressure, and mean free path profiles are plotted as a function of altitude in Fig. 8.[11] In comparing Fig. 8 to Fig. 1, there is only a six orders of magnitude change in pressure, and therefore the mean free path, over 500 km in altitude on Titan as opposed to over ten orders of magnitude change in pressure over only 300 km on Earth. In addition, the molecular composition as a function of altitude does not change very much. Nitrogen and methane are still the prominent constituents at high altitudes. Recent calculations confirm that the concentration of methane at the surface is about 3.3% but decreases slightly with increasing altitude.[6] Hydrogen is the next most prominent gas in the upper atmosphere, but is present in trace amounts (0.1%).[42]

For the DSMC simulations of Titan, a mixture of 95% nitrogen and 5% methane (trace gases were omitted in our analysis) was considered.

Similar to Mars, because of the low temperatures on Titan, the vibrational energies of nitrogen or methane are not active and do not contribute to the specific heat for the frequencies under consideration. Therefore, we only need to take into consideration the rotational internal energy of the molecules.

The rotational relaxation number for nitrogen-nitrogen collisions is again based on experimental values given by Eq. (2). Recent experimental values for the rotational relaxation number for methane are available for several rotational modes of methane for temperatures ranging from 100 to 296 K for pure methane and methane-nitrogen mixtures.[61] Based on these measurements the collision number for methane-methane collisions was set to 0.5 and 7.0 for methane-nitrogen collisions.[61]

Because of the large concentrations of nitrogen on both Titan and Earth, the molecular weights are comparable between the two planets. However, with the low temperatures on Titan the speed of sound is 60% of the speed of sound on Earth. When compared to Mars, where the molecular weight is 1.5 larger due to the abundance of carbon dioxide, the speed of sound on Titan is 88% of the speed of sound on Mars. In addition, due to an order of magnitude difference in acoustic impedance between Earth and Titan, pressure wave signals in Titan's dense atmosphere would be 10 dB higher than on Earth for the same acoustic pressure making Titan acoustically responsive.[14] This implies that it is easier to produce and sustain high-amplitude signals, possibly leading to significant nonlinear effects.

Predictions by Petculescu and Leuptow give the absorption of sound to be ten times smaller on Titan than on Earth because of the dense pressures and low temperatures.[14] At these low temperatures, the molecules are in their "frozen" state and absorption is therefore dominated by classical thermal-viscous losses. Because of the low absorption and high acoustic impedance, nonlinear waves can travel a long distance before being absorbed by the atmosphere.

A useful parameter in discussing the relative importance between nonlinearity and absorption is the Gol'dberg number Γ, which is defined as:

$$\Gamma = \frac{1}{\bar{x}\alpha_{\mathrm{cl}}}, \tag{3}$$

where $\bar{x}$ is the shock formation distance and α_{cl} is the low frequency classical absorption coefficient.[62] The shock formation distance is defined as

$$\bar{x} = \frac{1}{(\beta_{\mathrm{NL}}\epsilon k_0)}, \tag{4}$$

where β_{NL} is the coefficient of nonlinearity and is defined by $\beta_{\mathrm{NL}}=(\gamma+1)/2$, where γ is the ratio of specific heats, $\epsilon = u_0/c_0$ is the acoustic Mach number, u_0 is the acoustic velocity amplitude, k_0 is the acoustic wavenumber, and c_0 is the low frequency, low amplitude speed of sound. For Gol'dberg number values of $\Gamma > 1$ nonlinear effects dominate and shock formation is imminent. For $\Gamma < 1$ absorption dominates over nonlinearity and dissipative effects attenuate the acoustic signal enough to keep shock waves from forming. However, it should be noted that since the Gol'dberg number is defined by continuum parameters and the low frequency approximation of the classical absorption coefficient, the Gol'dberg number is not well defined for high Kn situations. In addition, it also may not be well suited for frequencies where the absorption due to a relaxation process dominates since by definition, the Gol'dberg number is a function of the classical absorption coefficient.

Given the order of magnitude difference between the absorption coefficient on Titan and Earth and the similarity between the coefficient of nonlinearities on both planets, the Gol'dberg number, given by Eq. (3), on Titan is an order of

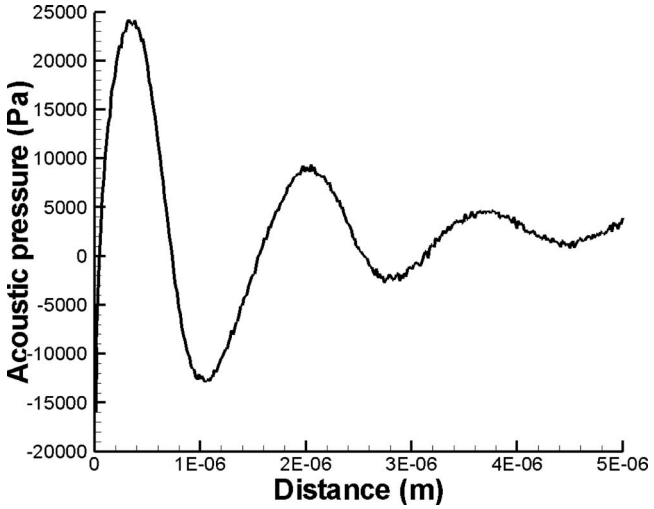


FIG. 9. Kn=0.042 waveform on Earth for ϵ=0.14 at p=101 000 Pa, T=273 K, and $\lambda_m=6.3\times10^{-8}$ m. Absorption dominates nonlinearity with little or no nonlinear effects visible.

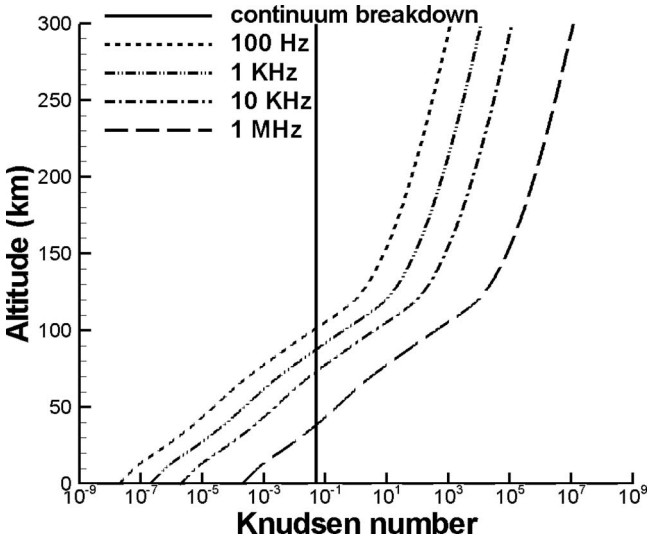


FIG. 11. Knudsen number as a function of altitude on Earth for frequencies of 100 Hz, 1 KHz, 10 KHz, and 1 MHz.

magnitude higher than on Earth for the same frequency and amplitude. As an example, DSMC simulations were performed at the surface of Earth (p=101 000 Pa, T=273 K, and $\lambda_m=6.3\times10^{-8}$ m) and Titan (p=151 500 Pa, T=95 K, and $\lambda_m=4^{-8}$ m) for a frequency of 230 MHz and acoustic Mach number ϵ=0.14 and are shown in Figs. 9 and 10. This corresponds to Knudsen numbers of 0.042 on Earth and 0.01 on Titan, well within the continuum approximation. For this frequency, classical thermal-viscous losses will dominate on both planets, so the Gol'dberg number is a good measure of the importance of nonlinearity. The Gol'dberg number is calculated to be 3.8 on Titan and 0.95 on Earth. This implies, and Fig. 10 suggests, that nonlinearity dominates for this frequency and amplitude on Titan as evidenced by significant wave steepening of the waveform. In contrast, Fig. 9 shows little, if any, wave steepening and significant attenuation on Earth, demonstrating the dominance of absorption over nonlinearity in this case. Therefore, nonlinear effects need to be considered when predicting and modeling sound propagation behavior on Titan.

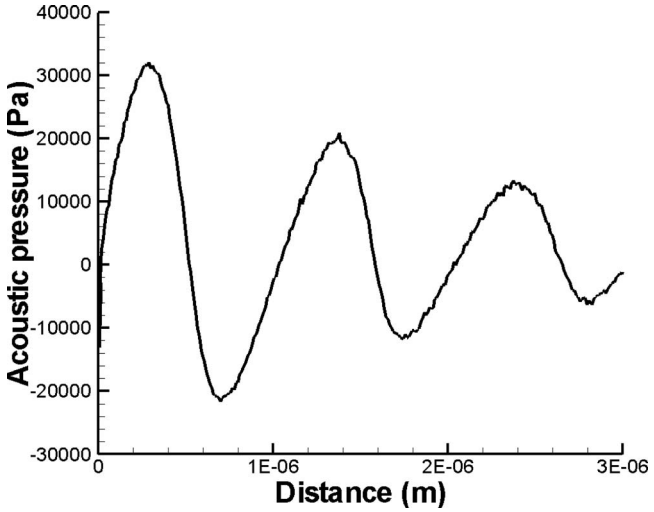


FIG. 10. Kn=0.01 waveform on Titan for ϵ=0.14 at p=151 500 Pa, T=95 K, and $\lambda_m=5\times10^{-8}$ m. Significant wave steepening can be observed.

D. Vertical profiles

On all three planets, atmospheric pressure decreases approximately exponentially with altitude, as shown in Figs. 1, 5, and 8, and the relationship between temperature and altitude varies among the different atmospheric layers. Close to the surface, the atmosphere can be assumed to be continuum for low frequencies and can be treated as a perfect gas. However, the continuum assumption gradually breaks down as altitude increases and diffusion and vertical transport become more important. In order to fully understand the acoustic environment as a function of altitude, particle methods are necessary when the Knudsen number reaches 0.05 in order to capture nonequilibrium and high Kn effects.

The Knudsen numbers for several frequencies as function of altitude are plotted for Earth, Mar, and Titan in Figs. 11–13, respectively. On Earth, all frequencies over 100 Hz reach the continuum limit by 100 km in altitude. Similar to Mars, the continuum limit is reached at approximately 90 km in altitude for the same frequencies. However, on Titan, the continuum assumption is still valid for frequencies less than approximately 1000 Hz at all altitudes because Titan's atmosphere is distended due to its low relative gravity. This implies that particle method solutions are required when investigating the absorption and dispersion of sound for frequencies above 100 Hz and 100 km on Earth and Mars because of the deviation from continuum theory. However, particle methods are necessary only for frequencies above 1 MHz at 100 km altitude on Titan.

Previous work for studying sound propagation as a function of altitude is given for Earth up to 160 km (Ref. 45) and for Earth, Venus, Mars, and Titan up to 120 km.[14] Sutherland and Bass used a theory by Greenspan[52] to model translational and rotational relaxation and an empirical adjustment to account for the deviation to theory at high Kn. Diffusion

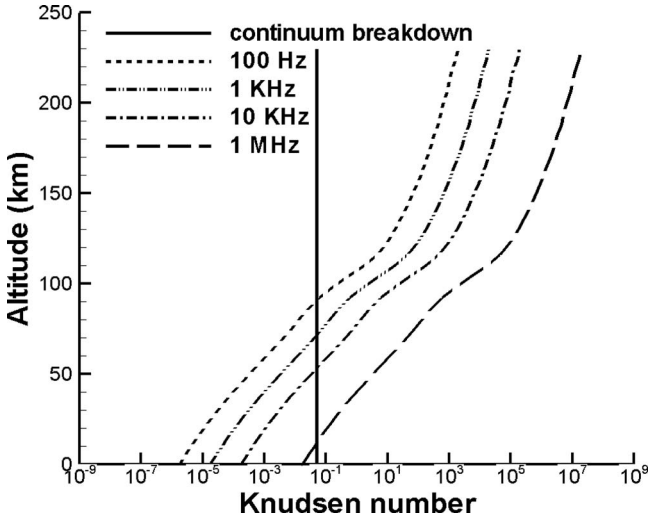


FIG. 12. Knudsen number as a function of altitude on Mars for frequencies of 100 Hz, 1 KHz, 10 KHz, and 1 MHz.

losses and vibrational relaxation losses are also accounted for based on estimates of the relaxation frequencies of the primary molecular constituents as a function of altitude. Temperature and molecular composition dependence were considered in mean atmospheric conditions, including humidity, viscosity, and the specific heat ratio. Rotational relaxation and classical thermal-viscous losses were found to be the prominent loss mechanisms at altitudes above 90 km.

In the predictions given by Petculescu and Lueptow,[14] an effective wavenumber approach developed by Dain and Lueptow[63] is used to model relaxation phenomenon. In this technique, the linearized continuum conservation equations in addition to temperature relaxation equations are used to form an eigenvalue problem. The relaxation equations include kinetic theory results for collision rates and quantum mechanics predictions for transition probabilities. The result can be formed into an effective wavenumber for determining the acoustic absorption and dispersion due to a relaxation process. Petculescu and Lueptow used this effective wavenumber theory in conjunction with the low frequency classical absorption to describe the acoustic properties in planetary

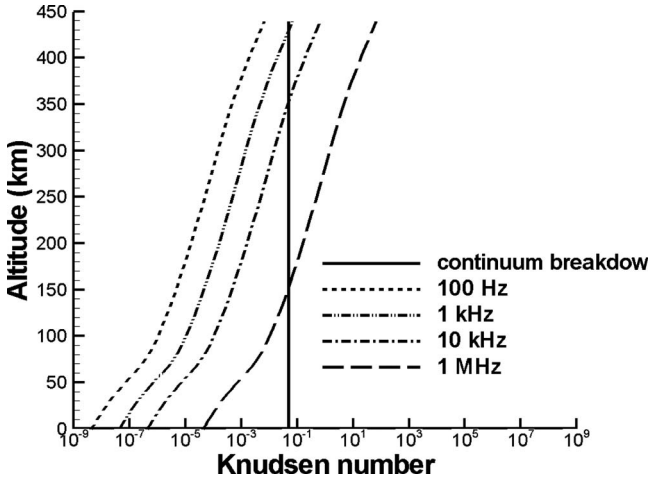


FIG. 13. Knudsen number as a function of altitude on Titan for frequencies of 100 Hz, 1 KHz, 10 KHz, and 1 MHz.

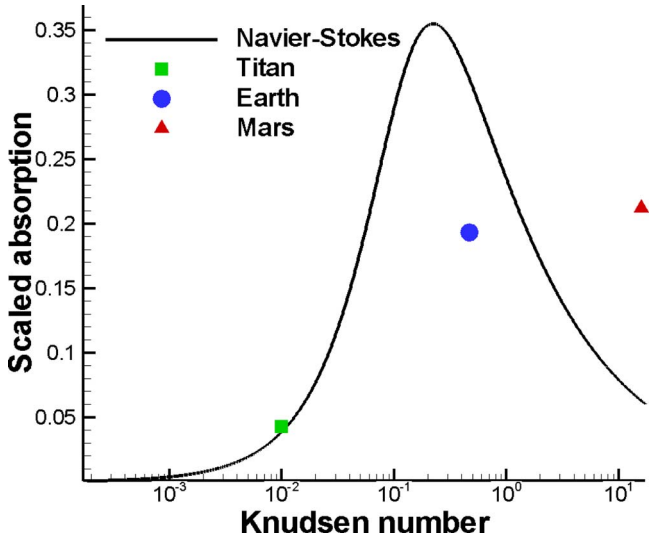


FIG. 14. (Color online) The scaled absorption for a 70 MHz signal on Earth (circle), Mars (triangle), and Titan (square) at an altitude of 25 km compared to Navier–Stokes predicted thermal-viscous losses (black line)

environments. While this method may prove to be computationally effective, the assumptions of using the linearized continuum equations in addition to the limit of low frequencies limit its usefulness. In addition, translational and rotational relaxations, which have been shown to be important in high Kn situations, were not considered. Thus, care must be taken when applying this effective wavenumber approach as a function of altitude.

DSMC results for the scaled absorption α/k_0 for a frequency of 70 MHz and an altitude of 25 km were computed for Earth, Mars and Titan based on general circulation model data. Given the atmospheric conditions of each planet, this choice of frequency results in values of Knudsen numbers that span a large range. Mars results in the largest Knudsen number of 16, high above the continuum breakdown and into the free-molecular flow regime. On Earth, the Knudsen number is smaller, but still above the continuum limit at 0.47, and Titan results in a Knudsen number of 0.01 at this frequency. The scaled absorption α/k_0 for each planet at an altitude of 25 km is shown in Fig. 14 and compared to Navier–Stokes predictions for classical thermal-viscous losses.[64] The trends shown by the data points follows the trend of other data available for high Kn behavior.[44,48] Deviation from Navier–Stokes is seen for results on Earth and Mars given the high Knudsen numbers associated with the frequency of 70 MHz as expected. Results for Titan, however, agree quite well with predicted values as expected for the Kn associated with this simulation. The results for Earth and Mars approach the free-molecular, frozen, limit for the scaled absorption of 0.2.[48]

The atmospheric composition of these planets overlap in several ways as a function of altitude allowing for the possibility of creating acoustic analogies between the atmospheres. For instance, the mean free path at Earth's surface is roughly equivalent to the mean free path on Titan at about 50 km altitude. In addition, the mean free path at Mars' surface is equivalent to the mean free path on Earth at about 40 km altitude or 120 km on Titan.

V. CONCLUSIONS

Using the DSMC method, it is possible to simulate sound for a wide range of planetary environments. Investigations on three different planets, Earth, Mars, and Titan, were performed to describe various sound propagation properties for a wide range of Kn. DSMC results for a simple molecular relaxation were performed in Earth-like conditions. Relaxation peaks for multiple relaxation collision numbers were seen. For frequencies above the vibrational relaxation frequency, results on Earth agreed very well with classical thermal-viscous theory. On Mars, DSMC results show that the absorption of sound is 100 times greater on Mars than on Earth and DSMC results agreed well with theory. Nonequilibrium effects are seen for large Kn on Mars. Titan's thick atmosphere is more conducive to making and sustaining nonlinear waves. The Gol'dberg number on Titan is larger than on Earth implying the importance of nonlinearity. Therefore, nonlinear effects need to be considered when predicting and modeling sound propagation behavior on Titan.

The Knudsen number is large in upper atmospheric conditions and requires a particle method solution due to the breakdown of the continuum assumption. Large deviations from continuum theory were seen for high Kn due to strong nonequilibrium effects. DSMC should be the method of choice for describing the acoustic phenomenon in high altitude situations.

Sound propagation through ionized gases has not been analyzed with the current DSMC model which could lead to further relaxation effects.[65] For instance, Earth's atmosphere begins to be ionized by solar radiation at about 65 km of altitude. On Mars, the ionosphere varies spatially throughout the planet, mostly confined to the southern hemisphere.[66] Titan's ionosphere is also heavily affected by Saturn's magnetosphere.[67] Inclusion of wave propagation through ionized gases would make this work more robust.

Computation and memory costs are high, in particular for low Kn simulations. Because of computer limitations, DSMC is currently limited to situations where $\mathrm{Kn} \geqslant 0.0008$. This implies that DSMC is currently not able to simulate audible frequencies or large scale simulations. Therefore, many of the results are for frequencies not commonly used in atmospheric conditions. In addition, the vibrational relaxation collision numbers used in this study are not physically realistic but where chosen based on the size of the system which is a computational restriction.

DSMC is a robust, physically realistic model that includes nonlinearity, absorption, relaxation, and internal energy effects which has been used to model sound propagation in a variety of planetary environments where experimentation is difficult. Despite large computation time and memory requirements, the use of DSMC to study planetary environments is the method best suited to study high altitude or where continuum theory is not valid.

ACKNOWLEDGMENTS

Support from the NASA GSRP Fellowship Program is gratefully acknowledged. In addition, the authors thank Patrick D. O'Connor, James B. Anderson, Feri Farassat, and Victor W. Sparrow for their support on this project.

[1] P. R. Mahaffy, "Intensive Titan exploration begins," Science **308**, 969–970 (2005).
[2] J.-P. Lebreton, O. Witasse, C. Sollazzo, T. Blancquaert, P. Couzin, A.-M. Schipper, J. B. Jones, D. L. Matson, L. I. Gurvits, D. H. Atkinson, B. Kazeminejad, and M. Pérez-Ayúcar, "An overview of the descent and landing of the Huygens probe on Titan," Nature (London) **438**, 758–764 (2005).
[3] J. C. Zarnecki, M. R. Leese, J. R. C. Garry, N. Ghafoor, and B. Hathi, "Huygens' surface science package," Space Sci. Rev. **104**, 593–611 (2002).
[4] J. C. Zarnecki, M. R. Leese, B. Hathi, A. J. Ball, A. Hagermann, M. C. Towner, R. D. Lorenz, A. M. McDonnell, S. F. Green, M. R. Patel, T. J. Ringrose, P. D. Rosenberg, K. R. Atkinson, M. D. Paton, M. Banaszkiewicz, B. C. Clark, F. Ferri, M. Fulchignoni, N. A. Ghafoor, G. Kargl, H. Svedhem, J. Delderfield, M. Grande, D. J. Parker, P. G. Challenor, and J. E. Geake, "A soft solid surface on Titan as revealed by the Huygens Surface Science Package," Nature (London) **438**, 792–795 (2005).
[5] M. C. Towner, J. R. C. Garry, R. D. Lorenz, A. Hagermann, B. Hathi, H. Svedhem, B. C. Clark, M. R. Leese, and J. C. Zarnecki, "Physical properties of Titan's surface at the Huygens landing site from the Surface Science Package Acoustic Properties sensor (API-S)," Icarus **185**, 457–465 (2006).
[6] A. Hagermann, P. D. Rosenberg, M. C. Towner, J. R. C. Garry, H. Svedhem, M. R. Leese, B. Hathi, R. D. Lorenz, and J. C. Zarnecki, "Speed of sound measurements and the methane abundance in Titan's atmosphere," Icarus **189**, 538–543 (2007).
[7] R. D. Lorenz, "Speed of sound in outer planet atmospheres," Planet. Space Sci. **47**, 67–77 (1998).
[8] Information on the Mars Microphone available at http://sprg.ssl.berkeley.edu/marsmic/welcome.html (Last viewed April, 2008).
[9] Online database for Earth available at http://www.spenvis.oma.be/ (Last viewed April, 2008).
[10] Online database for Mars available at http://www-mars.lmd.jussieu.fr/ (Last viewed April, 2008).
[11] Online database for Titan available at http://www.lmd.jussieu.fr/titanDbase/ (Last viewed April, 2008).
[12] H. E. Bass and J. P. Chambers, "Absorption of sound in the Martian atmosphere," J. Acoust. Soc. Am. **109**, 3069–3071 (2001).
[13] J.-P. Williams, "Acoustic environment of the Martian surface," J. Geophys. Res. **106**, 5033–5041 (2001).
[14] A. Petculescu and R. M. Lueptow, "Atmospheric acoustics of Titan, Mars, Venus, and Earth," Icarus **186**, 413–419 (2007).
[15] G. A. Bird, *Molecular Gas Dynamics and the Direct Simulation of Gas Flows* (Clarendon, Oxford, 1994).
[16] A. L. Danforth and L. N. Long, "Nonlinear acoustic simulations using direct simulation Monte Carlo," J. Acoust. Soc. Am. **116**, 1948–1955 (2004).
[17] F. J. Alexander and A. L. Garcia, "The direct simulation Monte Carlo method," Comput. Phys. **11**, 588 (1997).
[18] W. Wagner, "A convergence proof for Bird's direct simulation Monte Carlo method for the Boltzmann equation," J. Stat. Phys. **66**, 1011–1044 (1992).
[19] E. P. Muntz, "Rarefied gas dynamics," Annu. Rev. Fluid Mech. **21**, 387–422 (1989).
[20] G. A. Bird, "Monte Carlo simulation of gas flows," Annu. Rev. Fluid Mech. **10**, 11–31 (1978).
[21] E. Oran, C. Oh, and B. Cybyk, "Direct simulation Monte Carlo: Recent advances and applications," Annu. Rev. Fluid Mech. **30**, 403–441 (1998).
[22] G. A. Bird, "Direct simulation and the Boltzmann equation," Phys. Fluids **13**, 2676–2681 (1970).
[23] L. N. Long, "Navier stokes and Monte Carlo results for hypersonic flows," AIAA J. **29**, 200–207 (1991).
[24] A. D. Hanford, P. D. O'Connor, J. B. Anderson, and L. N. Long, "Predicting absorption and dispersion in acoustics by direct simulation Monte Carlo: Quantum and classical models for molecular relaxation," J. Acoust. Soc. Am. **123**, 4118–4126 (2008).
[25] N. G. Hadjiconstantinou and A. L. Garcia, "Molecular simulations of sound wave propagation in simple gases," Phys. Fluids **13**, 1040–1046 (2001).
[26] P. D. O'Connor, L. N. Long, and J. B. Anderson, "Accurate rate expressions for simulations of gas-phase chemical reactions," J. Comput. Phys.

227, 7664–7673 (2008).
[27]M. Ikegawa and J. Kobayashi, "Deposition profile simulation using the direct simulation Monte Carlo method," J. Electrochem. Soc. **136**, 303–310 (1989).
[28]D. G. Coronell and K. F. Jensen, "Monte Carlo simulations of very low pressure chemical vapor deposition," J. Comput.-Aided Mater. Des. **1**, 3–26 (1989).
[29]P. D. O'Connor, L. N. Long, and J. B. Anderson, "The direct simulation of detonations," in AIAA/ASME/SAE/ASEE Joint Propulsion Conference, (2006), AIAA Paper No. 2006–4411.
[30]J. B. Anderson and L. N. Long, "Direct Monte Carlo simulation of chemical reaction systems: Prediction of ultrafast detonations," J. Chem. Phys. **118**, 3102–3110 (2003).
[31]J. Zhang, D. B. Goldstein, P. L. Varghese, L. Trafton, C. Moore, and K. Miki, "Numerical modeling of ionian volcanic plumes with entrained particulates," Icarus **172**, 479–502 (2004).
[32]J. V. Austin and D. B. Goldstein, "Direct numerical simulation of circumplanetary winds on Io," Bull. Am. Astron. Soc. **29**, 1004 (1997).
[33]S. Chapman and T. G. Cowling, *The Mathematical Theory or Non-Uniform Gases*, 3rd ed. (Cambridge University Press, Cambridge, 1970).
[34]M. Greenspan, "Propagation of sound in five monatomic gases," J. Acoust. Soc. Am. **28**, 644–648 (1956).
[35]C. Borgnakke and P. S. Larsen, "Statistical collision model for Monte Carlo simulation of polyatomic gas mixture," J. Comput. Phys. **18**, 405–420 (1975).
[36]A. Danforth-Hanford, P. D. O'Connor, L. N. Long, and J. B. Anderson, "Molecular relaxation simulations in nonlinear acoustics using direct simulation Monte Carlo," in Proceedings from the 7th International Symposium on Nonlinear Acoustics (2005).
[37]A. D. Hanford, L. N. Long, and V. W. Sparrow, "The propagation of sound on Titan using the direct simulation Monte Carlo," J. Acoust. Soc. Am. **121**, 3117 (2007).
[38]A. D. Hanford and L. N. Long, "The absorption of sound on Mars using the direct simulation Monte Carlo," J. Acoust. Soc. Am. **119**, 3264 (2006).
[39]F. J. Alexander, A. L. Garcia, and B. J. Alder, "Cell size dependence of transport coefficients in stochastic particle algorithms," Phys. Fluids **10**, 1540–1542 (1998).
[40]F. Forget, F. Hourdin, R. Fournier, C. Hourdin, and O. Talagrand, "Improved general circulation models of the Martian atmosphere from the surface to above 80 km," J. Geophys. Res. **104**, 24155–24175 (1999).
[41]S. R. Lewis, M. Collins, P. L. Read, F. Forget, F. Hourdin, R. Fournier, C. Hourdin, O. Talagrand, and J.-P. Hout, "A climate database for Mars," J. Geophys. Res. **104**, 24177–24194 (1999).
[42]P. Rannou, S. Lebonnois, F. Hourdin, and D. Luz, "Titan atmosphere database," Adv. Space Res. **36**, 2194–2198 (2005).
[43]H. E. Bass, L. C. Sutherland, A. J. Zuckerwar, D. T. Blackstock, and D. M. Hester, "Atmospheric absorption of sound: Further developments," J. Acoust. Soc. Am. **97**, 680–683 (1995).
[44]M. Greenspan, "Rotational relaxation in nitrogen, oxygen, and air," J. Acoust. Soc. Am. **31**, 155–160 (1959).
[45]L. C. Sutherland and H. E. Bass, "Atmospheric absorption in the atmosphere up to 160 km," J. Acoust. Soc. Am. **115**, 1012–1032 (2004).
[46]J. K. Buckner and J. H. Ferziger, "Linearized boundary value problem for a gas and sound propagation," Phys. Fluids **9**, 2315–2322 (1966).
[47]L. Sirovich and J. K. Thurber, "Propagation of forced sound waves in rarefied gasdynamics," J. Acoust. Soc. Am. **37**, 329–339 (1965).
[48]R. Schotter, "Rarefied gas acoustics in the noble gases," Phys. Fluids **17**, 1163–1168 (1974).
[49]F. K. Lutgens and E. J. Tarbuck, *The Atmosphere*, 6th ed. (Prentice-Hall, Englewood Cliffs, NJ, 1995).
[50]H. E. Bass, L. C. Sutherland, J. Piercy, and L. Evans, *Physical Acoustics; Principles and Methods*, (Academic, New York, 1984), Vol. **XVII**, pp. 145–232.
[51]K. F. Herzfeld and T. A. Litovitz, *Absorption and Dispersion of Ultrasonic Waves* (Academic, New York, 1959).
[52]M. Greenspan, "Combined translational and relaxational dispersion of sound in gases," J. Acoust. Soc. Am. **26**, 70–73 (1954).
[53]T. Owen, K. Biemann, D. R. Rushneck, J. E. Biller, D. W. Howarth, and A. L. Lafleur, "The composition of the atmosphere at the surface of Mars," J. Geophys. Res. **82**, 4635–4639 (1977).
[54]J. E. Tillman, "Mars global atmospheric oscillations: Annually synchronized, transient normal-mode oscillations and the triggering of global dust storms," J. Geophys. Res., [Atmos.] **93**, 9433–9451 (1988).
[55]D. P. Hinson, R. A. Simpson, J. D. Twicken, G. L. Tyler, and F. M. Fasar, "Initial results from radio occultation measurements with Mars Global Surveyor," J. Geophys. Res., [Planets] **104**, 26297–27012 (1999).
[56]M. T. Lemmon, M. J. Wolff, M. D. Smith, R. T. Clancy, D. Banfield, G. A. Landis, A. Ghosh, P. H. Smith, N. Spanovich, B. Whitney, P. Whelley, and R. Greeley, "Atmospheric imaging results from the Mars Exploration Rovers: Spirit and Opportunity," Science **306**, 1753–1756 (2004).
[57]V. W. Sparrow, "Acoustics on the planet Mars: A preview," J. Acoust. Soc. Am. **106**, 2264–2264 (1999).
[58]P. H. Smith, M. T. Lemmon, R. D. Lorenz, L. A. Sromovsky, J. J. Caldwell, and M. D. Allison, "Titan's surface, revealed by HST imaging," Icarus **119**, 336–349 (1996).
[59]F. M. Flasar, R. K. Achterberg, B. J. Conrath, P. J. Gierasch, V. G. Kunde, C. A. Nixon, G. L. Bjoraker, D. E. Jennings, P. N. Romani, A. A. Simon-Miller, B. Bzard, A. Coustenis, P. G. J. Irwin, N. A. Teanby, J. Brasunas, J. C. Pearl, M. E. Segura, R. C. Carlson, A. Mamoutkine, and P. J. Schinder, "Titan's atmospheric temperatures, winds, and composition," Science **308**, 975–978 (2005).
[60]E. R. Stofan, C. Elachi, J. I. Lunine, R. D. Lorenz, B. Stiles, K. L. Mitchell, S. Ostro, L. Soderblom, C. Wood, H. Zebker, S. Wall, M. Janssen, R. Kirk, R. Lopes, F. Paganelli, J. Radebaugh, L. Wye, Y. Anderson, M. Allison, R. Boehmer, P. Callahan, P. Encrenaz, E. Flamini, G. Francescetti, Y. Gim, G. Hamilton, S. Hensley, W. T. K. Johnson, K. Kelleher, D. Muhleman, P. Paillou, G. Picardi, F. Posa, L. Roth, R. Seu, S. Shaffer, and S. V. nd R. West, "The lakes of Titan," Nature (London) **445**, 61–64 (2007).
[61]F. M. Bourcin, J. Menard, and C. Boursier, "Temperature dependence of rotational relaxation of methane in the $2v_3$, vibrational state by self- and nitrogen-collisions and comparison with tline broadening measurements," J. Mol. Spectrosc. **242**, 55–63 (2007).
[62]M. F. Hamilton and D. T. Blackstock, eds., *Nonlinear Acoustics* (Academic, San Diego, 1998).
[63]Y. Dain and R. M. Lueptow, "Acoustic attenuation in three-component gas mixtures—Theory," J. Acoust. Soc. Am. **109**, 1955–1964 (2001).
[64]L. Rayleigh, *Theory of Sound*, 2nd ed. (Dover, New York, 1945).
[65]M. Numano, "Relaxation effect on sound propagation in partially ionized gases," Plasma Phys. **18**, 519–524 (1976).
[66]J. G. Trotignon, H. C. Seran, C. Beghin, N. Meyer-Vernet, R. Manning, R. Grard, and H. Laakso, "In situ observations of the ionized environment of Mars: the antenna impedance measurements experiment, AIM, proposed as part of the Mars advanced radar for subsurface and ionospheric sounding," Planet. Space Sci. **49**, 155–164 (2001).
[67]J. E. Wahlund, R. Boström, G. Gustafsson, D. A. Gurnett, W. S. Kurth, A. Pedersen, T. F. Averkamp, G. B. Hospodarsky, A. M. Persoon, P. Canu, F. M. Neubauer, M. K. Dougherty, I. E. Ai, M. W. Morooka, F. Gill, M. André, L. Eliasson, and I. Müller-Wodarg, "Cassini measurements of cold plasma in the ionosphere of Titan," Science **308**, 986–689 (2005).

Mesoscale variations in acoustic signals induced by atmospheric gravity waves

Igor Chunchuzov, Sergey Kulichkov, and Vitaly Perepelkin
Obukhov Institute of Atmospheric Physics, 3 Pyzhevskii Per., 129164 Moscow, Russia

Astrid Ziemann, Klaus Arnold, and Anke Kniffka
LIM-Institute for Meteorology, University of Leipzig, Stephanstrasse 3, D-04103 Leipzig, Germany

(Received 21 February 2008; revised 27 November 2008; accepted 2 December 2008)

The results of acoustic tomographic monitoring of the coherent structures in the lower atmosphere and the effects of these structures on acoustic signal parameters are analyzed in the present study. From the measurements of acoustic travel time fluctuations (periods 1 min–1 h) with distant receivers, the temporal fluctuations of the effective sound speed and wind speed are retrieved along different ray paths connecting an acoustic pulse source and several receivers. By using a coherence analysis of the fluctuations near spatially distanced ray turning points, the internal wave-associated fluctuations are filtered and their spatial characteristics (coherences, horizontal phase velocities, and spatial scales) are estimated. The capability of acoustic tomography in estimating wind shear near ground is shown. A possible mechanism describing the temporal modulation of the near-ground wind field by ducted internal waves in the troposphere is proposed.
 [DOI: 10.1121/1.3056477]

PACS number(s): 43.28.Gq, 43.28.Dm, 43.28.Bj, 43.60.Rw [VEO] Pages: 651–663

I. INTRODUCTION

At present, little is known about the effect of the wind speed and temperature fluctuations caused by internal gravity waves (IGWs) on a long-range sound propagation in the atmosphere, particularly on the direction of propagation of acoustic wave front. The estimating of this effect is of practical importance for solving the problem of acoustic localization of different sources in the atmosphere, including explosions. Typical horizontal scales of the wind speed and temperature fluctuations induced by IGWs in the troposphere lie in the range from a hundred meters to a few kilometers, whereas temporal scales are in the range 1 min–1 h. There is also a problem in modeling of the dynamics of IGWs and eddy structures in the stably stratified atmospheric boundary layer (ABL) taking into account their interaction with small-scale turbulence (Finnigan *et al.*, 1984; Danilov and Chunchuzov, 1992; Chimonas, 1999; Chimonas, 2002; Sun *et al.*, 2004; Cooper *et al.*, 2006). Similarly, we are faced with this problem when trying to parametrize a stably stratified ABL in weather forecasting models of the atmosphere (Chimonas, 2002). To solve the problems outlined above, one needs to know the statistical characteristics of the wind speed and temperature fluctuations in the mesoscale range (1 min–1 h). This motivates an experimental study of the coherences and spectra of the mesoscale wind speed fluctuations and their effect on the parameters of acoustic signals (travel time, duration, and direction of propagation of the acoustic signals) which is presented in this paper. For this purpose, we applied different methods of acoustic tomography of the ABL (see, e.g., Ziemann *et al.*, 2001; Chunchuzov *et al.*, 1990; Chunchuzov *et al.*, 2005). Using the measurements of the fluctuations of travel time differences between different arrivals at a net of distant receivers, these methods allow us to retrieve the temporal effective sound speed fluctuations averaged over different acoustic ray paths with turning points at different heights. Based on the coherences and phase spectra of the retrieved variations within different spatial volumes (tomographic cells) of the ABL, the characteristic scales and horizontal speeds of propagation of the organized structures can be estimated (Chunchuzov *et al.*, 2005).

An advantage of the acoustic tomographic measurements of the wind fluctuations in the ABL over point measurements is the capability of obtaining the spatial characteristics of the organized structures whose direction of propagation usually does not coincide with that of the mean wind velocity (so that Taylor's hypothesis of "frozen turbulence" is not valid). By averaging the effective sound speed fluctuations over tomographic cells, the tomographic method automatically filters only those fluctuations whose spatial scales are greater than the cell's sizes and provides a retrieval of their temporal fluctuations in many regions of the ABL simultaneously. This feature turns the acoustic tomography into a promising method for studying the three-dimensional dynamics of the lower atmosphere.

The paper is organized as follows. In Sec. II A, the field experiment is described where we applied acoustic tomography to the lower atmosphere. The results of reconstruction of the temporal variations in the effective sound speed c_{eff} (sound speed plus wind velocity projection on a radius vector from a source to a given receiver) from measurements of travel time intervals between signal arrivals are presented in Sec. II B. The retrieved variations of the effective sound speed gradients, their coherences, and characteristic scales are analyzed in Sec. II C. In this section, the effect of internal wave-associated wind speed gradients on the parameters of acoustic signals (travel time interval between different arrivals, duration, and azimuth of propagation of the wave front) propagating through stable ABL is studied. Based on this

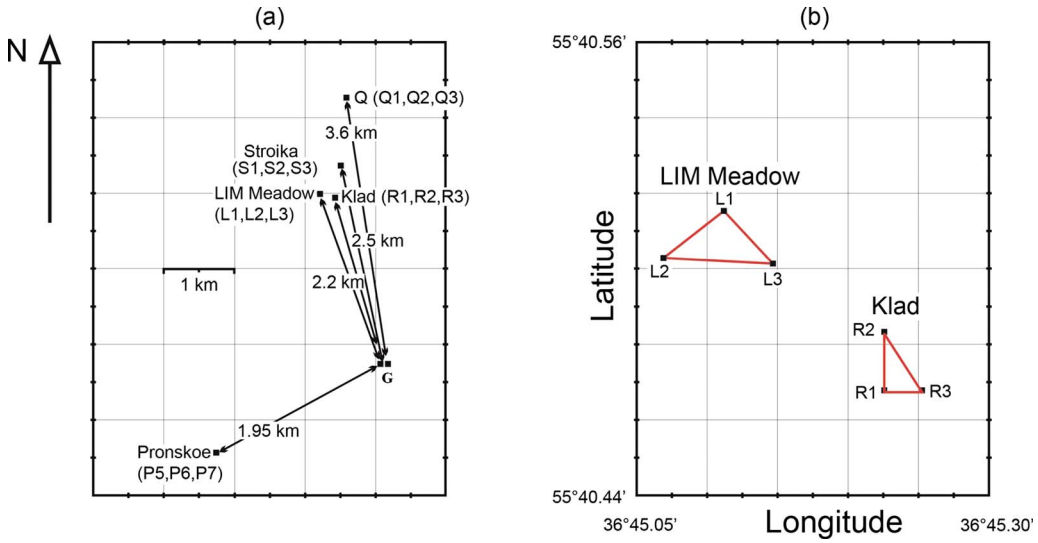


FIG. 1. (Color online) Positions of the source (g) and receivers during experiment near Zvenigorod. (a) The 30 m triangle arrays were located at Klad (receivers R1, R2, and R3), Stroika (receivers S1, S2, and S3), and Pronskoe (receivers P5, P6, and P7), and a larger-size triangle was located at LIM meadow (receivers L1, L2, and L3). One of the 30 m triangles was displaced on July 13, 2005 from Pronskoe to a new position Q (receivers Q1, Q2, and Q3). (b) The orientations of the sides of the triangle (R1, R2, and R3) and of the larger-size triangle (L1, L2, and L3).

study, a possible mechanism of modulation of the parameters of acoustic signals and of the intensity of turbulence by ducted internal waves in the lower troposphere is proposed in Sec. III.

II. ACOUSTIC MONITORING OF THE COHERENT STRUCTURES IN THE ABL

A. Field experiment

To study the influence of the mesoscale wind speed fluctuations on propagation of acoustic signals in the atmosphere, three field experiments were carried out during the 2 year period 2004–2005. These experiments were carried out by researchers from the Leipzig Institute of Meteorology (LIM) and the Oboukhov Institute of Atmospheric Physics (OIAP). Despite differences in the methods of acoustic tomography of the ABL used in the parallel experiments near Torgau, Germany (carried out by LIM in October 2004) and near Zvenigorod, Russia (carried out by OIAP in August 2004), the obtained results, reported by Kniffka *et al.* (2006), Arnold *et al.* (2005), and Chunchuzov *et al.* (2007), showed the existence of quasiperiodic oscillations in the sound travel time with periods of 16–20, 8–10, 4–5, and 2 min. The horizontal phase speeds and scales of these oscillations estimated in both experiments were in a good agreement with each other.

Below, the results of the joint experiment of the OIAP and LIM are described which was carried out in July 2005 at the OIAP's base near Zvenigorod. The stratification of the ABL was continuously measured with a SODAR and a temperature profiler developed by Kadygrov *et al.* (2003). The SODAR measurements with a vertical resolution of 50 m provided us with 5 min averaged wind speed profiles up to a height of 400 m, whereas the temperature profiler measured the 5 min averaged temperature (vertical resolution of 50 m) up to altitudes of about 700 m. Above these altitudes, the wind speed and temperature stratification of the troposphere up to an altitude of 20 km was measured twice a day (in the midnight and midday) by radiosondes with a height resolution of about 200 m.

To monitor organized structures of different scales, the three triangle microphone arrays of different sizes (from 30 to 85 m between receivers of a triangle) were placed at the sites indicated in Fig. 1(a) as Klad, LIM meadow, and Pronskoe. They had different azimuths but almost the same distance r (between 1.95 and 2.2 km) from the source. The triangle arrays in Klad and Pronskoe were of the same side length (about 30 m), whereas the triangles L1, L2, and L3 (LIM meadow) were of a larger size (about 85 m) [Fig. 1(b)]. The measurements were carried out mostly in the early morning and during night hours to ensure stable stratification of the ABL. In the majority of the analyzed cases, 1 h observation periods were chosen.

An example of the acoustic signals generated by a detonation pulse source with a repetition period of 60 s and signals received at distances r about 2 km is shown in Fig. 2. This figure illustrates a strong dependence of the signal's shape on the distance r and the angle ψ between a direction of sound propagation and a direction of the wind velocity. The initially single pulse near the source [Fig. 2(a)] "splits" at a distance of $r=2.2$ km (Klad) from the source into a set of arrivals A, B, C, and D [Fig. 2(b)] due to the formation of an acoustic waveguide. During the time period 23:45–23:50 Moscow local time (hereafter LT) the direction of mean wind relative to the direction from a source to Klad was changing from 20° at a height of 6 m (measured by sonic anemometer) to about 90° at a height of 100 m (at heights more than 20 m the wind was measured by SODAR), whereas mean wind speed was increasing from 0.5 to about 5.5 m/s.

The same signal, but received in Pronskoe and thus having another direction ψ [Fig. 1(a)], does not split into the arrivals at all, but its shape is significantly smoothed due to combined effect of antiwaveguide propagation of the signal

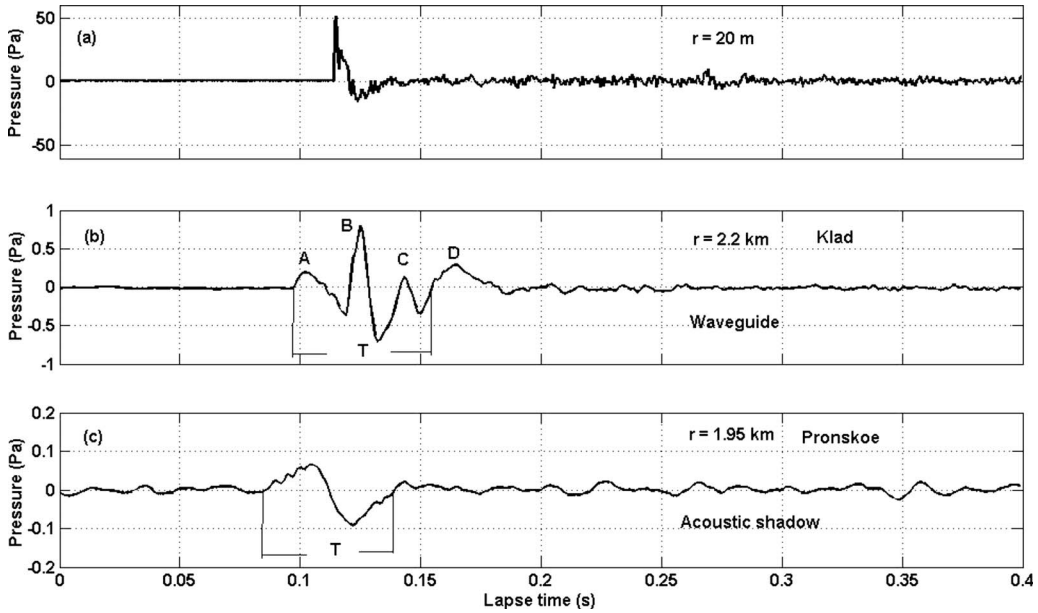


FIG. 2. Acoustic signal received at different distances r from the source on July 12, 2005 at 00:11:55 LT. (a) Signal at $r=20$ m. (b) Signal at $r=2.2$ km (Klad) downwind from the source, $\psi=20°-90°$. (b) Signal at $r=1.95$ km (Pronskoe) upwind from the source, $\psi=120°-190°$. The abscissa is the lapse time with respect to some initial moment of signal recording at the distance r. A, B, C, and D are arrivals of the signal, and T is the time interval.

in the upwind direction ($\psi=120°-190°$) and finite ground impedance (Don and Gramond, 1986; Chunchuzov *et al.*, 1990).

The wind stratification of the ABL played a major role in the existence of the observed azimuthal anisotropy of the signal's shape (Fig. 2) which is well illustrated in Fig. 3, where the profiles of the effective sound speed c_{eff} (averaged over the selected 5 min period of pulse sounding) and the corresponding acoustic ray trajectories are shown for both Pronskoe and Klad. In the case of waveguiding sound propagation [Fig. 3(a)], the refracted rays A, B, and C, having different heights of turning points, cause the corresponding arrivals of the signal [Fig. 3(c)]. Along the direction from the source to Pronskoe (−114° West to the North), the refracted rays bend upward causing the formation of the shadow zone [Fig. 3(b)]. Contrary to the ray approximation, wave theory predicts the existence of the diffracted field of the signal near ground in the upwind direction and explains its form in the shadow zone (Chunchuzov *et al.*, 1990). In this zone the peak amplitude of the signal [Fig. 2(c)] is one order less than that in the waveguide [Fig. 2(b)], and the time duration of the signal T [indicated in Fig. 2(c)] varies in time t with the varying vertical gradients of c_{eff} near ground. The variations in wind speed and temperature also cause the variations in the direction of propagation of acoustic signals whose frequency spectrum was calculated by Kulichkov *et al.* (2006).

To capture organized structures of different scales, we compared the coherences and phase spectra of the time series for the travel time interval $T(t)$ measured by small-size 30 m triangle arrays in Pronskoe [Fig. 2(c)] and Klad [Fig. 2(b)] with those measured by the larger-size triangle (R1, L1, and P5). The multicoherence function K_0 is defined as (see Grachev *et al.*, 1978)

$$K_0=\left[K_{12}^2+K_{23}^2+K_{31}^2-2K_{12}K_{23}K_{31}\cos\left(\sum\varphi\right)\right]^{1/2}, \quad (1)$$

where $K_{ij}^2=|R_{ij}(f)|^2/R_{ii}(f)R_{jj}(f)$ is the coherence function between the fluctuations of $T(t)$ measured by a given pair of the receivers of a triangle labeled by i and j with $i,j=1,2,3$, $R_{ij}=|R_{ij}|\exp(i\varphi_{ij})$ is the cross-spectral density, $\varphi_{ij}=\varphi_i-\varphi_j$ is its phase, φ_i is the phase of the Fourier spectrum of the time series measured by the ith receiver, R_{ii} is the power spectral density for the ith receiver, and $\Sigma\varphi=(\varphi_1-\varphi_2)+(\varphi_2-\varphi_3)+(\varphi_3-\varphi_1)$ is the sum of the phase differences over a triangle. The coherences and phase spectra were estimated by means of standard procedures (Bendat and Piersol, 1967). If K_{12}, K_{23}, and K_{31} are close to 1 and $\Sigma\varphi\approx0$ then $K_0\approx1$. Therefore, the existence within some frequency interval of the maximum of the multicoherence function K_0 along with the condition $\Sigma\varphi\approx0$ may be considered as an indication of the existence of wavelike fluctuations within the same frequency interval (Grachev *et al.*, 1978).

The coherences and phase spectra for the time series $T(t)$ at 30 m triangles in Pronskoe and Klad are shown in Fig. 4 (in Pronskoe the receivers are numbered by 5, 6, and 7 instead of 1, 2, and 3, therefore $i,j=5,6,7$). Because of the pulse repetition period (1 min) chosen in the experiment, the coherences and spectra are limited by an upper frequency of 1/120 Hz, i.e., periods no shorter than 2 min are resolved. One can find common (for both Pronskoe and Klad) frequency intervals 1.5–2, 3.5–4, and 6.5–8 mHz, for which a multicoherence function K_0 [black curves in Figs. 4(a) and 4(c)] reaches local maxima along with the condition $\Sigma\varphi\approx0$ for the existence of the wavelike fluctuations. For the frequencies of about 6.5 mHz (period of about 2.5 min), the estimated horizontal phase speeds of the wavelike fluctuations are low enough (about 1–1.5 m/s) to be detected by a

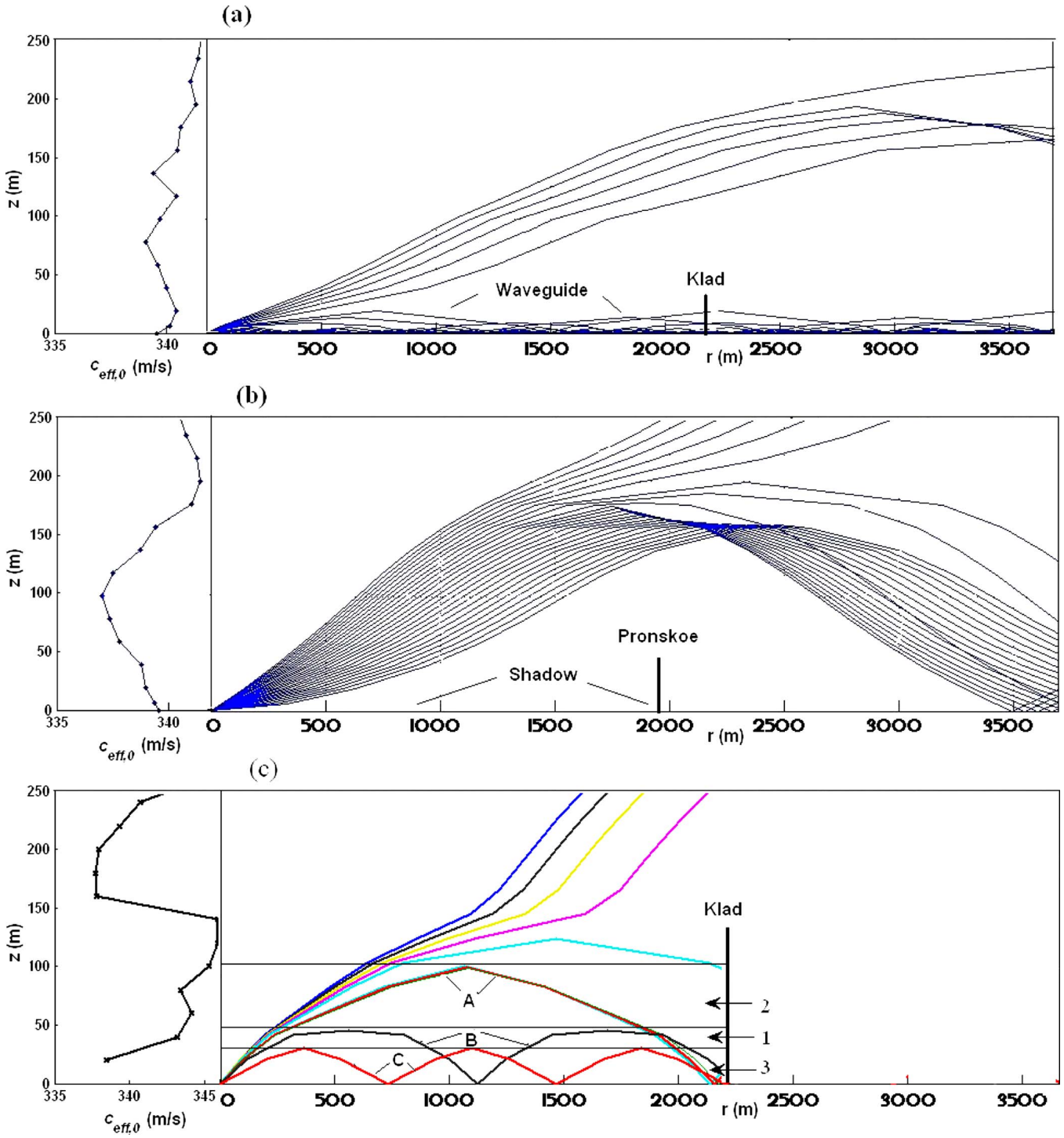


FIG. 3. (Color online) Effective sound speed profiles, $c_{\mathrm{eff},0}(z)$, along with the ray paths calculated for July 11/12, 2005. (a) 23:45–23:50 LT (Klad), waveguiding propagation. (b) 23:45–23:50 LT, antiwaveguide (Pronskoe). (c) 23:25–23:30 LT (Klad). Trajectories A, B, and C correspond to arrivals A, B, and C of the signal in Fig. 2(b). Rectangular cells 1, 2, and 3, for which the temporal fluctuations of c_{eff} were retrieved, are also shown.

30 m triangle from the phase differences, so that these structures have relatively small horizontal scales (about 150–250 m).

Based on the 30 m triangle, it is not possible to estimate the phase differences for higher speed structures such as IGWs. For this purpose, a triangle of larger size was used, such as Pronskoe-LIM-Klad. The distance between receivers L1 and R2 is about 134 m, whereas the distance between receivers L1 and P5 is about 3 km. The distances between the highest turning points of the ray paths, connecting a source and each receiver of the triangle Pronskoe-LIM-Klad, are two times less than the corresponding distances between the receivers of this triangle.

For the triangle Pronskoe-LIM-Klad, the maximum of coherence along with the condition $\Sigma\varphi \approx 0$ was found only for a rather low frequency $f=1.5$ mHz (period of 12 min), for which the estimated phase speed was about 3 m/s and the horizontal wavelength 2 km. These parameters are typical for

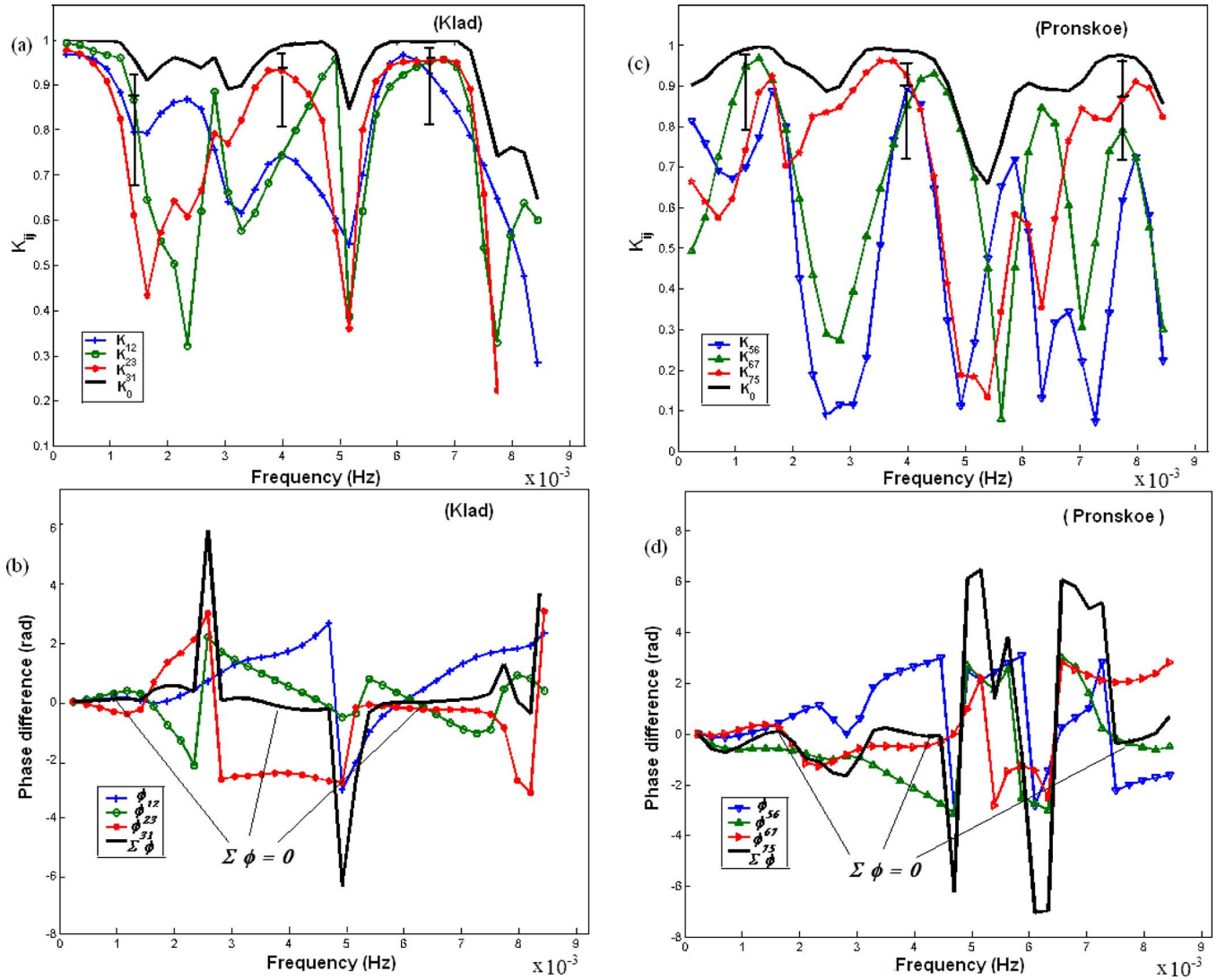


FIG. 4. (Color online) The coherences [(a) and (c)] and phase spectra [(b) and (d)] for the time variations in the travel time interval $T(t)$, obtained at 30 m triangles in Pronskoe (P5, P6, and P7) and Klad (R1, R2, and R3). There are common frequency intervals ~1.5–2, 3.5–4, and 6.5–8 mHz, for which a multicoherence function K_0 [black curve in Figs. 4(a) and 4(c) reaches local maxima along with the condition $\Sigma\varphi \approx 0$ [indicated in Figs. 4(b) and 4(d)].

IGWs, which can maintain a high horizontal coherence over the distances of a few wavelengths. At higher frequencies, the wavelike fluctuations (found for a 30 m triangle in Fig. 4) lost their coherence over the distance between Pronskoe and Klad. The use of triangles of different sizes allowed us to detect the coherent structures of different scales and to estimate these scales.

B. Effective sound speed fluctuations retrieved from travel time measurements

For small variations in the effective sound speed $\delta c_{\text{eff}}(t)$, $(|\delta c_{\text{eff}}/c_{\text{eff}}| \ll 1)$, one can express the temporal variations in the time interval $\delta T(t) = T(t) - T(t_0)$ between pulse arrivals [for instance, D and A, Fig. 2(b)] at one microphone as follows:

$$\delta T(t) = \int_{\Gamma_A} \delta c_{\text{eff}} c_{\text{eff},0}^{-2} ds, \tag{2}$$

where t_0 is some initial time moment, Γ_A is the sound ray path over which the integration is performed, $c_{\text{eff},0}$ is an initial field of the effective sound speed at $t = t_0$, and the second-order terms, $O(\delta c_{\text{eff}}^2/c_{\text{eff}}^2)$, are neglected.

If the field $c_{\text{eff},0}$ is known, then the small perturbations $\delta c_{\text{eff}}(t)$ are the solutions of the linear system of Eq. (2) derived for the selected ray path. This system is significantly simplified if $c_{\text{eff},0}$ varies mainly in the vertical direction z. This was the case for our experiment, during which the ABL was stably stratified, and the distances (about 2–3 km) between a source and the receivers were less than the characteristic horizontal scales of the variations in $c_{\text{eff},0}$. In this case $c_{\text{eff},0}$ can be considered as a function of z only, and can be obtained from the measured wind speed and temperature profiles or by using analytic profiles, such as exponential or Epstein profiles (Chunchuzov *et al.*, 1990). In the latter case, the parameters of the analytic profile of $c_{\text{eff},0}(z)$ are chosen in such a way that the best match between calculated and measured time intervals for the pulse arrivals is obtained.

In Fig. 5(a), the profile of $c_{\text{eff},0}(z)$ is shown which was calculated using 5 min averaged wind speed and temperature profiles in the stable ABL. The profile was obtained between 23:35 and 23:40 LT on July 13, 2005, which was an initial

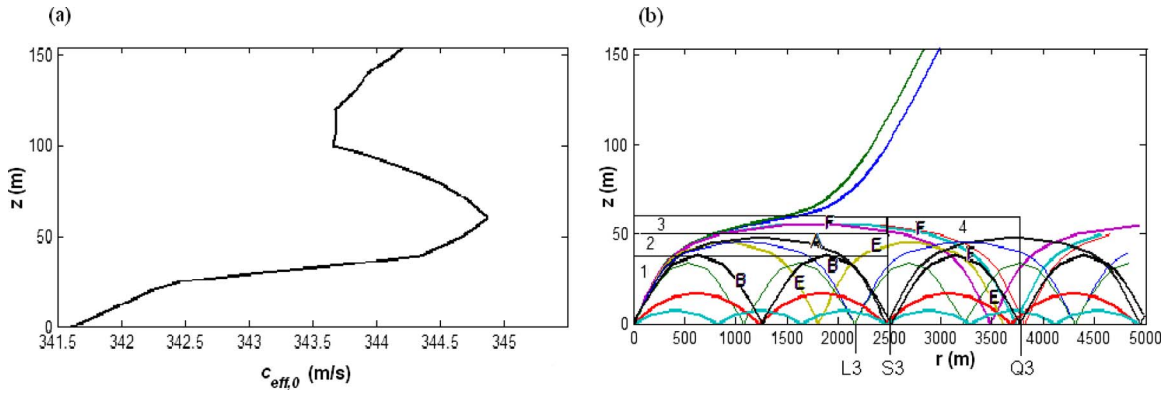


FIG. 5. (Color online) Initial profile, $c_{\mathrm{eff},0}(z)$, and the corresponding ray paths during period 23:35–23:40 LT of July 13, 2005. The selected ray paths A, B, E, and F connecting a source and the receivers S3 (r=2.5 km) and Q3 (r=3.6 km) are chosen to retrieve variations in c_{eff} in cells 1, 2, 3, and 4. The grazing angles for these ray paths are 6.690° (a), 6.530° (b), 7.010° (e), and 7.172° (f).

period of pulse sounding of the nocturnal ABL. The calculated ray trajectories [Fig. 5(b)] between the source and the receivers L3 (r=2157 m), S3 (r=2500 m), and Q3 (r=3620 m) have almost the same azimuth (between −13° and −15° relative to the northward direction). As in the case of July 11, 2005, shown in Fig. 3(b), the signal received by all nine receivers between 23:32 LT on July 13, 2005 and 00:36 LT on July 14, 2005 was composed of arrivals A, B, C, and of the latest arrival D that propagated horizontally along ground surface. For each of the receivers the corresponding ray paths [Fig. 5(b)] were found, for which the calculated time intervals between arrivals most closely matched with the measured time intervals between the same arrivals.

The time series $T(t)$ for the measured time intervals between arrivals A and D, and between B and D, obtained on July 13, 2005 by a 30 m triangle array at r=2.5 km (the receivers S1, S2, S3 in Fig. 1) are shown in Figs. 6(a) and 6(b). The same time series, but obtained at r=3.6 km with 30 m triangle arrays Q1, Q2, and Q3 [see Fig. 1(a)], are shown in Figs. 6(c) and 6(d) along with the time series $T(t)$ obtained by receiver L3 at r=2157 m.

To solve the system of Eq. (2), a spatial region containing ray trajectories within a narrow range of azimuths was divided into N spatial cells to present Eq. (2) in the following discrete form:

$$\delta T_m = \sum_{n=1}^{N} A_{mn} C_n, \quad m = 1,2,\ldots,M, \tag{3}$$

where n is the number prescribed to a given cell (n=1,2,…,N), m is the number corresponding to the ray path connecting a given source-receiver pair, C_n is the fluctuation $\delta c_{eff}(t)$ averaged over the volume of the nth cell, A_{mn} $=-\int_{\Gamma mn} ds(c_{eff,0}^{-2})$ is the integral taken over the portion of the mth ray trajectory Γ_{mn} within the nth cell.

Among the pulse arrivals detected by the receiver P6 at r=2480 m, only arrivals A and B do not overlap in the sense that the time interval between their peaks was greater than the time duration of each arrival. These arrivals correspond to ray paths A and B having one and two ray turning points, respectively [Fig. 5(b)]. For the receiver Q3 at r=3620 m, the ray paths corresponding to the same arrivals (A and B) are labeled by E and F. Four cells (N=4) were selected which cross ray trajectories A, B, E, and F, as shown in Fig. 5(b). For these cells, the system of Eq. (3) was solved with respect to C_n in which the numbers m=1, 2, 3, and 4 corresponded to ray paths B, A, F, and E, respectively.

Let us designate the temporal fluctuation of the time interval T_1 between arrival B along ray 1 and the last arrival D as $\delta T_1(t)$ [Fig. 6(b)]. Similarly, designate the fluctuations of the time intervals (with respect to arrival D) for rays 2, 3, and 4 as $\delta T_2(t)$, $\delta T_3(t)$, and $\delta T_4(t)$.

For the initial profile of $c_{\mathrm{eff},0}(z)$ shown in Fig. 5(a), the travel times $t_{mn}=\int_{\Gamma mn} ds(c_{\mathrm{eff},0}^{-1})$ and the coefficients A_{mn} for the portions of ray paths Γ_{mn} within the corresponding cells were calculated. For ray B (m=1), which is totally within the cell with n=1, the calculated nonzero coefficient is A_{11} =0.0212 s^2/m. Other nonzero elements of the matrix A_{mn} are as follows: A_{21}=0.0085 s^2/m and A_{22}=0.0124 s^2/m for ray A (m=2); A_{41}=0.012 s^2/m, A_{42}=0.0092 s^2/m, and A_{44} =0.0084 s^2/m for ray E (m=4); and A_{31}=0.00373 s^2/m, A_{32}=0.0034 s^2/m, A_{33}=0.016 s^2/m, and A_{34}=0.007 s^2/m for ray F (m=3).

As a result, the average fluctuations $C_1(t)$, $C_2(t)$, $C_3(t)$, and $C_4(t)$ were obtained which allowed us to calculate the variations in the effective sound speed $c_{\mathrm{eff}}(z_n,t)$ in each cell: $c_{\mathrm{eff}}(z_n,t)=c_{\mathrm{eff},0}(z_n)+C_n(t)$, n=1,2,3,4, where z_n is the altitude of the geometric center of the nth cell. It was assumed that the fluctuations of c_{eff} averaged over the cell's volumes and over the portions of the ray paths belonging to these cells are close to each other. Since the values of $c_{\mathrm{eff}}(z_n,t)$ are sensitive to the errors of measurement of δT_m and to the errors in calculating the coefficients A_{mn}, we compared the calculated travel time intervals for the obtained profile $c_{\mathrm{eff}}(z_n,t)$ at different time moments and the measured intervals at the same moments. If the absolute difference between the measured and calculated values of the interval T was greater than the accuracy of its measurement (about 2 ms), then the retrieved profile $c_{\mathrm{eff}}(z_n,t)$ was slightly perturbed to reach (after an iteration procedure) a closest match between the measured and calculated intervals T. After applying such an iteration procedure, the time series of the fluctuations $C_n(t)$ was obtained [see Fig. 7(a)]. These fluctuations were compared to the fluctuations of the northward component of

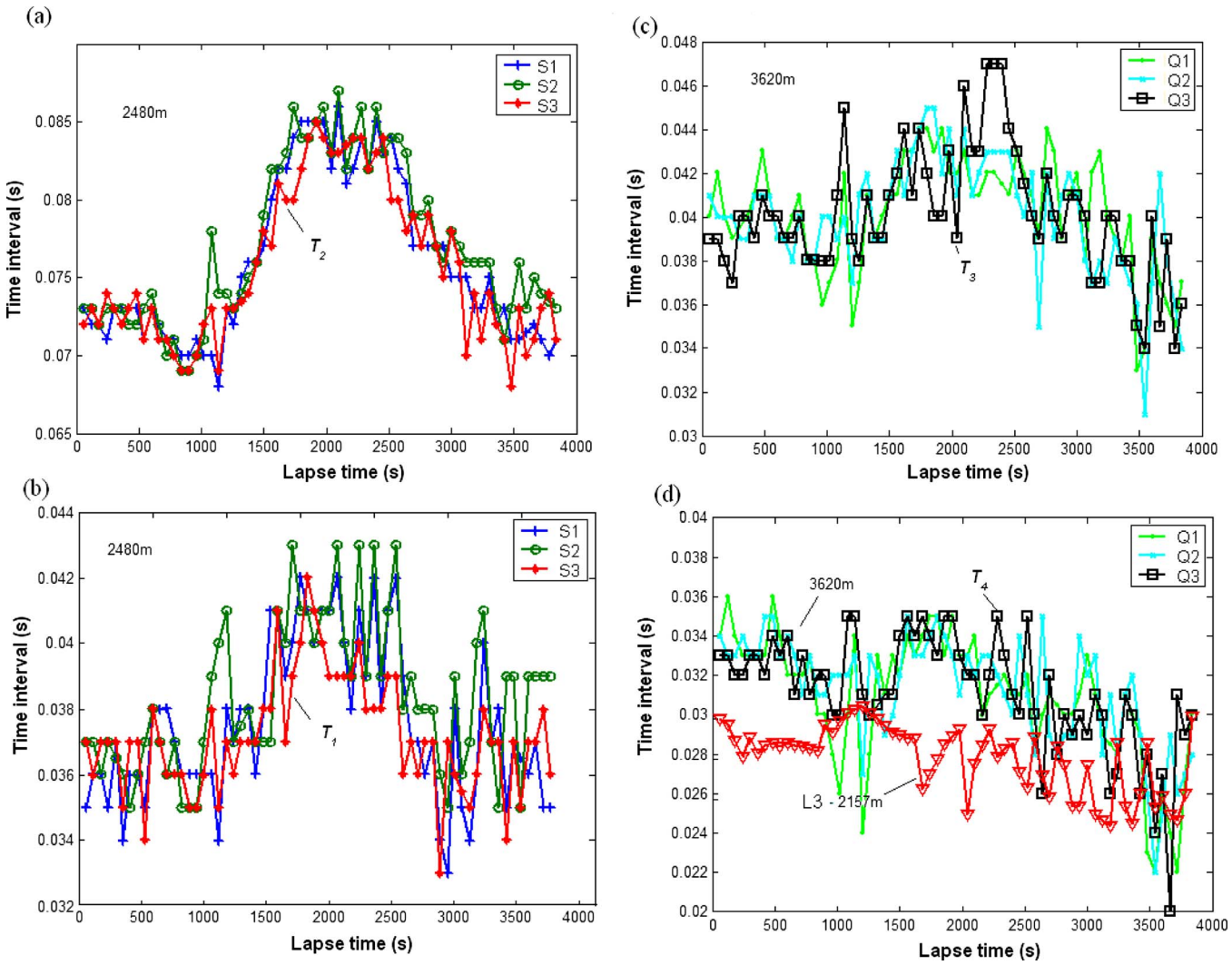


FIG. 6. (Color online) The time series of the interval $T(t)$ between arrivals A and D, and B and D (see Fig. 2) obtained on July 13, 2005 from the triangle arrays: (a) S1-S2-S3 ($r \sim 2.5$ km, intervals A–D). (b) S1-S2-S3 (intervals B–D). (c) Q1-Q2-Q3 ($r \sim 3.6$ km, intervals A–D); (d) Q1-Q2-Q3 (intervals B–D). Also shown is the time series at $r \sim 2.2$ km (receiver L1).

the wind speed $u'(t)$ measured by the sonic anemometer at a height of 56 m above ground and at a horizontal distance of about 4.5 km from the source [Fig. 7(b)].

Note that the fluctuations $C_3(t)$ are retrieved near the turning point of ray path F which is within the same height range (between 50 and 56 m) as the anemometer, but at a horizontal distance of about 2.8 km. Despite this distance between the turning point and the anemometer, the retrieved fluctuations $C_3(t)$ quite well reproduce the increase in the wind speed $u'(t)$ observed during the time period between 1500 and 2500 s, thereby showing a high horizontal coherence between $C_3(t)$ and $u'(t)$. However, the fluctuations $C_3(t)$, being averaged over some volume of space, reached peak values almost 1.5 times less than those of the wind speed fluctuations $u'(t)$ measured by the anemometer at a certain point. Despite such difference in the amplitudes of the fluctuations $C_3(t)$ and $u'(t)$, the obtained high coherence between them indicates the capability of the acoustic tomography in the retrieval of these fluctuations in different regions of the ABL. This capability is also confirmed by a good agreement between the fluctuations $C_1(t)$ and $C_2(t)$ retrieved for July 11, 2005 [Fig. 7(c)] in cells 1 and 2 [shown in Fig. 3(c)] and the wind fluctuations $u'(t)$ measured at a height of 56 m [Fig. 7(d)]. In the case of July 13, 2005, considered above, the number of the resolved rays (without overlapped arrivals) happened to be equal to the number of the selected cells. The same problem given by Eq. (3) can be also stated for the more general case $M > N$ which is less sensitive to the measurement errors. For this case, the plausibility of the retrieved data will be tested in a future study.

The relative variations in the effective sound speed can be presented as $\delta c_{\text{eff}}/c_{\text{eff},0} = \delta T'/(2T_0) + V_e/c_0$, where V_e is the fluctuation of the wind velocity projection on a radius vector from a source to a receiver, $\delta T'$ are the fluctuations of the temperature T', and c_0 and T_0 are the mean (over 60 min) sound speed and temperature near ground, respectively. Using the time series of the temperature and wind components, measured by the sonic anemometer/thermometer at 56 m height, the rms values (over the measurement period) of the temperature fluctuations $(\langle \delta T'^2/(2T_0)^2 \rangle)^{1/2}$ and wind speed fluctuations $(\langle V_e^2/c_0^2 \rangle)^{1/2}$ were estimated. For the night hours

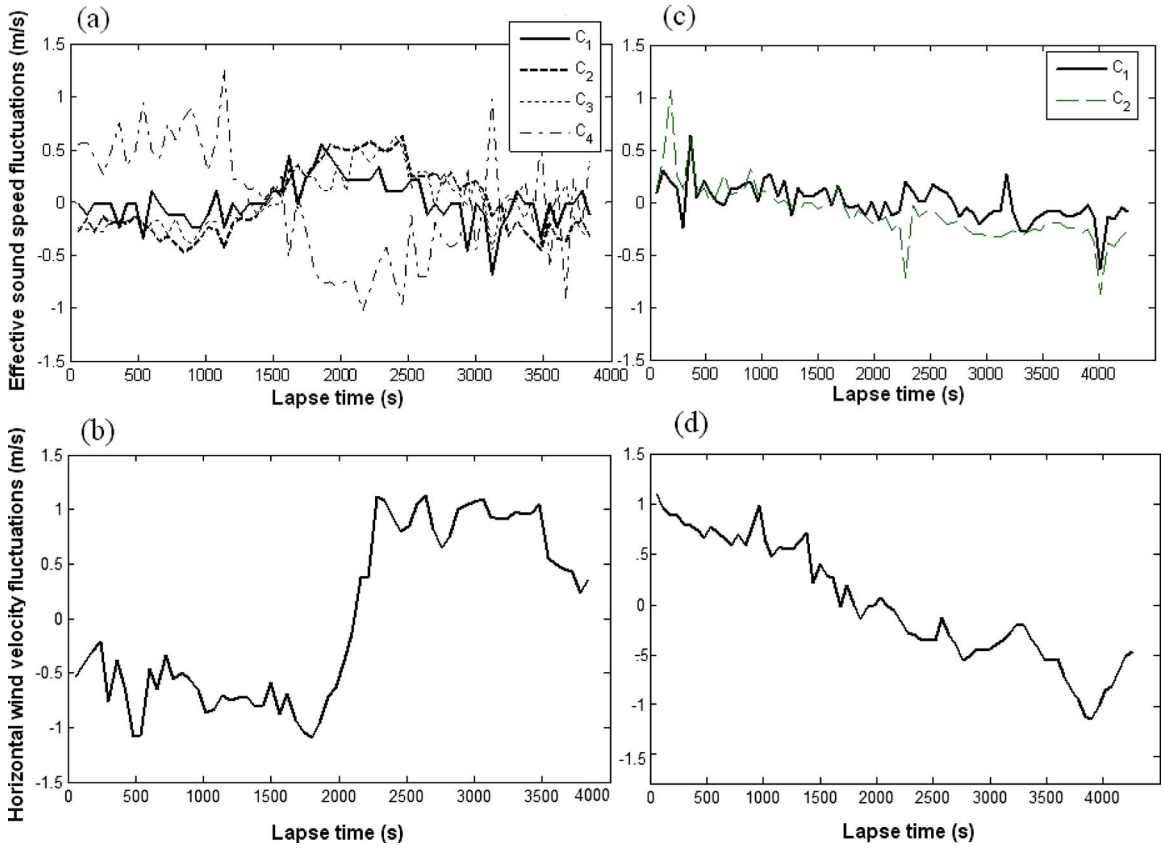


FIG. 7. (Color online) Retrieved effective sound speed fluctuations $C_i(t)$ (i is the cell's number) and the horizontal wind speed fluctuations $u'(t)$ measured by sonic at 56 m height. (a) $C_i(t)$ for July 13/14, 2005, 23:32–00:35 LT. The cells with $i=1,2,3,4$ are shown in Fig. 5. (b) $u'(t)$ for July 13/14, 2005, 23:32–00:35 LT. (c) $C_i(t)$ for July 11/12, 2005, 23:25–00:36 LT, receiver R2. The cells with $i=1,2,3$ are shown in Fig. 3. (d) $u'(t)$ for July 11/12, 2005, 23:25–00:36 LT.

of July 11, 2005 it was obtained that the contribution of the wind fluctuations $(\langle V_e^2/c_0^2\rangle)^{1/2}$ to the effective sound speed variations was noticeably greater than those of the temperature fluctuations$(\langle \delta T'^2/(2T_0)^2\rangle)^{1/2}$. For the height interval of $100<z<300$ m, the wind speed and temperature contributions were estimated by comparing the time series of the 5 min averaged wind speed variations measured by SODAR and the time series of the 5 min averaged temperature variations measured by temperature profiler during the same time period. The obtained wind contribution $(\langle V_e^2/c_0^2\rangle)^{1/2}$ at different heights was from two to five times greater than the temperature contribution $(\langle \delta T'^2/(2T_0)^2\rangle)^{1/2}$. These estimates are consistent with the theoretical estimates of the same contributions based on the energy spectrum of IGWs with random amplitudes and phases (Chunchuzov, 2003). The theory shows that the value of $(\langle V_e^2/c_0^2\rangle)^{1/2}$ induced by IGWs is about three times greater than $(\langle \delta T'^2/(2T_0)^2\rangle)^{1/2}$, so under stable stratification of the ABL, a main contribution to the effective sound speed fluctuations comes from the wind fluctuations.

Furthermore, an agreement was found between the theoretical and experimental estimates of the rms values of the azimuth fluctuations of an acoustic signal wave front which at $r=2-3$ km from a source were in the range $0.5°-1.0°$ (Kniffka *et al.*, 2006). This is an estimate of the error in determining of the azimuth of an acoustic source which should be taken into account when solving a problem of localization of different sources under stable stratification of the ABL. The IGWs, filtered in our experiments by a coherence analysis, significantly contribute to this error.

C. Retrieved variations in the effective sound speed gradient

The time series $c_{\text{eff}}(z_n,t)$ retrieved in the cells with different heights z_n were used to calculate the temporal variations of the vertical gradients $\Delta c_{\text{eff}}(z_n,t)/\Delta z_n$ between the altitudes of the centers of neighboring cells $(\Delta z_n=z_n-z_{n-1})$ and to estimate the rms values $(\langle(\Delta c_{\text{eff}}(z_n,t)/\Delta z_n)^2\rangle)^{1/2}$ over the observational time period (about 1 h in our case). Based on these data, the vertical wind shear can be derived. Monitoring of the wind shears within some spatial volume of the ABL is needed, particularly, for the aircraft navigation near airports, where an operative forecast of the periods of strong enhancement of the wind shears in the lower atmosphere is very important for aviation safety (Stoll, 1991).

Furthermore, using the obtained data one can estimate the horizontal coherence at a distance of about 2.8 km (Fig. 8) between the time variations in the retrieved gradients $\Delta c_{\text{eff}}(z_n,t)/\Delta z_n$ and the measured vertical gradients $\Delta u'/\Delta z$ of the wind component $u'(t)$ obtained from the anemometer data at the heights of 6 and 56 m. Note that such a rough estimate of the wind gradient from sonic measurements contains no information about the vertical variations in the local wind gradient values associated with the wind inhomogeneities whose vertical scales are less than 50 m. Nevertheless,

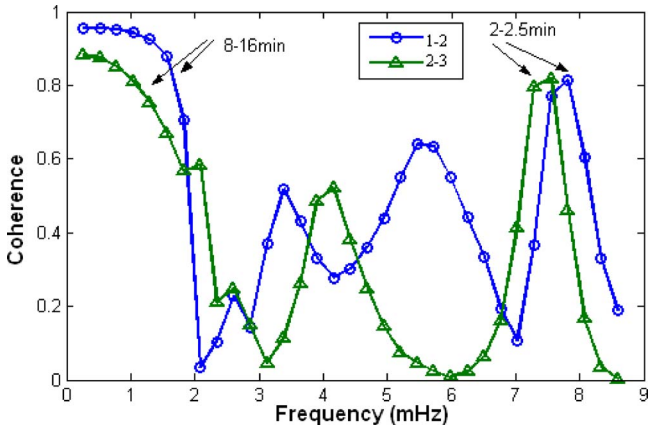


FIG. 8. (Color online) Coherence between the time fluctuations of the retrieved gradients $\Delta C_i(z,t)/\Delta z$ and the measured wind gradients $\Delta u'(z,t)/\Delta z$ (July 13/14, 2005, 23:32–00:35 LT). The coherences for the gradients between cells 1 and 2, and 2 and 3 are designated as (1–2) and (2–3), respectively.

the time series $\Delta u'(t)/\Delta z$ provided us with the valuable information about low-frequency temporal variations in the wind gradients in the surface layer needed for calculating their horizontal coherence. This coherence, as seen in Fig. 8, significantly increases up to the values (0.8–1) for the periods of 8–16 min and 2–2.5 min. The periods of 8–16 min are typical to the IGWs in the troposphere, whereas shorter periods, less than 8 min, are often in the ducted gravity waves (Chunchuzov *et al.*, 2005) and in the Kelvin–Helmholtz billows (Blumen *et al.*, 2001) existing in the surface inversion layer.

Thus, the coherent structures modulate the effective sound speed and wind gradients with a discrete set of temporal and spatial periods. The estimated rms values of the average gradients $(\langle(\Delta c_{\text{eff}}(z_n,t)/\Delta z_n)\rangle)^{1/2}$ between the first $(n=1)$ and second $(n=2)$ cells are about 0.065 s^{-1}, whereas between the second and the third cells they are about 0.02 s^{-1}. The obtained gradient $\Delta c_{\text{eff}}(z_n,t)/\Delta z_n$ depends on both the wind speed and temperature gradients whose mesoscale variations cause spatial and time perturbations of the local values of the Richardson number Ri and the buoyancy frequency N thereby modulating the stability of the ABL (Gossard and Hooke, 1975). Therefore, acoustic monitoring of the gradients $\Delta c_{\text{eff}}(z_n,t)/\Delta z_n$ in different regions of the ABL allows one to obtain information about the stability of these regions.

The mesoscale variations in the retrieved gradients $\Delta c_{\text{eff}}(z_n,t)/\Delta z_n$ were compared to the mesoscale variations in the local vertical turbulent fluxes of momentum $mf \equiv \langle w'u' \rangle$ and sensible heat $hf \equiv \langle \theta' u' \rangle$ whose time series were obtained from sonic anemometer-thermometer measurements at the 56 m mast. Here, $\theta'(t)$, $w'(t)$, and $u'(t)$ are the time series of the perturbations of the potential temperature relative to its mean value, and vertical and horizontal (northward) velocity components, respectively. The sampling period of the measurements of $\theta'(t)$, $w'(t)$, and $u'(t)$ was about 0.1 s, whereas the averaging time period for the estimates of the correlations $\langle w'u' \rangle$ and $\langle \theta' u' \rangle$ was about 1 min.

For July 11, 2005, the mesoscale variations in the turbulent momentum flux (mf) at 56 m are shown in Fig. 9(a) along with the time series of the vertical gradient of wind velocity $\Delta u'/\Delta z$, between 6 and 56 m. For the atmospheric layers within the height ranges 180–200, 200–220, and 220–240 m, the time series of the 5 min averaged vertical gradient $\Delta u'/\Delta z$ were derived from SODAR data and they are shown in Fig. 9(b). One can find in Figs. 9(a) and 9(b) the enhancement of the wind gradient $|\Delta u'/\Delta z|$ between 23:30 and 23:45 LT along with the enhancement of turbulent momentum flux (mf) [Fig. 9(a)]. During this time interval, the vertical wind shear defined by $\Delta V/\Delta z=[(\Delta u'/\Delta z)^2+(\Delta v'/\Delta z)^2]^{1/2}$, where v' is the eastward wind component, reached its critical values ($\sim$0.4 s^{-1}), for which the local Richardson number $\text{Ri}=N^2/(\Delta V/\Delta z)^2$, where N is the Brunt–Väisälä (BV) frequency, dropped below 1/4. It is expected that this condition should result in the shear instability of the wind field and generation of small-scale turbulence (Gossard and Hooke, 1975; Danilov and Chunchuzov, 1992). Indeed, some enhancement of the intensity of SODAR backscattered echo signal between the altitudes 180 and 260 m was observed during the same time interval with the wind shear enhancement (indicated by an arrow in [Fig. 9(c)]).

Similar episodes of simultaneous changes in the wind shear and turbulent fluxes mf and hf were also observed during the morning hours of July 13, 2005 between 7:35 and 7:50 LT, when a sharp increase in the wind shear [between 35 and 50 min in Fig. 9(d), dashed line, sonic] was accompanied by the sharp changes in the vertical turbulent fluxes mf and hf [between 35 and 50 min in Fig. 9(e), sonic 56 m] as well as in the intensity of the SODAR backscattered echo signal (not shown here). During the same time period, the simultaneous changes in the pulse travel time interval between arrivals $T(t)$ [Fig. 9(d), solid line] and the wind shear were observed. Note that an upper ray turning point for arrival A was at a height of 100 m and at a distance $r=1240$ m, whereas the wind shear was estimated at $r=4.5$ km from the source. Despite the large distance between the measurement points (more than 3 km) for the wind shear and travel time interval, they show a similar behavior in time [Fig. 9(d)]. This confirms a strong influence of the wind gradient variations in both sound travel time and local turbulent fluxes near ground.

III. GRAVITY WAVE MODEL

The cross spectra between fluctuations $T(t)$ [Fig. 10(a)], azimuth of propagation [Fig. 10(b)], turbulent fluxes mf and hf, retrieved effective sound speed $C_2(t)$ [Fig. 10(d)], and the autospectra of mf and hf, and of the turbulent drag coefficient C_d [Fig. 10(c)] have distinct peaks within almost the same frequency ranges: 2–3, 4–5, and 6–8 mHz. Thus the time variations in both signal parameters and turbulence parameters have several dominant periods under stable stratification of the ABL. One important consequence of this fact, which should be taken into account in wave-turbulence interaction models, is that the fluctuating parts of the wind shear, eddy diffusion coefficient, and of the turbulent fluxes

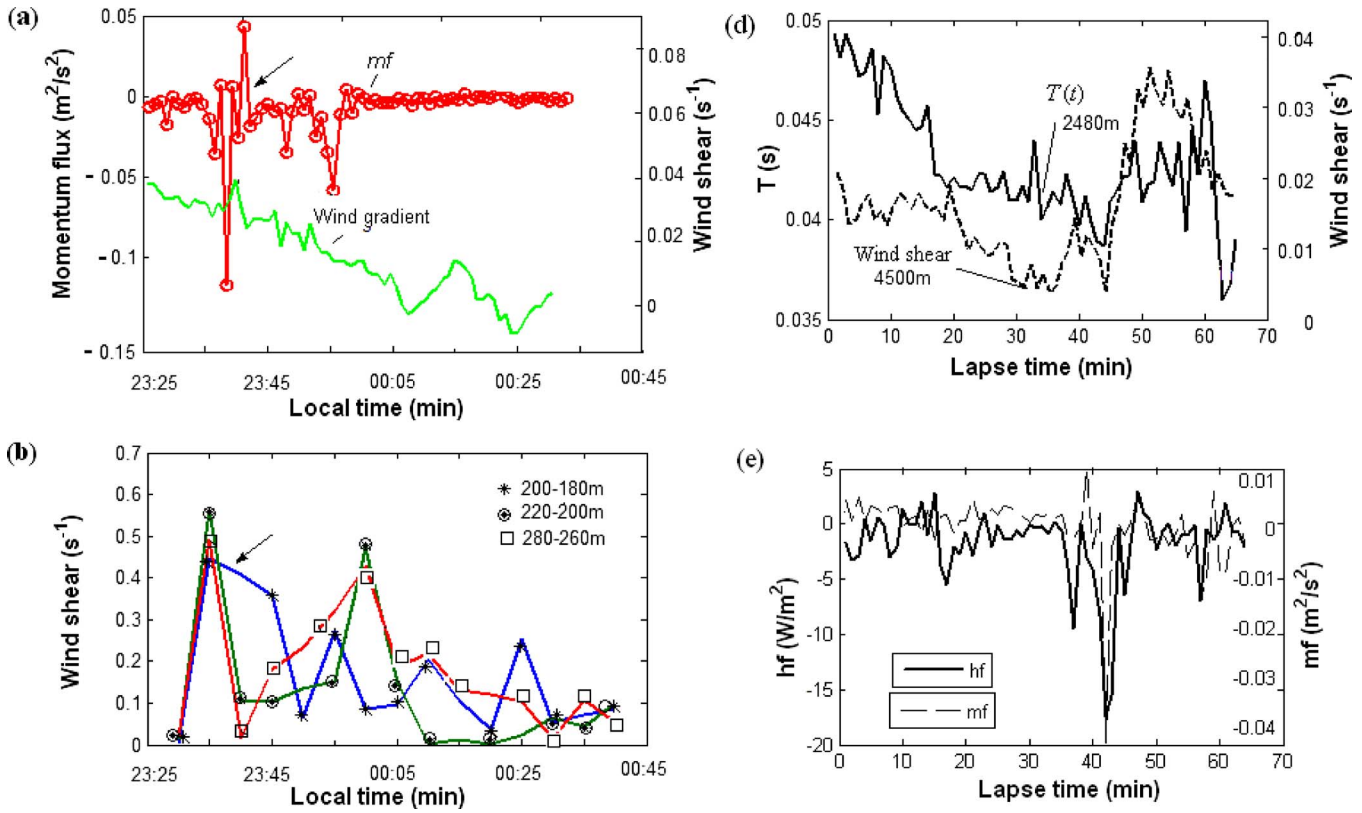


FIG. 9. (Color online) Episodes of sharp temporal changes in the wind gradient, $\Delta u'(z,t)/\Delta z$, acoustic travel time interval $T(t)$, and the vertical turbulent fluxes of momentum, mf, and sensible heat, hf, observed in the night of July 11/12, 2005 between 23:25 and 00:40 LT [(a)–(c)] and on July 13, 2005 after sunrise (07:00–08:10 LT) [(d)–(f)]. (a) 1 min averaged temporal variations in mf at $z=56$ m (sonic) and vertical gradient $\Delta u'(z,t)/\Delta z$ between 6 and 56 m (sonic). (b) 5 min averaged vertical gradient $\Delta u'(z,t)/\Delta z$ vs time for the atmospheric layers in the height ranges: 180–200, 200–220, 220–240 m (SODAR). The arrows indicate the enhancement in the wind speed gradients $|\Delta u'(z,t)/\Delta z|$ between 23:35 and 23:45 LT along with the enhancement of mf and the intensity of backscattered echo signal. (d) Variations in the vertical wind shear (sonic) and the travel time interval $T(t)$. (e) Turbulent fluxes hf and mf.

are caused by internal wave-packets with narrow frequency bands rather than by the monochromatic IGWs of a certain frequency.

Note that some of the dominant periods indicated above were also found in the wind speed fluctuations in a convective ABL (Petenko and Bezverkhnii, 1999). In our experiment, a comparison of the auto spectra for the daytime and nighttime series of the azimuth fluctuations showed the existence of the same spectral maximum at a frequency of 4–5 mHz under both unstable and stable stratifications of the ABL. Therefore, it is assumed that the existence of the dominant periods in the observed fluctuations is associated mostly with the stable stratification of the tropospheric layers that lie above the ABL. The mechanism of coupling between convective eddy structures in the ABL and gravity waves in the troposphere was modeled, for example, by Clark *et al.* (1986) and Sang (1991).

In the nocturnal ABL, the temperature inversion forms a wave duct for short-period gravity waves near ground, since the BV frequency $N(z)$ reaches a maximum near ground. This is readily identifiable from the profiles of $N(z)$ calculated by using the temperature profiler data of July 11, 2005 between 23:25 and 00:35 LT [Fig. 11(a)]. In the lower 100 m layer, $N^2(z)$ reaches a maximum of 0.0015 or 0.0035 rad^2/s^2 at $z=0$, so that the shortest periods of ducted gravity waves in the inversion layer are of about 2.5–3 min. There are different sources that excite IGWs in the lower atmosphere such as jet streams, wind shear instabilities, meteorological fronts, topography, and others. But whatever source is active during the observational period, a near-ground wave duct filters only a certain range of wave periods determined by the temperature and wind stratification near ground. This range is limited by the maximum and minimum values of the near-ground BV-frequency profile $N(z)$; therefore, the typical periods of ducted IGWs near the ground range from about 8–11 to 2–4 min (Chunchuzov *et al.*, 2005).

The analysis of the $N(z)$ profiles in the troposphere, obtained from the radiosonde data of July 11–12, 2005, revealed the existence of a second wave duct for IGWs in the troposphere above the ABL where $N^2(z)$ reached a local maximum [Fig. 11(b)]. As seen from Fig. 11(b), the layer with the maximum stability between altitudes of 2 and 3.5 km occurred both in the day time and night time hours. A similar tropospheric layer of high static stability between altitudes of 2 and 6 km was also found in an earlier experiment carried out during August 2002 and described by Chunchuzov *et al.* (2005).

The BV-frequency profiles for both days, July 11, 2005 and August 13, 2002, were approximated by steplike functions of z to calculate a ducted field of IGWs. One such approximation for August 13, 2002 with the chosen values of N in each atmospheric layer (designated by N_1, N_2, N_3, and N_4) is shown in Fig. 11(c). For the selected four-layer model of the atmosphere with a constant wind speed, the linear

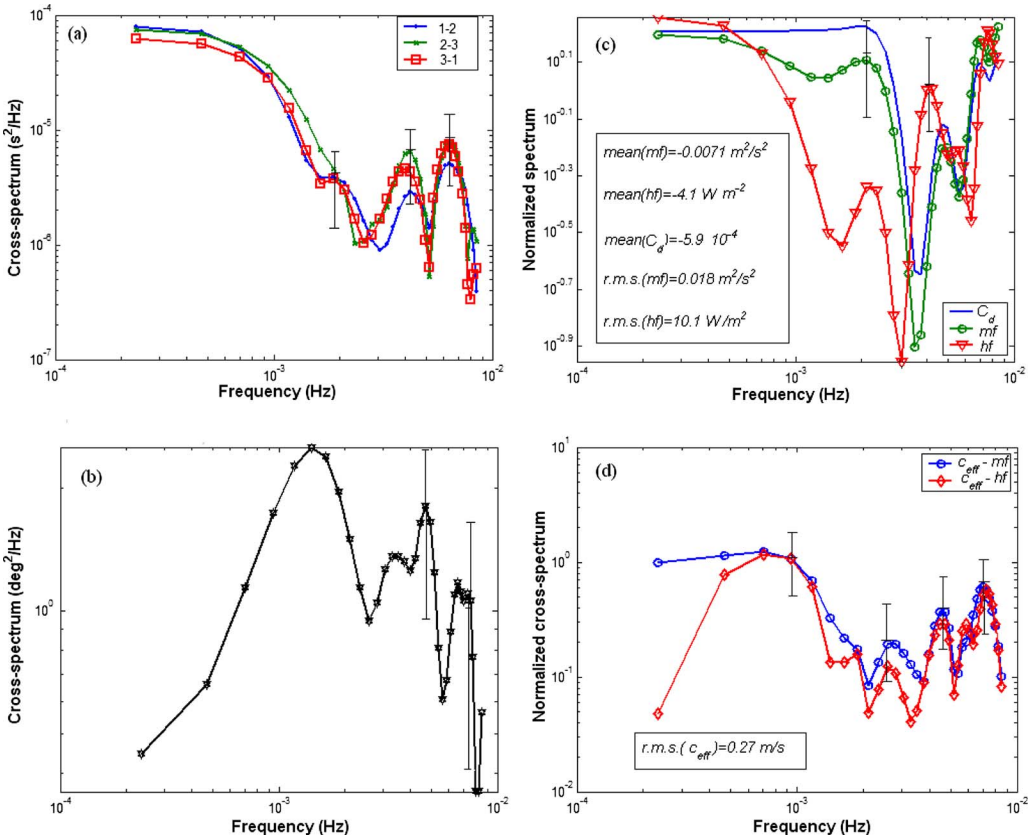


FIG. 10. (Color online) Dominant periods in the variations in the time interval $T(t)$, the azimuth of propagation of the acoustic signal, and the parameters of turbulence observed on July 11/12, 2005 (23:25–00:36 LT). (a) Cross spectra between fluctuations $T(t)$ measured by a 30 m triangle array R1-R2-R3 (Klad, see Fig. 1). (b) Cross spectrum between fluctuations of the azimuth of propagation of the signal's wave front measured at Pronskoe and Klad (3 km apart). (c) The autospectra of the variations in the vertical turbulent fluxes mf and hf and in the turbulent drag coefficient C_d normalized by the corresponding variances of the fluctuations (their mean and rms values are shown). (d) Normalized cross spectra between variations in mf, hf, and retrieved variations $C_2(t)$ [Fig. 3(c)].

vertical velocity field of ducted gravity modes was calculated by Chunchuzov *et al.* (2007) similar to those calculated by Gossard and Hooke (1975) for a three-layer model. The height dependence of the vertical velocity amplitude $W(z)$ of one of the calculated ducted modes is shown in Fig. 12. $W(z)$ oscillates with growing height z within the third layer with $N=N_3$, and outside this layer it decays exponentially downward to the ground surface. Despite the rapid downward decay of $W(z)$, its magnitude can remain large enough near ground to excite a gravity wave field within a near-ground surface wave duct whose maximum BV-frequency value is $N_1>N_2$. This takes place, when the frequency ω of the ducted wave mode approaches its lowest (cutoff) value N_2, because in this case the rate of the amplitude decay downward to the ground essentially decreases. For this particular frequency ($\omega \approx N_2$), the wave field penetrates from the troposphere to the near-ground surface layer and modulates the wind field near ground.

The phase speed of the wavelike fluctuations with the frequencies $f \approx 0.001-0.0015$ Hz was estimated to be about 2.7 m/s (see also Chunchuzov *et al.*, 2005) for the case shown in Fig. 12. These frequencies are close to the minimum values on $N(z)$ in the troposphere, $N_2/(2\pi)=0.001$ Hz (corresponding period is about 16 min); therefore the frequency $f=0.00101$ Hz and a modal phase speed of 2.7 m/s were chosen to calculate $W(z)$. These parameters select a certain mode with $f=0.00101$ Hz shown in Fig. 12, where $W(z)$ is normalized by the value of $W(2$ km). The normalized amplitude $W(z)/W(2$ km) for $f=0.0016$ Hz which is close to the maximum BV frequency, $N_3/(2\pi)=0.002$ Hz (corresponding period is about 8 min), in the layer 2 km $\leq z \leq 6$ km is also shown in Fig. 12. Within this layer $W(z)$ oscillates over z, but exponentially decays almost to 0 (since $W(0.2$ km$)/W(2$ km$) \sim 10^{-3}$), when z decreases from $z=2$ km to $z=200$ m. When f approaches N_2, which is a minimum of $N(z)$, the rate of the amplitude decay from $z=2$ km to the ground significantly decreases as compared to the case $f=0.0016$ Hz. As a result, the amplitude $W(z)$ at $z=200$ m becomes only one order less than its value at $z=2$ km. For example, if $W(2km) \approx 0.5-1$ m/s, which are typical values for such altitudes, then at $z=200$ m the amplitude $W(z) \approx 0.05-0.1$ m/s. The latter is close to the observed amplitudes of the vertical velocity oscillations near ground under stable stratification of the ABL.

Thus, even in the case of a broadband frequency spectrum of IGWs excited by their sources, the tropospheric wave duct filters only those spectral component of the spectrum which is close to the lowest frequency of the wave duct N_2. A strong nonlinearity of the gravity waves with the observed low horizontal phase speeds can lead to the nonresonant wave-wave interactions (Chunchuzov *et al.*, 2003).

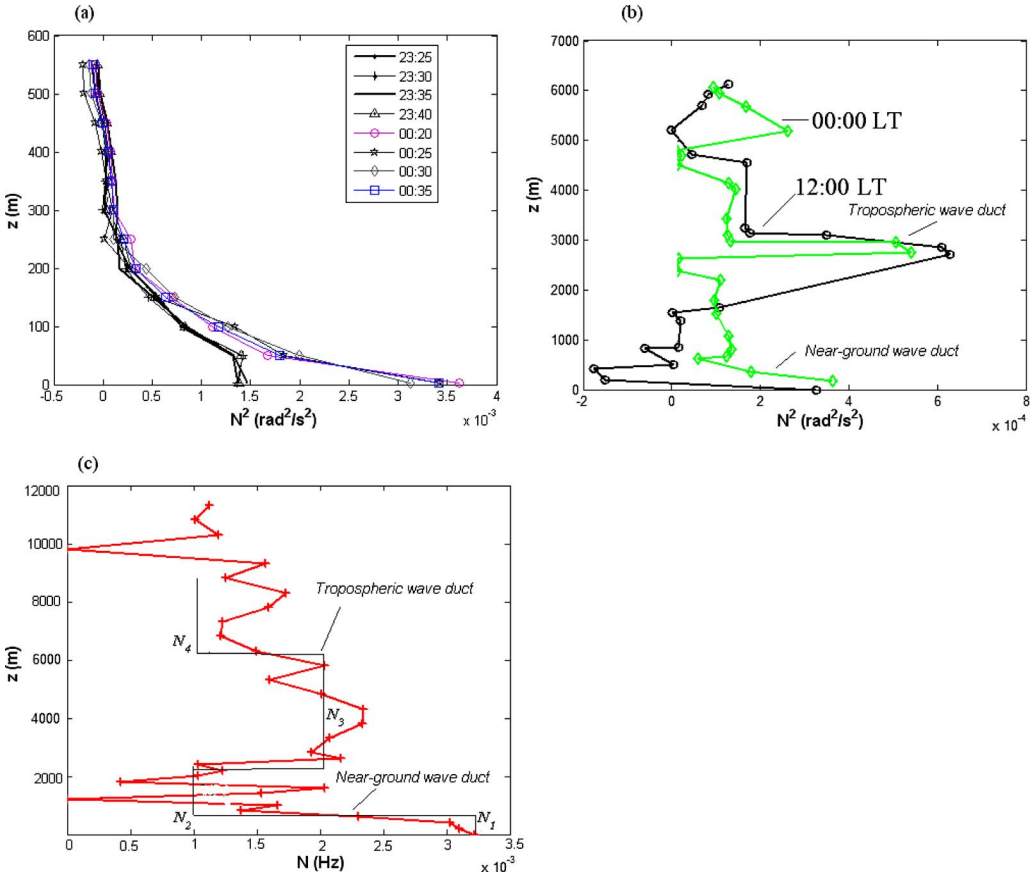


FIG. 11. (Color online) BV frequency profiles $N(z)$ during measurement periods. (a) 5 min averaged profiles of $N^2(z)$ in the ABL obtained on July 11/12, 2005 between 23:25 and 00:35 LT. (b) Day-time (12:00 LT) and night-time (24:00 LT) profiles of $N^2(z)$ in the troposphere obtained on July 11, 2005. (c) The profile $N(z)$ obtained on August 13, 2002 at 23:00 LT. The profile was approximated by a steplike function, as shown in Fig. 11(c).

These interactions are caused by advective nonlinearity of the Eulerian fluid motion equations described by the term $(\mathbf{V}\nabla)\mathbf{V}$ that becomes comparable to the linear term $\partial\mathbf{V}/\partial t$. The nonlinearity leads, particularly, to the generation of the harmonics with frequencies $2N_2 \sim 0.002$ Hz and $3N_2 \sim 0.003$ Hz, and of higher-order harmonics. At the same time, the nonlinearity can cause an interaction between the tropospheric wave fields, penetrating downward with $\omega \sim N_2$, with the short-period ducted waves in the near-ground wave duct. As a result, the harmonics of the tropospheric wave with $\omega \sim N_2$ along with those generated due to nonresonant interaction between tropospheric and near-ground ducted gravity waves will dominate in the variations in the wind field. This is a possible coupling mechanism between tropospheric and near-ground IGWs that can affect a short-period variability of the wind field, turbulence, and acoustic signal parameters near ground in the stably stratified ABL.

IV. CONCLUSIONS

One method of acoustic tomography has been applied here to the lower atmosphere for studying the influence of IGWs and other organized structures on the parameters of acoustic signals (travel time, duration, and arrival angles of the sound impulses) propagating along ground surface and their statistical characteristics. Based on the mesoscale fluctuations of acoustic travel time differences between signal arrivals (periods 1 min–1 h), measured by a net of distant receivers, temporal fluctuations of the effective sound speed averaged over different ray paths connecting a source and various receivers were retrieved. A major contribution to the derived fluctuations was found to be caused by wind speed fluctuations.

The calculated coherences between effective sound speed variations, retrieved near different ray turning points, showed an existence of wavelike fluctuations, which are typical for IGWs and eddy structures generated due to different types of the IGW's instabilities. The wavelike character of these fluctuations was confirmed by the fact that the sum of phase differences between the points of triangle arrays of different sizes was about zero for the periods, for which the peaks of coherences have been observed. A set of dominant periods was found both in the variations in acoustic signal parameters (travel time interval, duration, and azimuth of propagation of the signal's wave front) and of parameters of turbulence. The turbulent fluxes of momentum and sensible heat are modulated in time and space by spatially extended and highly coherent structures, such as IGWs and different types of eddy structures.

A number of episodes were observed, when the enhancement in the vertical gradients of wind speed up to their critical values (for which a Richardson number drops below 1/4) was accompanied by an enhancement of the intensity of turbulence. These episodes indicate that the local instabilities of

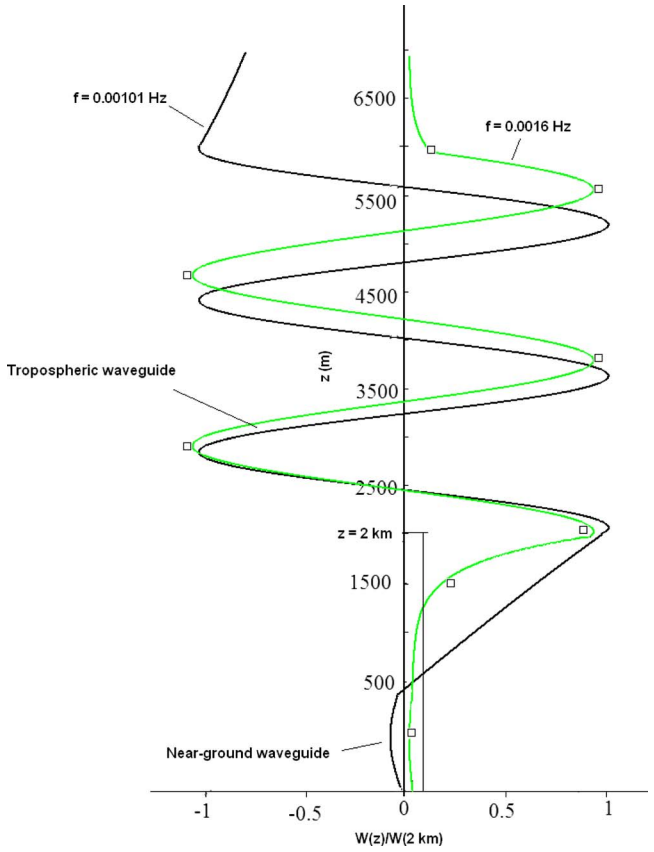


FIG. 12. (Color online) The normalized amplitude $W(z)/W(2\ \mathrm{km})$ of the vertical velocity oscillations caused by ducted gravity wave mode in the four-layer model of the atmosphere [see Fig. 11(c)].

IGWs can be a possible source of the intermittency of small-scale turbulence under conditions of stable stratification of the ABL.

A possible origin of the discrete periods in the variations in the parameters of acoustic signals under stable stratification of the ABL can be associated with the ducted gravity waves in the troposphere, capable of penetrating downward to the ground at a certain (cutoff) frequency, and with their nonlinear interaction with the short-period gravity waves in the near-ground wave duct caused by temperature inversion.

ACKNOWLEDGMENTS

We would like to thank our colleagues, M. Barth, D. Daniel, A. Raabe (Institute of Meteorology, Leipzig), R. Kuznetsov, M. Kallistratova, T. Kashkarova, N. Kulichkova, and N. Elansky (Institute of Atmospheric Physics, Moscow) for their help in conducting experiments, data processing, and useful discussions of the results. We are also very thankful to E. N. Kadygrov for providing us with the temperature profiler data. This work was supported by Russian Foundation for Basic Research (RFBR) Grant Nos. 06-05-64229, 05-05-64973, and 06-05-64357, by DFG-RFBR Grant Nos. 03-05-04001 and 07-05–91555, and by DFG Grant Nos. 436 RUS 113/737/0-1 (international co-operation) and Ra 569/9-1 (research grant).

Arnold, K., Balogh, K., Ziemann, A., Barth, M., Raabe, A., and Daniel, D. (**2005**). "Determination of meteorological quantities and sound attenuation via acoustic tomography," Proceedings of the Forum Acusticum Budapest 2005, Extended Abstracts on CD-ROM.

Bendat, J. S., and Piersol, A. G. (**1967**). *Measurement and Analysis of Random Data* (Wiley, New York).

Blumen, W., Banta, R., Durns, S., Fritts, D. C., Newsom, R., Poulos, G. S., and Sun, J. (**2001**). "Turbulence statistics of a Kelvin–Helmholtz billow event observed in the nighttime boundary layer during the CASES-99 field program," Dyn. Atmos. Oceans **34**, 189–204.

Chimonas, G. (**1999**). "Steps, waves and turbulence in the stably stratified planetary boundary layer," Boundary-Layer Meteorol. **90**, 397–421.

Chimonas, G. (**2002**). "On internal gravity waves associated with the stable boundary layer," Boundary-Layer Meteorol. **102**, 139–155.

Chunchuzov, I. (**2003**). "Influence of internal gravity waves on sound propagation in the lower atmosphere," Meteorol. Atmos. Phys. **34**, 1–16.

Chunchuzov, I., Kulichkov, S., Otrezov, A., and Perepelkin, V. (**2005**). "Acoustic pulse propagation through a fluctuating stably stratified atmospheric boundary layer," J. Acoust. Soc. Am. **117**, 1868–1879.

Chunchuzov, I., Kulichkov, S., Perepelkin, V., Ziemann, A., Arnold, K., and Kniffka, A. (**2007**). "Acoustic tomographic study of the mesoscale coherent structures in the lower atmosphere," available online http://scitation.aip.org/confst/ASA/pub/8/4pPA5.pdf (Last viewed August 2008).

Chunchuzov, I. P., Bush, G. A., and Kulichkov, S. N. (**1990**). "On acoustical impulse propagation in a moving inhomogeneous atmospheric layer," J. Acoust. Soc. Am. **88**, 455–466.

Clark, T. L., Hauf, T., and Kuettner, J. P. (**1986**). "Convectively forced internal waves: Results from two-dimensional numerical experiments," Q. J. R. Meteorol. Soc. **112**, 899–925.

Cooper, D. I., Leclerc, M. Y., Archuleta, J., Coulter, R., Eichinger, W. E., Kao, C. Y., and Nappo, C. J. (**2006**). "Mass exchange in the stable boundary layer by coherent structures," Agric. Forest Meteorol. **136**, 114–131.

Danilov, A., and Chunchuzov, I. (**1992**). "Possible mechanism of layered structure formation in a stably stratified atmospheric boundary layer," Izv., Acad. Sci., USSR, Atmos. Oceanic Phys. **28**, 684–688.

Don, C. G., and Gramond, A. J. (**1986**). "Creeping wave analysis of impulse propagation through a shadow boundary," J. Acoust. Soc. Am. **80**, 302–305.

Finnigan, J. J., Einaudi, F., and Fua, D. (**1984**). "The interaction between an internal gravity wave and turbulence in the stably-stratified nocturnal boundary layer," J. Atmos. Sci. **41**, 2409–2436.

Gossard, E. E., and Hooke, W. H., *Waves in the Atmosphere* (Elsevier, Amsterdam, **1975**).

Grachev, A. I., Zagoruiko, S. V., Matveyev, A. K., and Mordoukhovich, M. I. (**1978**). "Some results of recording of atmospheric infrasound waves," Izv., Acad. Sci., USSR, Atmos. Oceanic Phys. **14**, 474–483.

Kadygrov, E. N., Kuznetsova, I. N., and Viazankin, A. S. (**2003**). "Investigation of atmospheric boundary layer temperature, turbulence, and wind parameters on the basis of passive microwave remote sensing," Radio Sci. **38**(3), MAR 13–113–12.

Kniffka, A., Arnold, K., Barth, M., Ziemann, A., Chunchuzov, I., Kulichkov, S., and Perepelkin, V. (**2006**). "Internal gravity waves in the lower atmosphere: spatial and temporal characteristics," Proceedings of the International Symposium for the Advancement of Boundary Layer Remote Sensing (ISARS-2006), Garmisch-Partenkirchen, Germany, pp. 109–111.

Petenko, I. V., and Bezverkhnii, V. A. (**1999**). "Temporal scales of convective coherent structures derived from sodar data," Meteorol. Atmos. Phys. **71**, 105–116.

Sang, J. G. (**1991**). "On formation of convective roll vortices by internal gravity waves: A theoretical study," Meteorol. Atmos. Phys. **46**, 15–28.

Stoll, S. A. (**1991**). "Microburst detection by the low-level wind shear alert system," Weather **46**, 334–347.

Sun, J., Lenschow, D., Burns, S. P., Banta, R. M., Newsom, R. K., Coulter, R. F., Stephene, I. T., Nappo, C., Balsley, B., Jensen, M., Mahrt, L., Miller, D., and Skelly, B. (**2004**). "Atmospheric disturbances that generate intermittent turbulence in nocturnal boundary layers," Boundary-Layer Meteorol. **110**, 255–279.

Ziemann, A., Arnold, K., and Raabe, A. (**2001**). "Acoustic tomography as a method to identify small-scale land surface characteristics," Acta. Acust. Acust. **87**, 731–737.

Broadband impedance boundary conditions for the simulation of sound propagation in the time domain

Jonghoon Bin[a] and M. Yousuff Hussaini[b]
Department of Mathematics, Florida State University, Tallahassee, Florida 32306

Soogab Lee[c]
School of Mechanical and Aerospace Engineering, Seoul National University, Seoul 151-742, South Korea

(Received 4 March 2008; revised 17 July 2008; accepted 19 September 2008)

An accurate and practical surface impedance boundary condition in the time domain has been developed for application to broadband-frequency simulation in aeroacoustic problems. To show the capability of this method, two kinds of numerical simulations are performed and compared with the analytical/experimental results: one is acoustic wave reflection by a monopole source over an impedance surface and the other is acoustic wave propagation in a duct with a finite impedance wall. Both single-frequency and broadband-frequency simulations are performed within the framework of linearized Euler equations. A high-order dispersion-relation-preserving finite-difference method and a low-dissipation, low-dispersion Runge–Kutta method are used for spatial discretization and time integration, respectively. The results show excellent agreement with the analytical/experimental results at various frequencies. The method accurately predicts both the amplitude and the phase of acoustic pressure and ensures the well-posedness of the broadband time-domain impedance boundary condition. © *2009 Acoustical Society of America.* [DOI: 10.1121/1.2999339]

PACS number(s): 43.28.Js, 43.28.En, 43.20.Rz [GCL] Pages: 664–675

I. INTRODUCTION

The acoustic impedance condition in computational aeroacoustic (CAA) applications, such as the calculation of sound propagation and absorption through an engine inlet duct and the reflection of acoustic waves in outdoor propagation, is an important issue.[1] Until now, several analytical attempts have been made to solve problems of acoustic wave reflection above an impedance surface. Analytical solutions and asymptotic formulas are useful and efficient if one is interested in a single observer point in an acoustic field. Li and White[2] derived an analytical expression of acoustic waves by a harmonic point source above an impedance surface. Di and Gilbert[3] recently represented the total acoustic field due to a point source above a complex impedance plane as a sum of the free space field and an image-source field, and obtained an image integral, which is relatively simple and rapidly convergent compared to the usual Sommerfeld integral. In more realistic and complicated broadband-frequency problems, however, it can be considerably difficult, if not impossible, to find analytical or approximate expressions. Because experimental approaches can be laborious and expensive, time-domain numerical approaches provide an attractive alternative for analyzing the effect of an impedance surface on the propagation of acoustic waves generated by broadband sources, such as a high-speed impulsive source.

Recently, several attempts[4–14] have been made to implement the impedance boundary conditions in the context of the time-domain methodology in CAA. Davis[4] considered acoustic waves in an open-ended pipe and solved the relevant one-dimensional equations using a fourth-order compact difference scheme with low-dispersion error and no amplitude dissipation error. He obtained the impedance boundary condition at the open end of the pipe by the inverse Fourier transform of the standard frequency-dependent impedance for transients with predominant low-frequency content. Bottledooren[5] proposed a finite-difference time-domain method for the simulation of the acoustic field on an analytically generated quasi-Cartesian grid and a simple expression for the impedance boundary condition relating pressure to the corresponding normal velocity where the normal impedance is given. The advantage of the method for the curved impedance boundary is demonstrated. Tam and Auriault[6] developed time-domain single-frequency and broadband impedance boundary conditions for the three-parameter impedance model of the Helmholtz resonator type. Zheng and Zhuang[7,8] verified and validated their broadband time-domain impedance boundary condition in semi-infinite two- and three-dimensional ducts with acoustically treated walls. Their impedance model, however, is a simple resonant type that cannot be extended to the general impedance problem. A general impedance condition was proposed by Long and co-workers[9,10] based on the z-transform. They pointed out that although the impedance model in rational form provides "quite accurate resistance and reactance representations of the experimental impedance data used in this paper, these representations did not meet the stability and causality criteria" as it does not ensure that the poles of the impedance lie within the unit circle in the z-domain, and that the region of convergence lies outside the outermost pole. Furthermore,

[a]Electronic mail: jbin@scs.fsu.edu
[b]Electronic mail: myh@cespr.fsu.edu
[c]Electronic mail: solee@snu.ac.kr

for accuracy, it may require a high degree of the rational function, which in turn requires high-order derivatives of the pressure and the velocity resulting in increased computational time. Fung and Ju[11–13] discussed some issues concerning the modeling and implementation of the time-domain impedance boundary condition. They showed that the reflection process corresponding to a typical impedance model is a convolution of the incident waves and the reflection impulse (which is the inverse Fourier transform of the reflection coefficient). It is pointed out that a direct inversion of impedance into time-domain boundary operators generally leads to an unstable system, whereas the inversion of the corresponding reflection coefficient results in stable, easily implementable boundary operators for time-domain prediction of wave reflection. They validated their models with the available analytical and experimental results. Although this approach is convenient to approximate an impedance curve and to treat the impedance condition numerically, high-order polynomial expansions for the reflection coefficient may lead to severe over- or underpredictions beyond the limited frequency region of interest. Furthermore, it is difficult to represent the resonance phenomena in the impedance curve using this approach. The impedance modeling technique of Wilson *et al.*[14] cast the convolution integrals of their relaxation impedance model in a form amenable to numerical implementation, and it has been demonstrated on two-dimensional calculations of outdoor sound propagation involving hills, barriers, and ground surfaces with various material properties. This approach is both computation and memory intensive since convolution integrals are involved.

The objective of this paper is to develop a robust, accurate, and practical time-domain impedance boundary condition for acoustic simulation with broadband frequencies. To validate this condition, two kinds of numerical simulations are performed: one is acoustic propagation due to point sources over an impedance surface in an open field and the other is noise propagation in a duct with a finite impedance wall. Examples of a typical grass ground impedance and a wool-felt ground impedance are used to illustrate the practicality and effectiveness of the impedance model for the first case.[15] Both single-frequency and broadband-frequency calculations are performed. The second set of examples is for acoustic propagation in a two-dimensional duct with a finite ceramic tubular liner (CT73).[10,16] The numerical solutions are compared with the analytical/experimental results in both cases.

The paper is organized as follows. Section II provides brief mathematical preliminaries, and succinct derivation of the impedance boundary condition followed by a brief description of the discretization scheme. Section III describes the validation problems with their governing equations, boundary conditions, and discretization scheme. Numerical results are presented in Sec. IV.

II. BROADBAND IMPEDANCE BOUNDARY CONDITION

A. Mathematical preliminaries

The characteristics of acoustic impedance are measured and usually described in the frequency domain. Myers[17] and Ingard[18] derived a general acoustic impedance boundary condition assuming that an acoustically treated wall makes deformations in response to an incident acoustic field from the fluid and ignoring a possible hydrodynamic mode.[19] These deformations are assumed to be small perturbations compared to a stationary mean surface, and the corresponding fluid velocity field is a small perturbation about a mean base flow u_0. The linearized frequency-domain impedance boundary condition with mean flow is expressed as

$$\hat{u}(\omega)\cdot n=(i\omega+u_0\cdot\nabla-n\cdot(n\cdot\nabla u_0))\frac{\hat{p}(\omega)}{i\omega Z(\omega)}. \tag{1}$$

Here $\hat{u}(\omega)$ is the complex amplitude of the velocity perturbation, ω is the angular frequency, n is the normal vector to the wall that points into the wall, $\hat{p}(\omega)$ is the complex amplitude of the pressure perturbation, and $Z(\omega)=R(\omega)+iX(\omega)$ [where $R(\omega)$ and $X(\omega)$ are the frequency-dependent resistance and reactance, respectively] is the acoustic impedance. (All these quantities are nondimensionalized with respect to their corresponding characteristic values, which are defined in Sec. III.) The use of this condition is limited to linear unsteady flow problems. Since the mean flow mainly satisfies $u_0\cdot n=0$ on the wall, this equation can be recast into

$$\hat{u}(\omega)\cdot n=(i\omega+u_0\cdot\nabla+u_0\cdot(n\cdot\nabla n))\frac{\hat{p}(\omega)}{i\omega Z(\omega)}, \tag{2}$$

where $n\cdot\nabla n$ is tangential to the wall surface and vanishes for a flat surface. If $Z(\omega)$ is assumed to be independent of position, we get the following form by multiplying $i\omega Z(\omega)$ by both sides of Eq. (2):

$$i\omega Z(\omega)\hat{u}(\omega)\cdot n=i\omega\hat{p}(\omega)+u_0\cdot\nabla\hat{p}(\omega)+u_0\cdot(n\cdot\nabla n)\hat{p}(\omega). \tag{3}$$

By applying the inverse Fourier transform to Eq. (3) and considering the causality condition, we get the impedance boundary condition in the physical domain as follows:

$$\frac{1}{2\pi}\int_0^t z(t-\tau)\frac{\partial}{\partial\tau}u(\tau)\cdot nd\tau=\frac{\partial p}{\partial t}+u_0\cdot\nabla p+u_0\cdot(n\cdot\nabla n)p. \tag{4}$$

If there is no mean flow, the impedance condition of Eq. (3) can be simply expressed by

$$\hat{p}(\omega)=Z(\omega)(\hat{u}(\omega)\cdot n), \tag{5}$$

$$p(t)=\frac{1}{2\pi}\int_0^t z(t-\tau)u(\tau)\cdot nd\tau, \tag{6}$$

where $p(t)$ and $u(t)$ denote the inverse Fourier transforms of $\hat{p}(\omega)$ and $\hat{u}(\omega)$ at the impedance wall, respectively. The impedance $z(t)$ is given by

$$z(t)=\int_{-\infty}^{\infty}Z(\omega)e^{i\omega t}d\omega. \tag{7}$$

The evaluation of the convolution integrals in Eqs. (4) and (6) is computationally expensive and may become prohibitively expensive especially for multidimensional CAA prob-

lems. We propose a relatively efficient and robust broadband time-domain impedance model, which circumvents the convolution integral problem.

B. Derivation of the broadband time-domain impedance boundary condition

With $\hat{v}(\omega)=\hat{u}(\omega)\cdot n$, Eq. (3) reads

$$i\omega\hat{p}(\omega)+u_0\cdot\nabla\hat{p}(\omega)+u_0\cdot(n\cdot\nabla n)\hat{p}(\omega)=-i\omega Z(\omega)\hat{v}(\omega). \tag{8}$$

The term $u_0\cdot(n\cdot\nabla n)\hat{p}(\omega)$ is usually small, which is the case if the curvature of the wall does not vary significantly, and hence it is usually neglected. In this study, the impedance surface is a plane and it is rightly neglected. Then the acoustic impedance boundary condition in the frequency domain can be expressed as

$$i\omega\hat{p}(\omega)+u_0\cdot\nabla\hat{p}(\omega)=-i\omega Z(\omega)\hat{v}(\omega). \tag{9}$$

To proceed further, we need to model the impedance, $Z(\omega)$.

1. The impedance model

Care must be exercised in modeling the impedance so that it provides the impedance boundary condition in a form that is amenable to an efficient direct numerical simulation of acoustics involving response to broadband frequencies. As the second-order frequency response function (FRF) can act as a low-pass filter or a bandpass filter, we propose to represent the impedance as a linear sum of FRFs (and provide *a posteriori* justification) as follows:

$$Z(\omega)=\sum_{j=1}^{N}\frac{a_0^j(i\omega)+a_1^j}{b_0^j(i\omega)^2+b_1^j(i\omega)+b_2^j}, \tag{10}$$

where N is the number of FRFs and the a's and b's are the constants/parameters that are so determined as to yield the best approximation to the empirical data. These parameters can be determined by common optimization methods, such as a nonlinear least squares fit algorithm or the steepest descent method. In this study, the conjugate gradient method is used to obtain the optimal values of the parameters,[20,21] which yield accurate representations of the well-known two-parameter empirical impedance model[1,15] and of experimental data from NASA.[10,16] The stability analysis in the appendixes shows that the impedance model is stable if all model parameters are real and positive. With N equal to 4, we obtain a good approximation of engineering accuracy for the impedance of the grass ground and the wool-felt ground, respectively (Figs. 4 and 5). Again, the impedance model fits very well with the Langley experimental data of a ceramic tubular liner (CT73) (Fig. 6). Further discussion is deferred to Sec. IV.

After the substitution of $Z(\omega)$ from Eq. (10) in Eq. (9), with some algebra and manipulation, we obtain

$$i\omega\hat{p}(\omega)+u_0\cdot\nabla\hat{p}(\omega)=i\omega\left[\sum_{j}^{N}\hat{p}_j(\omega)\right], \tag{11a}$$

$$\hat{p}_j(\omega)=-\hat{\nu}(\omega)\frac{a_0^j(i\omega)+a_1^j}{b_0^j(i\omega)^2+b_1^j(i\omega)+b_2^j},\quad j=1,\ldots,N, \tag{11b}$$

which are obtained from Eqs. (5) and (10). The inverse Fourier transform of Eq. (11a) yields

$$\frac{\partial p}{\partial t}+u_0\cdot\nabla p=\sum_{j}^{N}\frac{\partial p_j}{\partial t}. \tag{12}$$

To implement this impedance boundary condition, we should obtain the values of pressure, p_j or their derivatives, $\partial p_j/\partial t$. The pressure p_j's are the auxiliary values introduced by the impedance model of Eq. (10) and only depend on the normal velocity perturbation at the impedance surface. Thus, each p_j can be computed separately from the physical values, such as the pressure and normal velocity at the wall. Now, $\hat{p}_j(\omega)$ can be rearranged into the following form after applying the inverse Fourier transform to Eq. (11b):

$$\left(b_0^j\frac{\partial^2 p_j}{\partial t^2}+b_1^j\frac{\partial p_j}{\partial t}+b_2^j p_j\right)=-\left(a_0^j\frac{\partial v}{\partial t}+a_1^j v\right),\quad j=1,\ldots,N, \tag{13}$$

where v is the normal velocity perturbation on the wall, and p_j is the pressure of the jth subcomponent in Eq. (12).

C. Discretization scheme

Using the second-order finite-difference scheme, Eq. (13) can be discretized as

$$\left(b_0^j\frac{p_j^{(n+1)}-2p_j^{(n)}+p_j^{(n-1)}}{\Delta t^2}+b_1^j\frac{p_j^{(n+1)}-p_j^{(n-1)}}{2\Delta t}+b_2^j p_j^{(n+1)}\right)=-\left(a_0^j\frac{v^{(n+1)}-v^{(n-1)}}{2\Delta t}+a_1^j v^{(n)}\right) \tag{14}$$

for $j=1,\ldots,N$. The solution of Eq. (14) for the acoustic pressure at the $(n+1)$ time step, $p_j^{(n+1)}$, requires the acoustic velocity at the $(n+1)$ time step, $v^{(n+1)}$ [and the acoustic pressure and velocity at the (n) and $(n-1)$ time steps], implying the implicit nature of the impedance condition. The implicit discretization of Eq. (13) may enhance the numerical stability, but results in additional complexity. In order for the impedance boundary condition to be implemented explicitly, the normal momentum equation at the impedance wall is applied. Substituting for $\partial v/\partial t$ from the y-momentum equation in Eq. (21), Eq. (14) becomes

$$\left(b_0^j\frac{p_j^{(n+1)}-2p_j^{(n)}+p_j^{(n-1)}}{\Delta t^2}+b_1^j\frac{p_j^{(n+1)}-p_j^{(n-1)}}{2\Delta t}+b_2^j p_j^{(n+1)}\right)=\left(a_0^j\left(M_x\frac{\partial v^{(n)}}{\partial x}+\frac{\partial p^{(n)}}{\partial y}\right)-a_1^j v^{(n)}\right) \tag{15}$$

for $j=1,\ldots,N$. All of the values except $p_j^{(n+1)}$ are known, or can be computed from the values in the interior computational domain and on the impedance surface at the previous time step. At time level $(n+1)$, $p_j^{(n+1)}$ can be obtained explicitly from the following equation:

$$\left(\frac{b_0^j}{\Delta t^2}+\frac{b_1^j}{2\Delta t}+b_2^j\right)p_j^{(n+1)}=\frac{2b_0^j}{\Delta t^2}p_j^{(n)}-\left(\frac{b_0^j}{\Delta t^2}-\frac{b_1^j}{2\Delta t}\right)p_j^{(n-1)}+\left(a_0^j\left(M_x\frac{\partial \nu^{(n)}}{\partial x}+\frac{\partial p^{(n)}}{\partial y}\right)-a_1^j\nu^{(n)}\right) \tag{16}$$

for $j=1,\ldots,N$. Using the pressure values, p_j, and substituting them into Eq. (12), we get

$$\frac{p^{(n+1)}-p^{(n-1)}}{2\Delta t}+M_x\frac{\partial p^{(n+1)}}{\partial x}=\mathrm{RHS}\left(=\sum_j^N\frac{p_j^{(n+1)}-p_j^{(n-1)}}{2\Delta t}\right). \tag{17}$$

In this study, the temporal and spatial derivatives of the pressure in the impedance condition are discretized using the second-order central difference scheme

$$Ap_{(k-1,l)}^{(n+1)}+Bp_{(k,l)}^{(n+1)}+Cp_{(k+1,l)}^{(n+1)}=\frac{p_{(k,l)}^{(n-1)}}{2\Delta t}+\mathrm{RHS}, \tag{18}$$

$$A=-\frac{M_x}{2\Delta x},\quad B=\frac{1}{2\Delta t},\quad C=\frac{M_x}{2\Delta x}, \tag{19}$$

where the subscript (k,l) in Eq. (19) is the grid point index in the x- and y-directions, respectively. Specifically, l represents the grid point at the impedance surface in Eq. (18). For simplicity, a uniform mesh is chosen with $\Delta x=\Delta y=\mathrm{const}$. This results in a tridiagonal matrix form. If fourth-order spatial discretization is used in Eq. (17), a pentadiagonal equation matrix system will result. If there is no mean flow in the computational domain ($M_x=0$), the impedance boundary condition, Eq. (12), can be simply expressed as

$$p^{(n+1)}=\sum_j^N p_j^{(n+1)}. \tag{20}$$

To make this impedance boundary condition satisfy the causality condition, acoustic perturbations are assumed to be absent for $t<0$ and all physical values are set to zero for $t<0$ in the numerical simulations. The computed results using the broadband time-domain impedance boundary condition will be compared with the analytical/experimental solutions.

III. VALIDATION PROBLEMS

To validate our time-domain broadband impedance methodology, we choose two example problems: (i) the computation of the sound field above an impedance ground relevant to outdoor sound propagation, and (ii) the computation of the sound field in a duct with mean flow relevant to acoustic fields inside and radiated from both inlet and exhaust of turbofan engines.

A. Governing equations

We assume that the linearized Euler equations govern the acoustic field

$$\frac{\partial U}{\partial t}+A_x\frac{\partial U}{\partial x}+B_y\frac{\partial U}{\partial y}+mSU=mF, \tag{21}$$

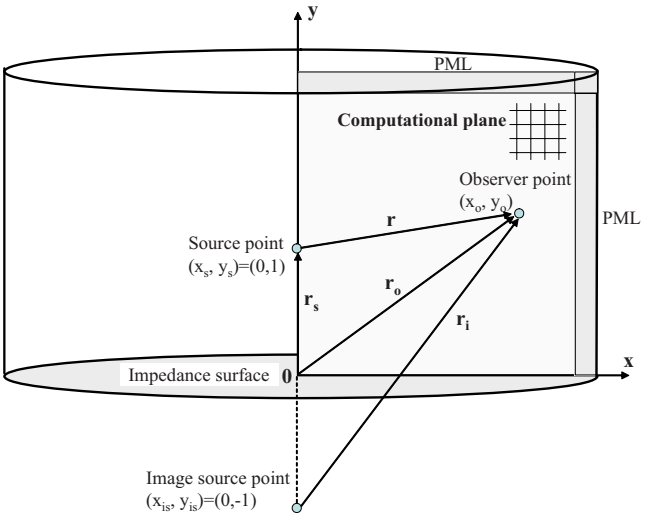


FIG. 1. (Color online) Coordinate system of a point monopole source over an impedance surface.

$$U=\begin{pmatrix}\rho\\ u\\ \nu\\ p\end{pmatrix},\quad A_x=\begin{pmatrix}M_x & 1 & 0 & 0\\ 0 & M_x & 0 & 1\\ 0 & 0 & M_x & 0\\ 0 & 1 & 0 & M_x\end{pmatrix},$$

$$B_y=\begin{pmatrix}M_y & 0 & 1 & 0\\ 0 & M_y & 0 & 0\\ 0 & 0 & M_y & 1\\ 0 & 0 & 1 & M_y\end{pmatrix},\quad S=\frac{1}{x}\begin{pmatrix}0 & 1 & 0 & 0\\ 0 & M_x & 0 & 0\\ 0 & 0 & 0 & 0\\ 0 & 1 & 0 & 0\end{pmatrix}, \tag{22}$$

where (x,y) represent the coordinates, ρ is the density perturbation, u and v are the velocity perturbations in the x- and y-directions, respectively, p is the pressure perturbation, $\mathbf{F}$ is the source vector, and M_x and M_y represent the Mach number in the x- and y-directions, respectively. For two-dimensional problems, $m=0$, and for axisymmetric problems, $m=1$, where (x,y) represent the radial and the axial coordinates, respectively. In this paper, density, velocity, and pressure are nondimensionalized by ρ_∞, c_∞, and $\rho_\infty c_\infty^2$, respectively. Quantities with the subscript ∞ denote the ambient conditions with c_∞ being the speed of sound. The coordinates (x,y) are scaled with respect to an appropriate length, L, which characterizes the problem. The characteristic length in Problem 1 is the height of the monopole source above the impedance surface, and in Problem 2 it is the duct height.

Problem 1: Acoustic wave reflection on an impedance surface (no mean flow). The sound field due to an acoustic monopole source in a homogeneous medium without mean flow above an impedance ground is cylindrically symmetric, in that there is no azimuthal variation. In order to consider three-dimensional acoustic wave propagation, cylindrical coordinates are used, where x is the radial coordinate, y is the axial coordinate, and $y=0$ is the impedance plane. Figure 1 shows the schematic of the coordinate system and the computational domain. The height of the monopole source, H^*, from the impedance plane is the characteristic length of the

problem, with which all lengths are scaled, and thus the acoustic source is located at $(x_s, y_s)=(0,1)$. An image-source point is used to obtain the analytic solution at an observer point.[3]

To simulate the acoustic field, Eq. (21) (with the source vector defined as $\mathbf{F}=[f_m,0,0,f_m]^T$) is solved with the following initial and boundary conditions.

Initial condition: $U=(\rho,u,v,p)^T=0$.

Boundary conditions: At the impedance boundary, $y=0$, the time-domain impedance boundary condition [Eq. (17)] is applied. At the radial open boundary, $x=x_\infty$, and the axial open boundary, $y=y_\infty$, the perfectly matched layer (PML) method[22–25] is applied. Axial symmetry condition is applied at $x=0$.

We consider both single-frequency and broadband-frequency cases.

Case (a): Periodic reflection at a single frequency. A monopole source is introduced to generate an acoustic wave in the computational domain. Following Ju and Fung,[3] we introduce a monopole source equivalent to a point monopole source strength, which is written as

$$f_m(r,t)=\frac{1}{B_w^3\pi^{3/2}}e^{-(r/B_w)^2}e^{(kB_w)^2/4}\sin(\omega t), \tag{23}$$

where $r=|r_s-r_o|$ is the distance between the observer at r_o and the source point at r_s, B_w is the half-width of a Gaussian distribution, and k is the wave number. The half pulse-width relative to the characteristic length $\bar{B}_w=B_w/H^*$ used for a monopole is very small ($\bar{B}_w=0.0375$) to maintain solution smoothness and to ensure equivalence with a point monopole source. The nondimensional coordinates of the monopole source are $(x,y)=(0,1)$, where the coordinates have been scaled with respect to the characteristic length represented by the height, H^*, of the source above the impedance surface. Here H^* is assumed to be equal to $2\lambda^*$, where $\lambda^*=c_\infty/f$. We chose the frequency $f=3$ kHz as the human ear is most sensitive around 2–5 kHz, and many acoustic simulations and applications to sound reduction have used the frequency range around it.

Case (b): Excess attenuation at broadband frequencies. For broadband-frequency simulations, we introduce monopole sources containing several frequencies of interest, and obtain the response over an impedance wall. For this, the monopole source term is defined as

$$f_m(x,y,t)=0.01\sum_{m=1}^{45}e^{-\ln 2(((x-x_s)^2+(y-y_s)^2)/B^2)}\times\cos\left[\left(\omega_0+2\pi(m-1)\frac{100}{c_\infty}\right)t+2(m-1)\pi/45\right], \tag{24}$$

$$x_s=0,\quad y_s=1.0,\quad B=\frac{2\pi}{2[\omega_0+2\pi(m-1)\times 100/c_\infty]}, \tag{25}$$

where $\omega_0=2\pi f_0$ with $f_0=500/c_\infty$ and $c_\infty=340$ m/s, which corresponds to 500 Hz.

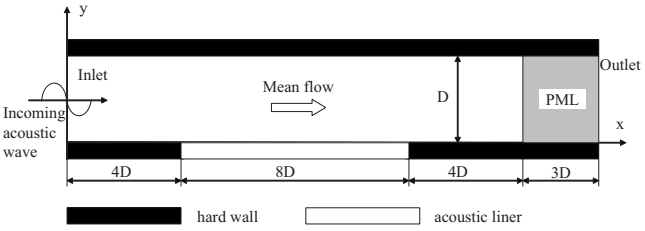


FIG. 2. The schematic of the flow-impedance duct.

Problem 2: Sound propagation in a duct with a finite impedance wall (with mean flow). We consider sound propagation in a semi-infinite two-dimensional duct with a finite impedance wall and unidirectional mean flow ($M_x=0.1$, $M_y=0.0$). The characteristic length is naturally the height of the duct, D. Figure 2 is the schematic of the computational domain. Equation (21) is solved with the following initial and boundary conditions.

Initial condition: $U=(\rho,u,v,p)^T=0$.

Inflow boundary condition: Assuming a uniform mean flow in the x-direction, the acoustic disturbances would satisfy the following radiation boundary condition at inflow ($x=0$):

$$\left\{\frac{\partial}{\partial t}-(1-M_x)\frac{\partial}{\partial x}\right\}\begin{pmatrix}\rho\\u\\v\\p\end{pmatrix}=-2\frac{\partial}{\partial x}\begin{pmatrix}\rho_d\\u_d\\v_d\\p_d\end{pmatrix}. \tag{26}$$

The downstream propagating wave at the inflow region $[\rho_d\ u_d\ v_d\ p_d]^T$ can be written as

$$\begin{pmatrix}\rho_d\\u_d\\v_d\\p_d\end{pmatrix}=A(t)\begin{pmatrix}1\\1\\0\\1\end{pmatrix}\sin\left\{\omega\left(\frac{x}{1+M_x}-t\right)+\varphi\right\}, \tag{27a}$$

$$A(t)=\text{Amp}\times\exp\left[-\ln 2\frac{(y-1)^2}{\{3\times\Delta y\times(1+t)\}^2}\right], \tag{27b}$$

where $A(t)$ and φ denote the amplitude and the phase of the incoming acoustic waves, respectively. The constant Amp is set to $\sqrt{2}p_{\text{ref}}10^{\text{SPL}/20}/\rho_\infty c_\infty^2$ with $p_{\text{ref}}=2.0\times10^{-5}$ Pa and $\varphi=0$. The implementation of the time-domain impedance boundary condition requires physical information from the previous time step. To avoid the transient effect of incoming waves on the response of an impedance surface, a Gaussian distributed incoming wave amplitude, as expressed in Eq. (27b), is introduced, which precludes abrupt jumps of physical values in the initial stage near the impedance wall.

Outflow boundary condition. At the damping layer, ($16\leqslant x\leqslant 19$), the two-dimensional PML method[22–25] is used to damp out the waves to prevent their reflection back into the domain of computation.

At the upper rigid wall ($y=1$), the normal velocity is set to zero. At the bottom impedance boundary ($y=0$), the time-domain impedance boundary condition [Eq. (17)] is imposed.

B. Numerical algorithm

In this work, the seven-point stencil dispersion-relation-preserving (DRP) scheme is used for spatial discretization.[26] A low-dissipation, low-dispersion Runge–Kutta method is used for time integration.[27]

1. Numerical damping

The DRP scheme is a central difference scheme with fourth-order accuracy with zero intrinsic dissipation. In order to eliminate spurious short waves and to improve numerical stability, two kinds of damping terms are used in this computation. The first kind is artificial selective damping proposed by Tam and Webb,[26] which is added to the discretized finite-difference equations, as shown in Eq. (29). An inverse mesh Reynolds number $R_\Delta^{-1}=\nu_a/(a_\infty\Delta)$, where ν_a and Δ are the artificial kinematic viscosity and mesh size, respectively, is prescribed over the whole computational domain to damp out the spurious waves.

$$U_{k,l}^{(n+1)} = U_{k,l}^{(n)} + \Delta t \sum_{j=0}^{3} b_j K_{k,l}^{(n-j)}, \tag{28}$$

$$K_{k,l}^{(n)} = -\frac{1}{\Delta x} A_x \sum_{j=-3}^{3} a_j U_{k+j,l}^{(n)} - \frac{1}{\Delta y} B_y \sum_{j=-3}^{3} a_j U_{k,l+j}^{(n)} - mSU_{k,l}^{(n)} + \bar{D}_{k,l}(U) + D_{k,l}(U), \tag{29}$$

$$\bar{D}_{k,l}(U) = (\bar{D}_x)_{k,l} + (\bar{D}_y)_{k,l} = -\frac{L}{\Delta x R_\Delta} \sum_{j=-3}^{3} d_j (U_{k+j,l}^{(n)} + U_{k,l+j}^{(n)}). \tag{30}$$

Near the wall a reduced number of points are used for damping, i.e., $j=-2,\ldots,2$ at $l=3$ and $j=-1,\ldots,1$ at $l=2$ in $(\bar{D}_y)_{k,l}$. The coefficients and the details can be found in Ref. 26.

On the impedance surface, there exists a mismatch of velocity and pressure because of their relationship in the acoustic impedance condition. Therefore, the amplitudes and the phases of the incoming waves to the impedance wall will be changed after the reflection from the surface. Furthermore, due to the use of the second- or fourth-order finite-difference discretization for the impedance condition in Eqs. (14) and (17), all of the time impedance information cannot be included in this discretization. To remove unphysical surface waves, which may be generated in this process, and to obtain the stable solution of the acoustic reflection from the surface, the following dissipation is used in the interior domain. Fourth-order dissipation is used with the second-order computations of the impedance condition, and sixth-order dissipation is used with the fourth-order computations

$$D_{k,l}^{(4)}(U) = -\kappa^{(4)}(\delta_x^{(4)} + \delta_y^{(4)})U, \quad D_{k,l}^{(6)}(U) = \kappa^{(6)}(\delta_x^{(6)} + \delta_y^{(6)})U, \tag{31a}$$

$$\delta_x^{(4)} U = \sum_{i=-2}^{2} w_i U_{k+i,l} \quad \text{with } w_i = [1\ \ -4\ \ 6\ \ -4\ \ 1] \quad \text{for } i=-2,\ldots,2, \tag{31b}$$

$$\delta_x^{(6)} U = \sum_{i=-3}^{3} w_i U_{k+i,l} \quad \text{with } w_i = [1\ \ -6\ \ 15\ \ -20\ \ 15\ \ -6\ \ 1] \quad \text{for } i=-3,\ldots,3. \tag{31c}$$

Typical values of the constants $\kappa^{(4)}$ and $\kappa^{(6)}$ are $1/128$ and $1/512$, respectively. The y-direction operators are defined in a similar manner. Dissipation is set to zero if values beyond a boundary are needed in the computation. It was found from the numerical simulation that the use of two kinds of damping can effectively eliminate the small amplitude three-point oscillations without affecting the accuracy of the physical solution and was critical in obtaining stable numerical solutions.

TABLE I. The parameter values used in the computation (Ref. 15.)

	Effective flow resistivity (σ_e) (kPa s m^{-2})	Effective rate of change of porosity (α_e) (m^{-1})
Grass-covered ground	100	20
Wool-felt material	38	15

IV. RESULTS AND DISCUSSION

A. The impedance model

Problem 1: Acoustic wave reflection over an impedance surface. In the following calculations, an empirical two-parameter impedance model[15] is used for the impedance data, where the normalized specific impedance Z of the ground is given by

$$Z = 0.436(1-i)\left(\frac{\sigma_e}{f}\right)^{0.5} - 19.48i\frac{\alpha_e}{f}, \tag{32}$$

where f is the frequency, and σ_e and α_e are the effective flow resistivity and the effective rate of change of porosity with depth, respectively. This model has often been used to fit excess attenuation data, especially that measured outdoors.[15] In this study, two types of parameter values are used for the validation of the impedance boundary condition. The parameters are given in Table I.

The objective is to determine the impedance function $Z(\omega)$ in Eq. (10), which is the best approximation to the empirical curve [Eq. (32)]. To that end, we choose frequencies ω_l, $l=1,2,3,\ldots,N_{\text{total}}$ from the frequency range of interest, and for each frequency we construct a FRF whose peak equals the impedance value at that frequency. Specifically, a second-order bandpass filter has the peak frequency and bandwidth, respectively, given by $\omega_0^2=b_2/b_0$ and $\beta=b_1/b_0$. The initial values of a's are arbitrarily assumed to be unity. Figure 3 shows the schematic of the initial FRFs and the experimental impedance curve. With this initial guess of impedance,

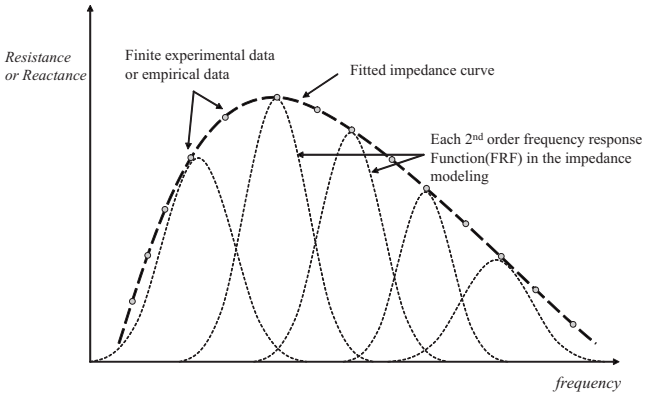


FIG. 3. The schematic of a modeled impedance function.

$$Z_l = Z(\omega_l) = \sum_{j=1}^{N} \frac{a_0^j (i\omega_l) + a_1^j}{b_0^j (i\omega_l)^2 + b_1^j (i\omega_l) + b_2^j}, \tag{33}$$

we start the minimization process of the objective function $I=\Sigma_{l=1}^{N_{\text{total}}}(Z_l-Z_l)^2$ (where Z_l are the empirical or experimental data) to obtain the optimal values of the parameters a's and b's. In the present case, we initially chose four uniformly spaced frequencies (specifically, $\omega_l=2,4,6,8$ kHz) in the frequency range of interest for both the grass and wool-felt grounds. With these as peak frequencies, initial FRFs are constructed to form the initial Z_l, and start the optimization process. Initial guess is important to ensure convergence of the optimization process to the right critical point. The convergence tolerance on the parameter value is on the order of 10^{-6}. The optimal parameter values are given in Appendix A.

Figures 4 and 5 show the comparison between the two-parameter impedance model[15] and the fitted impedance model used in this study for a grass ground and a wool-felt ground. There is an excellent agreement between the empirical curve and the fitted model. In these figures, individual

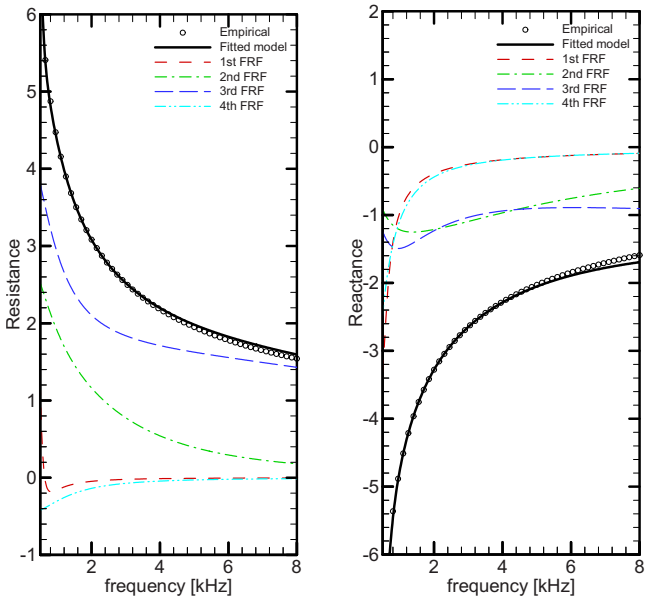


FIG. 4. (Color online) Fitted impedance model (solid line) and empirical model (circles): surface impedance of the typical grass ground. The individual FRFs are shown for comparison of their impedance content.

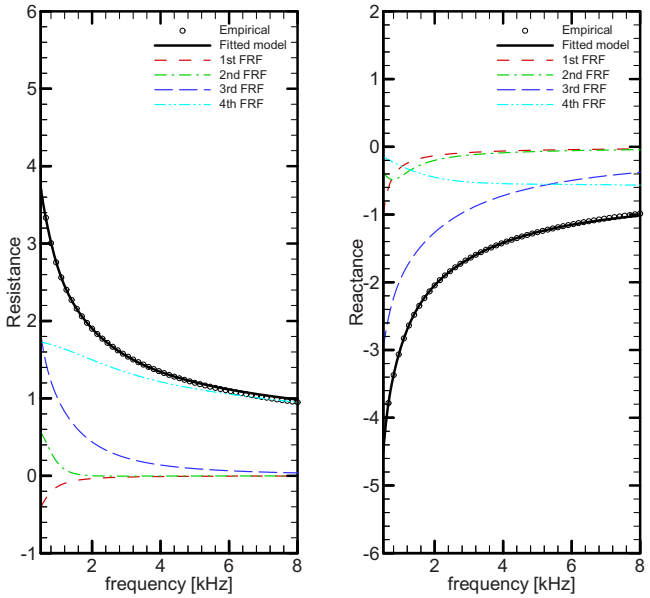


FIG. 5. (Color online) Fitted impedance model (solid line) and empirical model (circles): surface impedance of the typical wool-felt material. The individual FRFs are included to show their relative impedance content.

FRFs are also plotted to show that most FRFs contain the contributions of all the frequencies of interest (from 500 to 8 kHz) in both the resistance and the reactance, owing to broadband characteristic of the impedance surface.

Problem 2: Acoustic wave propagation in a duct with a finite acoustic liner. The input data used to extract the impedance of the test specimen were obtained from measurements using a flow-impedance tube at the NASA Langley Flow-Impedance Test Laboratory.[9,10,16] To fit our impedance model to the experimental data, we proceeded as in the case of Problem 1 except that we set the peaks of the FRFs at $\omega_l=1000$, 1500, 2000, 2500, and 3000 Hz. The optimal values of the model parameters are given in Appendix A. Figure 6 shows the comparison between the experimental values and the fitted model. Excellent agreement is observed. The

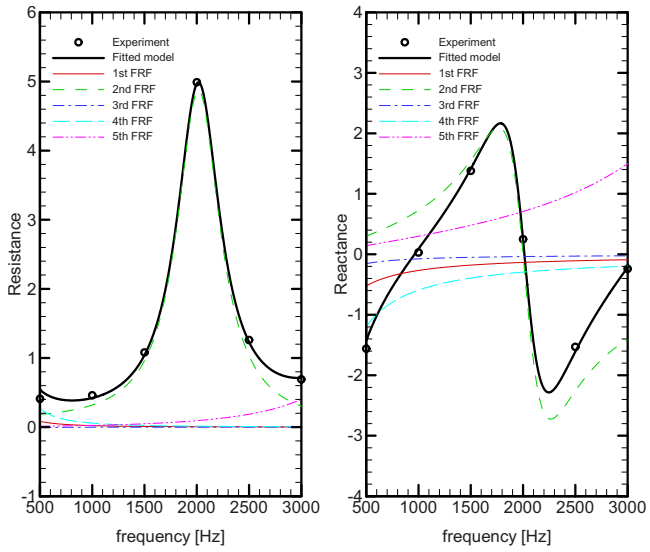


FIG. 6. (Color online) Fitted impedance model (solid line) and experimental data (circles) for resistance and reactance of a ceramic tubular liner (CT73).

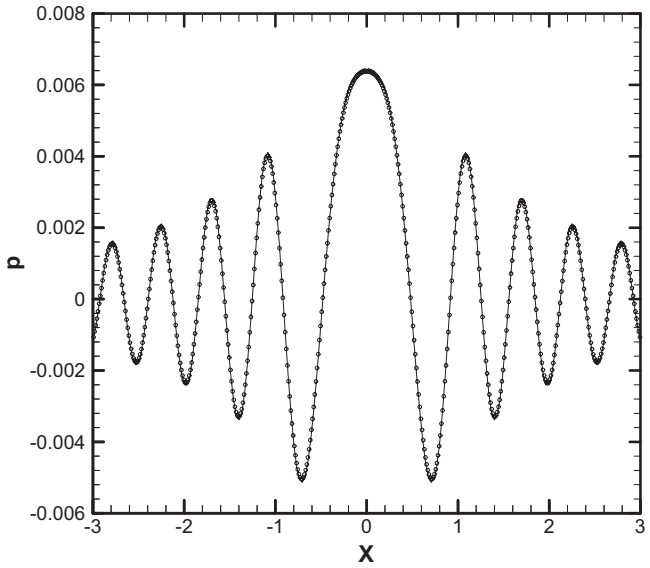


FIG. 7. Comparison of harmonic pressure distribution computed in time domain (symbols) with analytical solution (solid line) on the impedance surface ($y=0$) at $t=50$ T at $f=3$ kHz over a grass ground ($Z=2.52-2.65i$).

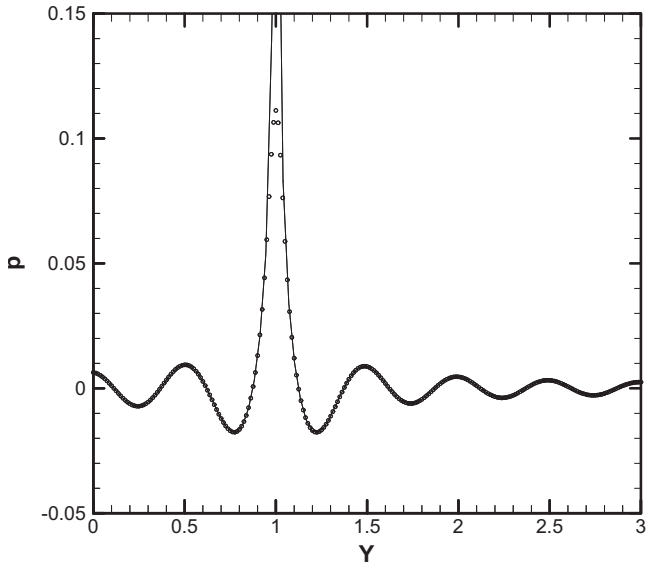


FIG. 8. Comparison of harmonic pressure distribution computed in the time domain (symbols) with analytical solution (solid line) along the line perpendicular to the impedance surface, i.e., axisymmetric line ($x=0$) at $t=50$ T at $f=3$ kHz over a grass ground.

impedance characteristic shows the resonance phenomenon in the impedance curve. The contribution of the second FRF term among the five FRFs that make up the impedance curve is dominant when compared with the other terms. Improvement in the impedance model can be obtained by increasing the number of FRFs, but we obtain satisfactory results using four or five FRFs in each problem.

B. Numerical validation

Case (a): Periodic reflection at a single frequency (Problem 1). In order to validate the impedance boundary condition, the numerical simulation of sound reflection over an impedance wall at a single frequency is performed. In this computation, the grid spacing and dimensions are $\Delta x=\Delta y=1/80$ and $D_x\times D_y=3\times 3$ in the x- and y-directions, respectively. Computations are performed using a nondimensional time step of $\Delta t=\mathrm{CFL}\Delta x$ where the Courant–Friedrich–Lewy (CFL) number has the value of 0.05.

Figures 7 and 8 show comparisons of the harmonic pressure distribution computed in the time domain with the analytical solution on the impedance surface ($y=0$) and along the line perpendicular to the impedance surface ($x=0$) at $t=50$ T at $f=3$ kHz over a grass ground ($Z=2.52-2.65i$). Figures 9 and 10 compare the analytical solution with the numerical solution using the present impedance boundary condition for an impedance value of wool felt at $f=3$ kHz ($Z=1.55-1.65i$). These figures show excellent agreement between the analytical and numerical solutions.

Case (b): Excess attenuation at broadband frequencies (Problem 1). To show that the impedance boundary condition works well in broadband problems, excess attenuation at broadband frequencies is performed and compared with an analytical solution in this section. The numerical simulation was performed on a 461×241 equally spaced grid in both the x- and y-directions ($\Delta x=\Delta y=1.13\times 10^{-2}$) with a CFL number of 0.05. In this computation, monopole sources containing 45 frequencies are considered, as is obvious from Eq. (24). Time signals stored at the observer points are used to compute the sound pressure level (SPL) (in decibels) for comparing the excess attenuation. The time data required to compute the SPL are stored after the sound field reaches the periodic state. Figure 11 compares the prediction of excess attenuation, which is defined as the total sound field relative to the direct field, for a monopole. The source and observer height are 1.0 m and 0.5 m, respectively, and the separation range along the x-direction for the comparison is 1.0 m. Figure 11 shows the comparison of excess attenuation between the computed result and the analytical solution for a grass ground and for a wool-felt ground. A fast Fourier transform

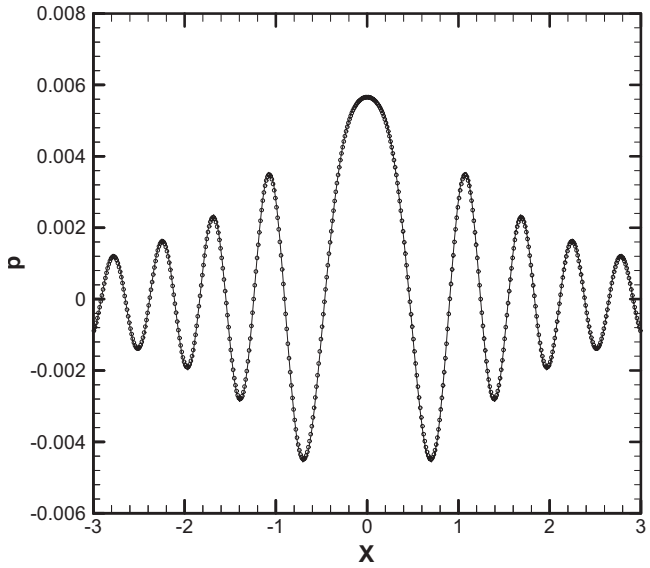


FIG. 9. Comparison of harmonic pressure distribution computed in the time domain (symbols) with analytical solution (solid line) on the impedance surface ($y=0$) at $t=50$ T at $f=3$ kHz over a wool-felt ground ($Z=1.55-1.65i$).

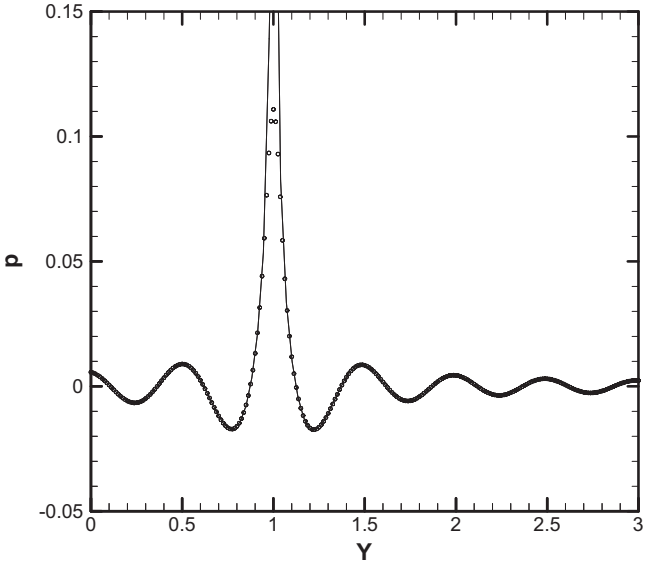


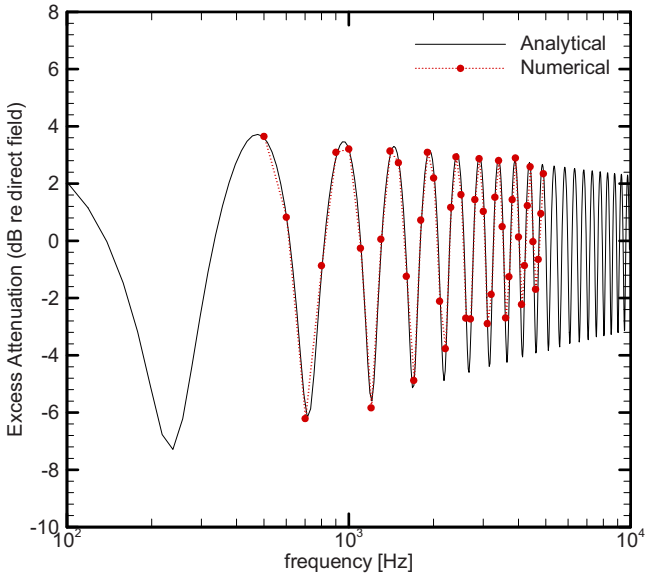
FIG. 10. Comparison of harmonic pressure distribution computed in the time domain (symbols) with analytical solution (solid line) along the line perpendicular to the impedance surface, i.e., axisymmetric line (x=0) at t =50 T at f=3 kHz over a wool-felt ground.

is used to obtain the SPL for each individual frequency from the time accurate signals. To obtain the SPL of the direct field, additional numerical computations without an impedance wall were performed. It can be seen from these figures that there is a little difference between the predicted and the analytical excess attenuation spectra for a monopole source. The agreement between the calculations and the analytical solution gives us confidence that it is possible to get accurate broadband numerical solutions using the broadband time-domain impedance boundary condition developed in this paper.

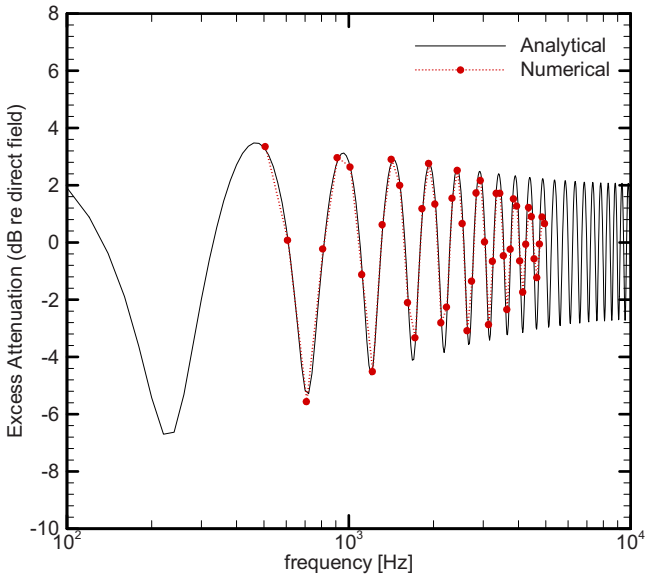
1. Absorption in an acoustic liner with mean flow (Problem 2)

The numerical simulations were performed on a 571 $\times$31 equally spaced grid in both the x- and y-directions ($\Delta x=\Delta y$) with a CFL number of 0.02. An acoustic wave of SPL=130 dB is introduced at the inflow boundary. The acoustic pressure signals required to compute the SPL along the upper wall are collected after the transients leave the computational domain and the field becomes periodic. Numerical computations were performed at six different frequencies from 0.5 to 3.0 kHz at 0.5 kHz increments. Fourth-order spatial discretization is employed in both the x- and y-directions for the governing equations and second-order discretization is used for the broadband impedance boundary condition, as mentioned in Eq. (15).

Figure 12 shows the comparison of the upper wall SPL results for the current calculations with the measured data. The symbols in these figures indicate the experimental data. The agreement between the measurements and the current results is excellent. In Fig. 12, it is shown that sound absorption in an acoustic liner is a function of frequency and that the large absorption occurs at 1.0 kHz and the small absorption occurs at 2.0 kHz.



(a) Grass ground



(b) Wool ground

FIG. 11. (Color online) Comparison of excess attenuation of sound due to a monopole between computed result (symbols) and analytical solution (solid line). The source-receiver geometry is y_s=1.0, y_o=0.5, and range equals 1.0. (a) Grass ground; (b) wool ground.

V. SUMMARY AND CONCLUSIONS

In this paper, a broadband time-domain impedance boundary condition has been developed and validated. The frequency-domain impedance condition was represented as a linear sum of second-order frequency response functions, assuming that the impedance is independent of the location on the surface. This allowed the construction of a bandpass filter and a low-pass filter type response function as the approximation to the expensive convolution integral in the conventional time-domain impedance condition. This frequency response function utilizes the past pressure and velocity outputs, and the present acoustic pressure inputs recursively. Two-dimensional numerical experiments reveal that there is

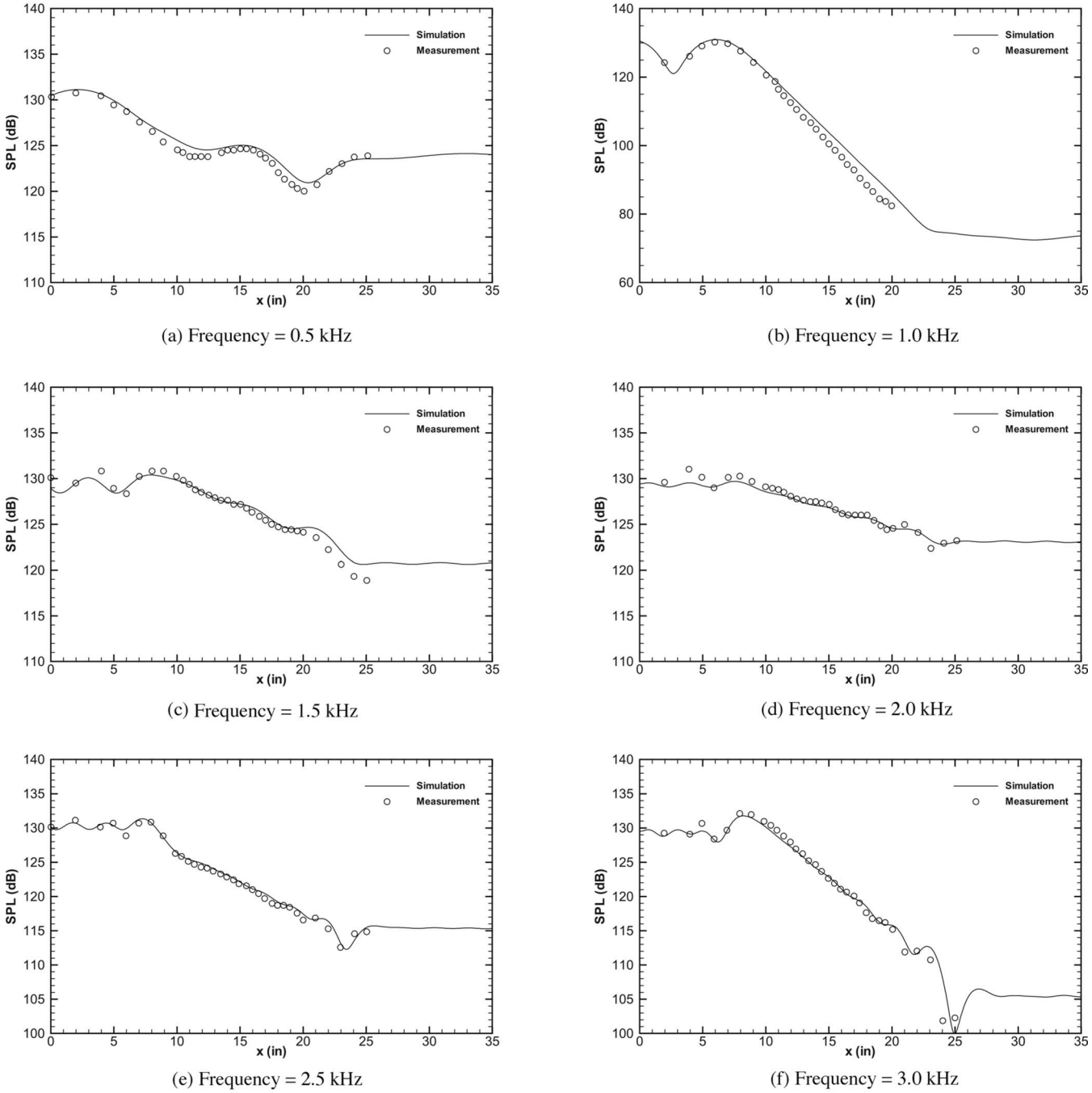


FIG. 12. Upper wall SPLs given by the single-frequency simulations. $M=0.1$: (a) $f=0.5$ kHz, (b) $f=1.0$ kHz, (c) $f=1.5$ kHz, (d) $f=2.0$ kHz, (e) $f=2.5$ kHz, and (f) $f=3.0$ kHz.

good agreement between the numerical results and analytical/experimental solutions and indicate that the present method is capable of accurately simulating the physical broadband phenomena over acoustically treated surfaces.

It stands to reason that the frequency-domain methods in situations where they are readily implementable may outperform the present time-domain impedance boundary condition or any time-domain approach. But then the present method is intended for real-world problems that may not be amenable to frequency-domain methods or current time-domain approaches. Having demonstrated the validity, robustness, and practicality of the method, its relative performance vis-à-vis other time-domain methods remains to be established in the context of acoustic simulations involving broadband frequencies, such as impulsive noise, turbulence noise, etc. It is but proper to mention that it may be difficult if not impossible to apply the current methodology to such problems. It will be the topic of future work.

ACKNOWLEDGMENT

J.B. gratefully acknowledges the financial assistance from the Office of the Provost.

APPENDIX A: BROADBAND-FREQUENCY IMPEDANCE FUNCTION

The curves shown in Figs. 4–6 were obtained by applying a conjugate gradient method[20,21] to the frequency-domain impedance function given by Eq. (10). A total of four FRFs were used in both cases of a grass ground and a wool-felt ground. A total of five FRFs were employed for the case of an acoustic liner. The parameters a_0^j to b_2^j of this equation were found to be

$$Z(\omega) = \sum_{j=1}^{4} \frac{a_0^j(i\omega) + a_1^j}{b_0^j(i\omega)^2 + b_1^j(i\omega) + b_2^j},$$

$$\omega = 2\pi f \quad \text{where } f \text{ is in Kilohertz.} \tag{A1}$$

1. The coefficients for the grass ground

$a_0^1 = 0.704\,173\,53/2\pi$, $a_1^1 = 0.387\,007\,09$,

$b_0^1 = 0.957\,422\,22/(2\pi)^2$, $b_1^1 = 0.274\,703\,90/2\pi$, $b_2^1 = 0.183\,504\,06$,

$a_0^2 = 0.771\,817\,02/2\pi$, $a_1^2 = 1.193\,269\,66$,

$b_0^2 = 0.144\,220\,47/(2\pi)^2$, $b_1^2 = 0.574\,349\,11/2\pi$, $b_2^2 = 0.404\,762\,86$,

$a_0^3 = 0.579\,089\,01/2\pi$, $a_1^3 = 1.311\,837\,80$,

$b_0^3 = 1.724\,860\,51 \times 10^{-2}/(2\pi)^2$, $b_1^3 = 0.340\,938\,39/2\pi$, $b_2^3 = 0.293\,515\,78$,

$a_0^4 = 0.500\,873\,21/2\pi$, $a_1^4 = 1.162\,788\,45$,

$b_0^4 = 0.706\,389\,34/(2\pi)^2$, $b_1^4 = 0.877\,132\,11/2\pi$, $b_2^4 = 1.393\,761\,61 \times 10^{-3}$.

2. The coefficients for the wool-felt material ground

$a_0^1 = 0.840\,338\,25/2\pi$, $a_1^1 = 0.730\,370\,99$,

$b_0^1 = 3.400\,127\,98/(2\pi)^2$, $b_1^1 = 0.965\,755\,94/2\pi$, $b_2^1 = 0.234\,801\,97$,

$a_0^2 = 1.692\,764\,18/2\pi$, $a_1^2 = 2.358\,337\,75$,

$b_0^2 = 5.064\,137\,35/(2\pi)^2$, $b_1^2 = 5.960\,787\,85/2\pi$, $b_2^2 = 3.448\,883\,64$,

$a_0^3 = 2.978\,427\,78/2\pi$, $a_1^3 = 2.118\,550\,71$,

$b_0^3 = 0.972\,494\,18/(2\pi)^2$, $b_1^3 = 1.476\,833\,16/2\pi$, $b_2^3 = 0.200\,666\,12$,

$a_0^4 = 0.506\,580\,59/2\pi$, $a_1^4 = 2.442\,884\,62$,

$b_0^4 = 2.142\,249\,30 \times 10^{-2}/(2\pi)^2$, $b_1^4 = 0.534\,803\,50/2\pi$, $b_2^4 = 1.395\,599\,61$.

3. The coefficients for the ceramic tubular liner (CT73)

$a_0^1 = 15.142\,262\,85$, $a_1^1 = 2.201\,562\,75$,

$b_0^1 = 8.958\,174\,74$, $b_1^1 = 5.807\,517\,92$, $b_2^1 = 0.472\,299\,49$,

$a_0^2 = 1.429\,013\,44$, $a_1^2 = 2.352\,884\,52$,

$b_0^2 = 9.351\,699\,97 \times 10^{-2}$, $b_1^2 = 0.294\,350\,28$, $b_2^2 = 15.251\,511\,51$,

$a_0^3 = 2.509\,711\,50$, $a_1^3 = 1.319\,491\,80$,

$b_0^3 = 5.325\,179\,34$, $b_1^3 = 1.907\,520\,80$, $b_2^3 = 1.004\,487\,05$,

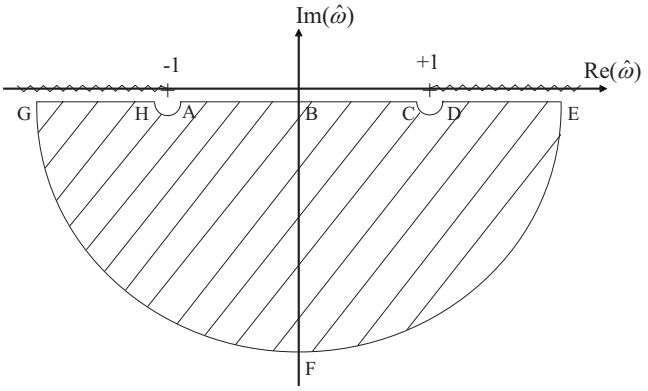


FIG. 13. Area of $\hat{\omega}$ plane.

$a_0^4 = 2.781\,151\,08$, $a_1^4 = 3.844\,384\,27$,

$b_0^4 = 0.753\,678\,58$, $b_1^4 = 1.474\,136\,38$, $b_2^4 = 1.155\,048\,05$,

$a_0^5 = 2.628\,979\,68 \times 10^{-2}$, $a_1^5 = 5.682\,839\,61 \times 10^{-3}$,

$b_0^5 = 7.627\,031\,04 \times 10^{-4}$, $b_1^5 = 4.203\,121\,55 \times 10^{-3}$, $b_2^5 = 0.580\,376\,55$.

APPENDIX B: THE WELL-POSEDNESS OF THE BROADBAND FREQUENCIES FOR IMPEDANCE SURFACES

Assuming the impedance surface is located in the x-z plane and applying the Fourier–Laplace transform to the governing equations, which are the linearized Euler equations, it can be found that the solution satisfies the outgoing wave condition at $y \to \infty$ as follows: Consider the separable solutions

$$\begin{bmatrix} p(x,y,z,t) \\ u(x,y,z,t) \\ v(x,y,z,t) \\ w(x,y,z,t) \end{bmatrix} = \begin{bmatrix} \tilde{p}(y) \\ \tilde{u}(y) \\ \tilde{v}(y) \\ \tilde{w}(y) \end{bmatrix} \times e^{i(\Omega t - \alpha x - \beta z)}. \tag{B1}$$

By substituting these equations into the governing equations, the corresponding solutions can be obtained in the following form:

$$\begin{bmatrix} \tilde{p}(y) \\ \tilde{u}(y) \\ \tilde{v}(y) \\ \tilde{w}(y) \end{bmatrix} = A \begin{bmatrix} 1 \\ \alpha/\Omega \\ -(\hat{\omega}^2 - 1)^{1/2}/\hat{\omega} \\ \beta/\Omega \end{bmatrix} \times e^{ik(\hat{\omega}^2 - 1)^{1/2} y}, \tag{B2}$$

where $\hat{\omega} = \Omega/k$, $k = (\alpha^2 + \beta^2)^{1/2}$, and $A = \text{const}$. The branch cuts of the function $(\hat{\omega}^2 - 1)^{1/2}$ are taken to be $0 \leqslant \arg(\hat{\omega}^2 - 1)^{1/2} \leqslant \pi$, as shown in Fig. 13.

The Fourier–Laplace transforms of the impedance condition are expressed as

$$\tilde{p} = -\frac{a_0 + a_1/(i\Omega)}{b_0(i\Omega) + b_1 + b_2/(i\Omega)} \tilde{v}. \tag{B3}$$

Substitution of Eq. (B2) into Eq. (B3) leads to the following dispersion relation:

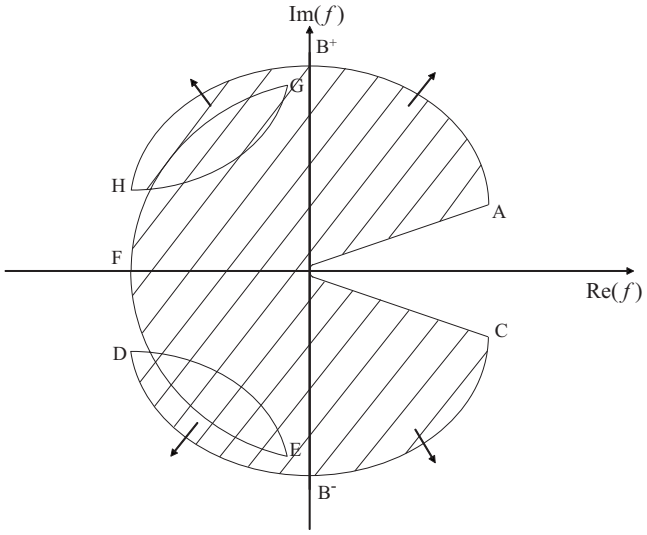


FIG. 14. Map of the lower half of $\hat{\omega}$ in the $f(\hat{\omega})$ plane.

$$\left(kb_0(i\hat{\omega}) + b_1 + \frac{b_2/k}{i\hat{\omega}}\right)\left(\frac{\hat{\omega}}{(\hat{\omega}^2-1)^{1/2}}\right) + \frac{a_1}{k\hat{\omega}}i = a_0. \quad \text{(B4)}$$

In this equation all of the coefficients of a's and b's are real, positive numbers. This boundary treatment is well posed if this equation has no solutions in the lower half of the $\hat{\omega}$ plane depicted in Fig. 13. Let the left-hand side of this equation be expressed by $f(\hat{\omega})$.

$$f(\hat{\omega}) = \left(kb_0(i\hat{\omega}) + b_1 + \frac{b_2/k}{i\hat{\omega}}\right)\left(\frac{\hat{\omega}}{(\hat{\omega}^2-1)^{1/2}}\right) + \frac{a_1}{k\hat{\omega}}i. \quad \text{(B5)}$$

Figure 14 shows the map of the lower half $\hat{\omega}$ plane in the f plane. For the case of $a_0 > 0$, there is no value of $\hat{\omega}$ in the lower half of the $\hat{\omega}$ plane that can satisfy the dispersion relation of Eq. (B4), since the right-hand side of Eq. (B4) is real and positive. So it can be proven that there is no stability problem in the case where all of the coefficient values are real and positive.

[1] A. D. Pierce, *Acoustics—An Introduction to Its Physical Principles and Applications* (Acoustical Society of America, New York, 1989).
[2] Y. L. Li and M. J. White, "Near-field computation for sound propagation above ground-using complex image theory," J. Acoust. Soc. Am. **99**, 755–760 (1996).
[3] X. Di and K. E. Gilbert, "An exact laplace transform formulation for a point source above a ground surface," J. Acoust. Soc. Am. **93**, 714–720 (1993).
[4] S. Davis, "Low-dispersion finite difference methods for acoustic waves in a pipe," J. Acoust. Soc. Am. **90**, 2775–2781 (1991).
[5] D. Botteldooren, "Acoustical finite-difference time-domain simulation in a quasi-cartesian grid," J. Acoust. Soc. Am. **95**, 2313–2319 (1994).
[6] C. K. W. Tam and L. Auriault, "Time-domain impedance boundary conditions for computational aeroacoustics," AIAA J. **34**, 917–923 (1996).
[7] S. Zheng and M. Zhuang, "Three-dimensional benchmark problem for broadband time-domain impedance boundary conditions," AIAA J. **42**, 405–407 (2004).
[8] S. Zheng and M. Zhuang, "Verification and validation of time-domain impedance boundary condition in lined duct," AIAA J. **43**, 306–313 (2005).
[9] Y. Özyörük, L. N. Long, and M. G. Jones, "Time-domain numerical simulation of a flow-impedance tube," J. Comput. Phys. **146**, 29–57 (1998).
[10] Y. Özyörük and L. N. Long, "A time-domain implementation of surface acoustic impedance condition with and without flow," J. Comput. Phys. **5**, 277–296 (1997).
[11] K.-Y. Fung and H. Ju, "Broadband time-domain impedance models," AIAA J. **39**, 1449–1454 (2001).
[12] H. Ju and K.-Y. Fung, "Time-domain impedance boundary conditions with mean flow effects," AIAA J. **39**, 1683–1690 (2001).
[13] H. Ju and K.-Y. Fung, "Time-domain simulation of acoustic sources over an impedance plane," J. Comput. Acoust. **10**, 311–329 (2002).
[14] D. K. Wilson, V. E. Ostashev, S. L. Collier, N. P. Symons, D. F. Aldridge, and D. H. Marlin, "Time-domain calculations of sound interactions with outdoor ground surfaces," Appl. Acoust. **68**, 175–200 (2007).
[15] K. Attenborough, "Ground parameter information for propagation modeling," J. Acoust. Soc. Am. **92**, 418–427 (1992).
[16] T. L. Parrott, W. R. Watson, and M. G. Jones, "Experimental validation of a two-dimensional shear-flow model for determining acoustic impedance," Technical Report No. TP-2679 (NASA, Washington, DC, 1987).
[17] M. K. Myers, "On the acoustic boundary condition in the presence of flow," J. Sound Vib. **71**, 429–434 (1980).
[18] U. Ingard, "Influence of fluid motion past a plane boundary on sound reflection, absorption, and transmission," J. Acoust. Soc. Am. **31**, 1035–1036 (1959).
[19] S. W. Rienstra, "Impedance models in time domain including the extended helmholtz resonator model," 12th AIAA/CEAS Aeroacoustics Conference, Cambridge, MA, 8–10 May 2006.
[20] W. H. Press, B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling, *Numerical Recipes in FORTRAN 77: The Art of Scientific Computing*, 2nd ed. (Cambridge University Press, Cambridge, 1992).
[21] E. K. P. Chong and S. H. Zak, *An Introduction to Optimization*, 2nd ed. (Wiley, New York, 2001).
[22] H. E. Hayder, F. Q. Hu, and M. Y. Hussaini, "Towards perfectly absorbing boundary conditions for the Euler equations," AIAA J. **37**, 3135–3144 (1999).
[23] F. Q. Hu, "On absorbing boundary conditions for the linearized Euler equations by a perfectly matched layer," J. Comput. Phys. **129**, 201–219 (1996).
[24] F. Q. Hu, "A stable, perfectly matched layer for linearized Euler equations in unsplit physical variables," J. Comput. Phys. **173**, 455–480 (2001).
[25] S. Abarbanel, D. Stanescu, and M. Y. Hussaini, "Unsplit variables perfectly matched layers for the shallow water equations with Coriolis forces," Comput. Geosci. **7**, 275–294 (2003).
[26] C. K. W. Tam and J. C. Webb, "Dispersion-relation-preserving finite difference schemes for computational aeroacoustics," J. Comput. Phys. **107**, 262–281 (1993).
[27] F. Q. Hu, M. Y. Hussaini, and J. L. Manthey, "Low-dissipation and low-dispersion Runge–Kutta schemes for computational acoustics," J. Comput. Phys. **124**, 177–191 (1996).

Transition scattering in stochastically inhomogeneous media

V. Pavlov[a] and E. P. Tito
California Institute of Technology, Mail Code 252-21, Pasadena, California 91125

(Received 30 April 2008; revised 3 November 2008; accepted 6 December 2008)

When a physical object ("a source") without its own eigenfrequency moves through an acoustically homogeneous medium, the only possible form of acoustic radiation is the emission of Mach shock waves, which appear when the source velocity surpasses sonic speed. In nonhomogeneous media, in nonstationary media, or in the neighborhood of such media, the source motion is accompanied by the so-called "transition" radiation (diffraction or scattering), which has place even when the source moves with subsonic velocity. Key features pertaining to the formation of the acoustical transition scattering in media with fluctuating acoustical parameters are established. To analytically study the effect, the Green's function method formulated in terms of functional derivatives is used. The relationship between the wave number and frequency, $k=k(\omega)$, for acoustic waves is found. The results serve to determine the phasing conditions necessary for opening the transition scattering and Cherenkov radiation channel and to establish the physical explanation for the phenomenon—scattering (transformation) on inhomogeneities of the accompanied source field; i.e., formation of radiation appears when the attached field readjusts back to the equilibrium state after being deformed while passing through the fluctuations of the medium.
 [DOI: 10.1121/1.3058633]

PACS number(s): 43.28.Ra, 43.28.Mw, 43.20.Bi, 43.20.Px [RMW] Pages: 676–689

I. INTRODUCTION

This paper investigates the phenomenon of transition radiation (scattering) of acoustical waves by an object (not possessing its own proper frequency)[1] uniformly moving in a stochastically nonhomogeneous medium in sub- and supersonic regimes.

Generation of waves caused by the source moving in various types of medium has attracted particular interest because of the wide range and importance of its applications to acoustics, optics, geophysics, and other areas of physics and due to its critical role in the broader theory of wave propagation (see, for example, Refs. 2–5 and references therein). It has been established that if the source moves uniformly at a supersonic speed in a homogeneous medium, it generates what is called "Cherenkov radiation." If the source moves at a constant but subsonic speed, it generates waves if it crosses the interface of two media with different properties ("transition radiation") or if it moves near some boundary ("transition diffraction") (see Ref. 6 and references therein). The source can also radiate if it accelerates.

To outline the geometry of the effect, consider Fig. 1 commonly used to describe Cherenkov radiation.[7,8] Figure 1 illustrates a point source moving from point A toward point B with constant velocity $\mathbf{V}$. At point A the source generates a wave with phase speed in medium c. By the time the source reaches point B at distance Vt, the spherical wave from point A propagates at distance ct.

Next look at line A-D in the direction of wave vector $\mathbf{k}$ of a field spectral component. The phase difference $\Delta\psi(\omega)$ between spherical waves, $\sim\exp(-i\omega t+ikr)/r$, generated at points A and B and observed at "infinity" along angle θ to the trajectory of the body, is given by expression $\Delta\psi=k(DA-CA)=k(Vt\cos\theta-ct)$ because $\psi^B(\omega)=\psi^C(\omega)$. At large distances, we can neglect the difference between $r_{A\infty}$ and $r_{B\infty}$ when considering amplitudes, but for phase relationships this distinction is essential. The waves do not cancel each other at infinity if $\Delta\psi=k(V\cos\theta-c)t\ll\pi$. For any t, this condition is realized *only* for $\cos\theta=c/V$, i.e., when $V>c$. This is the Cherenkov effect.

However, if the medium properties are not uniform, but rather stochastically fluctuating, the phases of waves at points C and B (Fig. 1) do not necessarily have the same values as above. Therefore, the phasing conditions for superposing waves at infinity may not hold. To derive the phase difference leading to radiation requires special calculations incorporating information about the medium.

Figure 2 illustrates the physics behind the formation of the "transition scattering." On the left side of Fig. 2, our source moves uniformly in a homogeneous medium and is accompanied by an attached wave field,[9] distributed in accordance with the properties of the surrounding region. It is precisely this attached wave field and not the source itself (whether a point or nonpoint) that leads to the additional radiation [circle (1) represents the surfaces of equal stress]. As the source travels through region D filled with inhomogeneities, such that the equilibrium parameters of the medium differ from the values describing the homogeneous region, the attached field becomes deformed by the fluctuating parameters within the region [circle (2)]. Past the inhomogeneous region, the attached field begins to rearrange itself toward the configuration corresponding to the equilibrium parameters of the surrounding medium [circle (3)]. Because the energy and source velocity remain constant in this process, the rearrangement of the attached field gives rise to an

[a]Also at UFR de Mathématiques Pures et Appliquées, Université de Lille 1, 59655 Villeneuve d'Ascq, France.

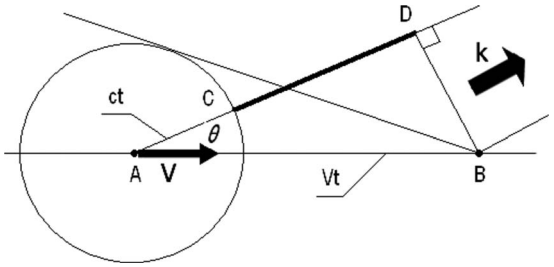


FIG. 1. Phasing condition for Cherenkov radiation.

additional field, namely, the radiation field. The resulting radiation is the transition scattering. Clearly transition scattering can be viewed as a subset of transition (transformation) radiation.

The distinction in the nature of transition scattering and transition radiation can also be seen in the difference of the process durations. Transition scattering is a continuous process caused by the rearrangement of the attached field, and therefore, it lasts effectively infinitely. Transition radiation, on the other hand, has a short-term impulselike character. It occurs when the object transits through the variation in the medium. While extensive literature[7,10] has addressed both transition radiation and transition scattering problems in electrodynamics, acoustical transition scattering has not been investigated to such depths. Problems of acoustic transition radiation have been surveyed in detail by Pavlov and Sukhorukov.[6] The problem of transition scattering in a turbulent medium (i.e., the one where the medium fluctuations are caused by fluctuations in velocities of fluid particles), but not the medium with fluctuating state parameters (such as sound speed, pressure, or density), was considered by Pavlov.[11] The effect of transition scattering near a rough surface was analyzed by Pavlov and Sukhorukov.[12] Lipovskii and Tamoikin[13] studied a specific model for the fluctuating parameter medium using a model relationship between the Fourier transform of the average momentum per unit volume of the medium and the velocity Fourier component. Our current paper aims at establishing a clear and unambiguous description, in terms of the Green's function, for the radiation field created by a moving source in stochastically fluctuating media with varying sound speed. Along with finding the characteristics of radiation by sources moving in fluctuating media, the development of the functional Green's function method produces a general methodology that can be applied to a variety of problems with differing specifics and assumptions.

The remainder of this paper is organized as follows. In Sec. II we define our model and the basic equations and address intensity, energy flux, and energy density relationships in the frequency domain. We propose a simple method (Sec. III) for analyzing wave propagation through a fluctuating medium, for calculating wave characteristics based on the Green's function method with functional derivatives, and for low-magnitude fluctuation approximations. We analytically calculate the dispersion relationship and coefficient of attenuation for the averaged component of the wave (acoustical) field. The results obtained at this step of our analysis serve to establish the phase conditions necessary for the opening of the Cherenkov radiation channel (Sec. IV). The angular-spectral power of the scattering radiation is considered in Sec. V. Section VI summarizes our results. In the medium with strongly fluctuating sound speed, the conditions for Cherenkov radiation can change drastically. The expression, obtained in this paper, shows that in such fluctuating medium the radiation channel opens for the *subsonic* Mach numbers, $M<1$. The shock wave with a *sharp* front does not form in this case because different spectral components are radiated under different angles. The relationship between the radiated angles for short and long waves makes experimental verification possible. This relationship should be taken into consideration when the power of transition radiation is derived.

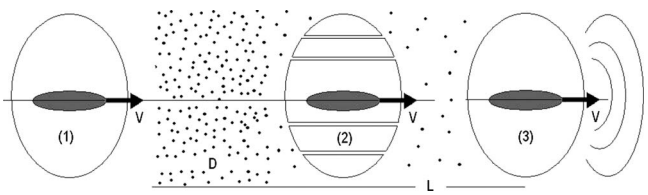


FIG. 2. Formation of transition scattering.

II. BASIC EQUATIONS AND ENERGETIC RELATIONS

As noted above, transition scattering radiation arises when a source moves through a medium whose properties are such that the speed of sound fluctuates. These fluctuations can be caused by a variety of natural phenomena. Frequently, sound speed fluctuations occur due to random variations of density, but such variations are typically small, and the resulting transition scattering effect is rather weak. Only near the phase transition points does the effect become significant. However, sound speed fluctuations may be rather significant in the medium of mixed nature such as when air bubbles are present in water (as in an upper oceanic layer or a jet wake). Then the transition scattering effect becomes much more pronounced. (Appendix A discusses both of the mentioned scenarios in more detail.)

Two factors determine the conditions under which a source moving in a fluctuating medium radiates acoustical waves—the dispersion relationship between wave number and frequency, which is described by the averaged Green's function, and the phasing condition between superposing emitted waves. To study the effect, we will start by considering the Green's function for a simple model (Appendix A). Consider the equation describing a scalar (for example, acoustical) field generated by a localized (point) source,

$$\Delta\phi - \frac{1}{c^2}(1+\epsilon(\mathbf{x}))\partial_{tt}\phi = F(t)\delta(\mathbf{x}-\mathbf{x}_0(t)). \tag{1}$$

Here, positions of the source and the receiver are defined by coordinates $\mathbf{x}_0$ and $\mathbf{x}$, respectively. Operator Δ is the three-dimensional (3D)-Laplacian operator, and $\delta(\mathbf{x}-\mathbf{x}_0)$ is the 3D-Dirac function used to describe the source as a point. If ϕ denotes standard velocity potential, then "observable" physical variables—acoustical pressure and velocity—are described by the rule $p_1=-\rho_0\partial_t\phi$, $\mathbf{v}_1=\nabla\phi$. The source term

describes a volume injection term or a thermic source.[6] The productivity of the source is characterized by function $F(t)$ (which is in units of volume per unit time). The local sound speed is defined by the state equation $c_{\mathrm{loc}}^2=(\partial p/\partial\rho)$. The wave speed profile $c_{\mathrm{loc}}(\mathbf{x})$ contains all information about the medium that is necessary to describe the process. We assume that the speed fluctuates: $c_{\mathrm{loc}}^{-2}=c^{-2}(1+\epsilon(\mathbf{x}))$. Quantity $\epsilon(\mathbf{x})$ is an acoustical fluctuating parameter. In principle, it can be any operator. Further, we will disregard temporal dependence of the variables due to fluctuations. The fluctuations are described by the zero average, $\langle\epsilon(\mathbf{x})\rangle=0$, and correlation function, $\langle\epsilon(\mathbf{x})\epsilon(\mathbf{x}')\rangle=B(\mathbf{x},\mathbf{x}')$, such that $B(0)=\langle\epsilon^2\rangle<1$. For a spatially homogeneous medium, the correlation function is a function of only coordinate difference, $B(\mathbf{x},\mathbf{x}')\equiv B(\mathbf{x}-\mathbf{x}')$. The correlation function is characterized by two parameters: the mean square fluctuation ($\langle\epsilon^2\rangle$) and the radius of correlation (l) describing the characteristic distance over which fluctuation correlation vanishes.[2,3] Obviously, more complex models can be constructed.

When multiplied by $\partial_t\phi$ and integrated with respect to volume, Eq. (1) becomes

$$\partial_t E' + \int d\mathbf{x}\,\mathrm{div}\,\mathbf{S}' = -\int d\mathbf{x}\partial_t\phi(\mathbf{x},t)F(t)\delta(\mathbf{x}-\mathbf{x}_0(t)), \tag{2}$$

where

$$E' = \frac{1}{2}\int d\mathbf{x}\left[(\nabla\phi)^2 + \frac{1}{c^2}(1+\epsilon(\mathbf{x}))(\partial_t\phi)^2\right] \tag{3}$$

and $\mathbf{S}'=-\partial_t\phi\nabla\phi$. Here, the field energy includes both the energy of free and attached to the source fields.[14] It also includes the energy of field interaction with fluctuations $E_{\mathrm{int}}=(2c^2)^{-1}\int d\mathbf{x}\epsilon(\mathbf{x})(\partial_t\phi)^2$. The term

$$A_r = -\int d\mathbf{x}\partial_t\phi(\mathbf{x},t)F(t)\delta(\mathbf{x}-\mathbf{x}_0(t)) \tag{4}$$

describes the work performed by the moving source against the radiation friction force.[15] Vector $\mathbf{S}'=-\partial_t\phi\nabla\phi$ defines the density of the energy flux. The integral

$$W(t) = \int d\mathbf{x}\,\mathrm{div}\,\mathbf{S} \equiv \oint_\Sigma d\mathbf{f}\cdot\mathbf{S} \tag{5}$$

defines the power of wave radiation. The surface integral is calculated over the "wave zone"—the integrable surface Σ placed at such a distance from the origin of coordinates (where the source moves but does not cross the surface) that the generated field has a structure of divergent spherical waves. For a chosen pulsation ω of a spectral component ϕ_ω, radius r of such a sphere is confined by the condition $c/\omega \ll r \ll \gamma^{-1}$. Here, γ is a dissipative factor (logarithmic decrement) since dissipation always exists in media. The time-averaged work of external forces against radiation damping is compensated by losses on radiation, $\bar{A}_r$, and the reorganization of the attached field of the source. The time averaging is defined by the integral $\overline{A(t)}=\lim_{T\to\infty}T^{-1}\int_0^T dtA(t)$.

Using Fourier transformation with respect to time, we obtain a time-averaged expression for the radiation power,

$$\begin{aligned}\bar{W} &= \lim_{T\to\infty}\frac{1}{T}\int dt\left[-\oint d\mathbf{f}\cdot\partial_t\phi\nabla\phi\right]\\ &= \lim_{T\to\infty}\frac{1}{T}\int\frac{d\omega}{2\pi}\frac{d\omega'}{2\pi}\int dt e^{-i(\omega-\omega')t}\\ &\quad\times\oint d\mathbf{f}\cdot(i\omega\phi_\omega\nabla\phi^*_{\omega'})\\ &= \int_0^{+\infty}d\omega\left[\lim_{T\to\infty}\frac{1}{2\pi T}\oint d\mathbf{f}\cdot(i\omega\phi_\omega\nabla\phi^*_\omega)+\mathrm{c.c.}\right].\end{aligned} \tag{6}$$

At large distances from the origin, the radiated intensity is defined as the amount of energy flowing per unit time through the area element on a spherical surface with radius r and centered at the origin. The surface integral can be transformed into an integral with respect to the solid angle $d\mathbf{f}[\cdots]=r^2d\Omega[\cdots]$, where $d\Omega=\sin\theta d\theta d\varphi$ is a solid-angle element. The angular-spectral density of radiation is defined by the following expression:

$$\bar{W}_{\omega,\mathbf{n}} = \lim_{T\to\infty}\frac{r^2}{T}\left[\frac{i}{2\pi}ck\phi_\omega\partial_r\phi^*_\omega+\mathrm{c.c.}\right], \tag{7}$$

normalized as $\bar{W}=\int_0^\infty d\omega\int d\Omega\bar{W}_{\omega,\mathbf{n}}$. Here, $k=\omega/c$, and $\mathbf{n}$ is the unit vector drawn from the origin to the point of observation in the radiation direction. When averaged with respect to the fluctuation, this expression gives the spectral density of radiation. It is clear that expression (7) describes two processes: Cherenkov and scattering radiations. In fact, the field is decomposed into the regular and fluctuating components, $\phi=\langle\phi\rangle+\phi'$, and, therefore, $\langle|\phi_\omega|^2\rangle=|\langle\phi_\omega\rangle|^2+\langle\phi'_\omega\phi'^*_\omega\rangle$. To simplify formula notations, below we will use the notation $\phi_\omega\equiv\phi_k$ to describe the temporal Fourier transform with $k=\omega/c$. (Do not confuse it with the spatial Fourier transformation!)

Thus, Cherenkov radiation is expressed via the formula

$$\langle\bar{W}_{\omega,\mathbf{n}}\rangle^{\mathrm{Ch}} = \lim_{T\to\infty}\frac{r^2}{T}\left[\frac{i}{2\pi}ck\langle\phi_k\rangle\partial_r\langle\phi^*_k\rangle+\mathrm{c.c.}\right], \tag{8}$$

and the scattering radiation is described by

$$\langle\bar{W}_{\omega,\mathbf{n}}\rangle^{\mathrm{sc}} = \lim_{T\to\infty}\frac{r^2}{T}\left[\frac{i}{2\pi}ck\langle\phi'_k\partial_r\phi'^{*}_k\rangle+\mathrm{c.c.}\right]. \tag{9}$$

All these quantities are calculated in the wave zone, and therefore, radius r has to be chosen to satisfy condition $k^{-1}\ll r\ll\gamma^{-1}$, where γ is the logarithmic decrement. In this case, $\partial_r\phi_k\simeq\Gamma_k\phi_k$. The analytical expressions for Γ and real part $\Re\Gamma$ are calculated in following sections. Equations (8) and (9) become

$$\langle\bar{W}_{\omega,\mathbf{n}}\rangle^{\mathrm{Ch}} \simeq \lim_{T\to\infty}\frac{r^2}{T}\left[\frac{1}{\pi}ck(\Re\Gamma_k)|\langle\phi_k\rangle|^2\right] \tag{10}$$

and

$$\langle\bar{W}_{\omega,\mathbf{n}}\rangle^{\mathrm{sc}} \simeq \lim_{T\to\infty}\frac{r^2}{T}\left[\frac{1}{\pi}ck(\Re\Gamma_k)\langle|\phi'_k|^2\rangle\right]. \tag{11}$$

III. GREEN'S FUNCTION

A. Averaged Green's function

Applying the Fourier transformation with respect to time to Eq. (1), we obtain a straightforward expression for a spectral component of the field,

$$\Delta\phi_k + k^2(1+\epsilon(\mathbf{x}))\phi_k = \hat{\mathcal{F}}_k[F(t)\delta(\mathbf{x}-\mathbf{x}_0(t))]. \tag{12}$$

Parameter $k=\omega/c$ denotes the wave number, and $\hat{\mathcal{F}}_k[\cdots]=\int dt[\cdots]\exp(+i\omega t)$ is the temporal Fourier transform of the argument. Our preliminary goal is to find the averaged field $\langle\phi_k\rangle$.

From Eq. (1), we find

$$\phi_k(\mathbf{x}) = \int d\mathbf{x}_1 G(\mathbf{x},\mathbf{x}_1)\hat{\mathcal{F}}_k[F(t)\delta(\mathbf{x}_1-\mathbf{x}_0(t))]. \tag{13}$$

Here, $G(\mathbf{x},\mathbf{x}_0)$ is the Green's function that satisfies

$$\Delta G(\mathbf{x},\mathbf{x}') + k^2(1+\epsilon(\mathbf{x}))G(\mathbf{x},\mathbf{x}') = \delta(\mathbf{x}-\mathbf{x}'). \tag{14}$$

The field follows from Eq. (13) by averaging with respect to fluctuations ϵ,

$$\langle\phi_k(\mathbf{x})\rangle = \int d\mathbf{x}_1\langle G(\mathbf{x},\mathbf{x}_1)\rangle\hat{\mathcal{F}}_k[F(t)\delta(\mathbf{x}_1-\mathbf{x}_0(t))]$$

$$= \hat{\mathcal{F}}_k[F(t)\langle G(\mathbf{x},\mathbf{x}_0(t))\rangle]. \tag{15}$$

To find the averaged Green's function $\langle G\rangle$, there exists a number of different methods.[2–4,16] Using the method based on functional derivatives (see Appendix B), we derive

$$\langle G(|\mathbf{r}-\mathbf{r}_0|)\rangle = \int \frac{d\mathbf{q}}{(2\pi)^3} e^{i\mathbf{q}\cdot|\mathbf{r}-\mathbf{r}_0|} \times\left[k^2-q^2-\int d\mathbf{x}\Sigma(\mathbf{x})\exp(-i\mathbf{q}\cdot\mathbf{x})\right]^{-1}. \tag{16}$$

Kernel $\Sigma(\mathbf{x},\mathbf{z})$ contains only irreducible diagrams (Fig. 3, see Appendix B and, for example, Ref. 3, p. 358). For isotropic and homogeneous medium, $\Sigma(\mathbf{x})=\Sigma(r)$, where $r=|\mathbf{x}|$. In a spherical coordinate system, after integrating over the angles, Eq. (16) takes the following form:

$$\langle G(R)\rangle = \frac{1}{i4\pi^2 R}\int_{-\infty}^{+\infty} dqqe^{iqR} \times\left[k^2-q^2-4\pi q^{-1}\int_0^\infty drr\Sigma(r)\sin qr\right]^{-1}. \tag{17}$$

Further analysis requires knowing the explicit form for $\Sigma(r)$. If we retain only the leading first term of Eq. (B13),[17] which is proportional to $\langle\epsilon^2\rangle$, we obtain

$$\Sigma(r) \simeq k^4 B(r) G_0(r). \tag{18}$$

B. Approximation for low-magnitude fluctuations

The poles of the integrand function in Eq. (17) determine the effective wave number of the mean field. They can be found by solving

$$k^2 - q^2 + \frac{k^4}{q}\int_0^\infty drB(r)e^{ikr}\sin qr = 0. \tag{19}$$

For small $\langle\epsilon^2\rangle$, the equation is solved by iterations. In the zero-order approximation with respect to $\langle\epsilon^2\rangle$, we have $q^{(0)}=k$. The next-order approximation is found from

$$k^2 - q^{(1)2} + k^3\int_0^\infty drB(r)e^{ikr}\sin kr = 0, \tag{20}$$

which gives

$$\Gamma \equiv q^{(1)} = k\left[1+\frac{k}{4}\int_0^\infty drB(r)\sin 2kr + i\frac{k}{2}\int_0^\infty drB(r)\sin^2 kr\right] + \cdots. \tag{21}$$

Here, we used $(1+2x)^{1/2}\simeq 1+x+\cdots$ for small x. For practical applications, it is convenient to use Fourier transforms of the correlation function: $B(\mathbf{s})=\int d\mathbf{x}B(r)e^{-i\mathbf{s}\cdot\mathbf{x}}$. After simple calculations in spherical coordinates, we find that $B(\mathbf{s})\equiv B(s)$,

$$B(s) = \frac{4\pi}{s}\int_0^\infty drrB(r)\sin sr \rightleftharpoons$$

$$B(r) = \frac{1}{2\pi^2 r}\int_0^\infty dssB(s)\sin sr. \tag{22}$$

By substituting $B(r)$ from Eq. (22) into Eq. (19), we obtain expressions where the following integrals are present: $I_1=\int_0^\infty drr^{-1}\sin sr\sin 2kr$, $I_2=\int_0^\infty drr^{-1}\sin sr\sin^2 kr$. These integrals indeed *can* be calculated analytically, and their calculation presents some methodological interest. Let us use the fact that $\int_0^\infty dr\exp(i\xi-\alpha)r=(\alpha-i\xi)^{-1}$ for $\alpha>0$. Consider $\alpha\to+0$. In this case, we can write that

$$\int_0^\infty dre^{i(2k\pm s)r} = \frac{1}{+0-i(2k\pm s)} \rightarrow$$

$$\int_0^\infty dr\sin sre^{i2kr} = \frac{1}{2i}\left[\frac{1}{+0-i(2k+s)} - \frac{1}{+0-i(2k-s)}\right].$$

By separating the real and imaginary parts of the integral, and calculating with respect to parameter k, we find

$$\int_0^\infty \frac{dr}{r}\sin sr\sin 2kr = \Re\frac{1}{2}\ln\left[\frac{(+0-i(2k+s))}{(+0-i(2k-s))}\frac{(+0+is)}{(+0-is)}\right] = \frac{1}{2}\ln\left|\frac{(2k+s)}{(s-2k)}\right|,$$

$$\int_0^\infty \frac{dr}{r}\sin sr\sin^2 kr = \Im\frac{1}{4}\ln\left[\frac{(+0-i(2k+s))}{(+0-i(2k-s))}\frac{(+0+is)}{(+0-is)}\right] = -\frac{1}{4}\arg(2k-s) = \frac{\pi}{4}H(2k-s). \tag{23}$$

Here, $H(z)$ is the Heaviside function: $H(z)=1$ for $z>0$ and $H(z)=0$ for $z<0$. We used the fact that $\ln z=\ln|z|+i\arg z$ and $\arg(s-2k)=-\pi$ for $s<2k$ because the branch point is at $s=2k+i0$, which we must contour *clockwise*. By substituting

Eq. (22) into Eq. (19) and using Eq. (23), we find that expression (19) takes form

$$\Gamma \equiv K_k + i\gamma_k \equiv k[1+\kappa_k + i\delta_k] = k\left[1 + \frac{k}{8\pi^2}\frac{1}{2}\int_0^\infty dssB(s)\ln\left|\frac{2k+s}{2k-s}\right| + i\frac{k}{4\pi^2}\frac{\pi}{4}\int_0^{2k} dssB(s)\right]. \quad (24)$$

C. Dispersion relationship and coefficient of attenuation

By combining Eqs. (24) and (17), we find that the averaged Green's function is approximated by

$$\langle G(\mathbf{x}-\mathbf{z})\rangle \simeq -\frac{1}{4\pi}e^{-\gamma_k|\mathbf{x}-\mathbf{z}|}\left.\frac{\exp(+iK_k|\mathbf{x}-\mathbf{z}|)}{|\mathbf{r}-\mathbf{z}|}\right|_{r\gg|\mathbf{z}|} \to -\frac{1}{4\pi}e^{-\gamma_k r}\frac{\exp(+iK_k r)}{r}\exp(-iK_k\mathbf{n}\cdot\mathbf{z}). \quad (25)$$

The dispersion relationship between the wave number and frequency and the attenuation coefficient of the averaged field amplitude are given by

$$K_k = k\left[1 + \frac{k}{8\pi^2}\frac{1}{2}\int_0^\infty dssB(s)\ln\left|\frac{2k+s}{2k-s}\right|\right],$$

$$\gamma_k = \frac{1}{16\pi}k^2\int_0^{2k} dssB(s). \quad (26)$$

Here, $k=\omega/c$. Notice that if there is dissipation, then there is also dispersion: different spectral components of the averaged field propagate with different phase speeds. Subsequent calculations require knowing the exact structure of the correlation function. However, we do know some general properties of this function and can obtain certain insights without such precise expressions. Any correlation function $B(r)$ has a maximum at $r=0$, vanishes when $r\to\infty$, and is characterized by two parameters: the mean square fluctuation, $\langle\epsilon^2\rangle$, and the correlation radius, l, describing the characteristic distance over which fluctuations are no more correlated. As an example, consider $B(s)=\pi^{3/2}\langle\epsilon^2\rangle l^3\exp(-s^2l^2/4)$. This expression corresponds to the correlation function $B(r)=\langle\epsilon^2\rangle\exp(-r^2/l^2)$ commonly used in many practical situations.[18] Simple calculations lead to the following expression for the attenuation coefficient:

$$\gamma_k = \frac{\sqrt{\pi}}{8}\langle\epsilon^2\rangle lk^2(1-\exp(-k^2l^2)). \quad (27)$$

Its limit cases are $\gamma_k=(\sqrt{\pi}/8)\langle\epsilon^2\rangle l^3k^4$ for $k^2l^2\ll 1$ and $\gamma_k=(\sqrt{\pi}/8)\langle\epsilon^2\rangle lk^2$ for $k^2l^2\gg 1$. In this case, the dispersion relationship (wave number K_k of the propagating wave expressed as a function of pulsation $k(=\omega/c)$) has the following form:

$$K_k = k + \frac{k^2}{8\pi^2}\frac{1}{2}\pi^{3/2}\langle\epsilon^2\rangle l^3\int_0^\infty dsse^{-s^2l^2/4}\ln\left|\frac{2k+s}{2k-s}\right| = k + \frac{k^4}{4\sqrt{\pi}}\langle\epsilon^2\rangle l^3\int_0^\infty dxxe^{-(kl)^2x^2}\ln\left|\frac{1+x}{1-x}\right|. \quad (28)$$

This integral can be calculated analytically. However, its expression in terms of special functions is very cumbersome. For this reason, we will analyze the behavior of this function only in two limit cases—for long and short waves—and will derive approximate interpolation expressions.

Consider $kl\ll 1$ (long waves). The principal contribution to the integral in Eq. (28) comes from the interval $0<x<(kl)^{-1}$ because exponent $e^{-(kl)^2x^2}$ is a rapidly decreasing function for $x>(kl)^{-1}$. For this reason, we present the integral in Eq. (28) as a sum of two integrals: $\int_0^\infty dx\cdots=\int_0^M dx\cdots+\int_M^\infty dx\cdots\equiv I_1+I_2=I$, where parameter M satisfies the condition $1\ll M<(kl)^{-1}$. In the first integral, I_1, we can replace the exponential $e^{-(kl)^2x^2}$ with 1. In the second integral, I_2, where $x>M\gg 1$, we can write $\ln|(1+x)/(1-x)|\simeq 2/x$. After this simplification both integrals can be calculated analytically. In fact, the first integral gives[19] $I_1=M+\frac{1}{2}(M^2-1)\ln[(M+1)/|M-1|)]\simeq 2M-(2/3M)$. The second gives $2\int_M^\infty dxe^{-(kl)^2x^2}\simeq e^{-(kl)^2M^2}/(kl)^2M\simeq 1/(kl)^2M$. The sum I_1+I_2 is a weakly dependent function of M when $\partial_M I=0$. This helps find $M=(\sqrt{2}kl)^{-1}\gg 1$. Finally, collecting the results of calculations leads to $K_k\simeq k+a_1\langle\epsilon^2\rangle l^2k^3$, which differs drastically from the linear dependence $K_k=k\equiv\omega/c$ existing for a nonfluctuating medium. Here, $a_1=3/(4\sqrt{2\pi})$ is a numerical constant.

If $kl\gg 1$ (short waves), the integral is approximately evaluated as $\simeq(kl)^{-3}$. Indeed, the principal contribution comes from the domain $0<x<(kl)^{-1}\ll 1$, where $\ln|(1+x)/(1-x)|\simeq 2x$. The integral is $2\int dxx^2\exp(-(kl)^2x^2)=\frac{\sqrt{\pi}}{2}(kl)^{-3}$ (see Ref. 20). By collecting the coefficients, we find that the dispersion relationship is written as $K_k=k[1+a_2\langle\epsilon^2\rangle]$, where $a_2=1/8$ is a numerical constant. The obtained expression shows that the averaged component of an acoustical wave propagates in a fluctuating medium with smaller speed, $c_{\rm ph}=c/[1+a_2\langle\epsilon^2\rangle]<c$, than if the medium does not fluctuate. This effect can be interpreted as scattering and superposition of multiple waves interacting with inhomogeneities.

A close approximation covering both limit cases can be obtained by the following interpolation of Eq. (28):

$$K_k \equiv k(1+\delta_k) = k\left[1 + \langle\epsilon^2\rangle\frac{a_1a_2(kl)^2}{a_2+a_1(kl)^2}\right]. \quad (29)$$

The phase speed of a regular component of a scalar field is defined thus from the approximate expression $c^{-1}(k)=c^{-1}[1+\langle\epsilon^2\rangle a_1a_2(kl)^2/(a_2+a_1(kl)^2)]$.

IV. CHERENKOV RADIATION

Let us now use the obtained results to establish the key conditions for the appearance of Cherenkov radiation for a general case with an inhomogeneous medium. Consider radiation by a point source with constant productivity F_0 trav-

eling at constant velocity $V=cM$ (M is the Mach number) along the x-axis. In this case, the right-hand side of Eq. (1) can be written as $V^{-1}F_0\delta_\perp(\mathbf{x}_\perp)\exp(iM^{-1}kx)$; i.e., the basic equation has the form

$$\Delta\phi_k + k^2(1+\epsilon(\mathbf{x}))\phi_k = \frac{1}{Mc}F_0\delta_\perp(\mathbf{x}_\perp)e^{iM^{-1}kx}. \tag{30}$$

Combining Eqs. (15) and (25), we find that the averaged field component is proportional to the Dirac function,

$$\langle\phi_k(\mathbf{x})\rangle \simeq -\frac{F_0}{4\pi Mc}e^{-\gamma_k r}\frac{e^{+iK_k r}}{r}\int dx e^{i(-K_k\cos\theta+M^{-1}k)x}$$

$$\simeq -\frac{F_0}{2Mc}\frac{e^{+iK_k r}}{r}\delta(-K_k\cos\theta+M^{-1}k), \tag{31}$$

since $\int_{-\infty}^{\infty}dx\exp isx=2\pi\delta(s)$. Here, θ is the direction of radiation defined by $\mathbf{K}_k\cdot\mathbf{x}=K_kx\cos\theta$. The presence of the delta function in the right side of the expression indicates that wave radiation takes place only if the Dirac function argument is zero. This requirement determines the phasing condition for possible radiation directions in a fluctuating medium, $V\cos\theta_k=c(k)$, resulting in

$$\cos\theta_k = \frac{c}{V}\left[1+\langle\epsilon^2\rangle\frac{a_1a_2(kl)^2}{a_2+a_1(kl)^2}\right]^{-1}; \tag{32}$$

i.e., spectral components with different frequencies radiate at different angles. In the limit case when the medium is uniform and has no fluctuations, i.e., when $\langle\epsilon^2\rangle\to 0$, we naturally obtain the classical result: $\cos\theta_k=M^{-1}$ for any spectral components.

The angular-spectral power of Cherenkov radiation is calculated from Eq. (8),

$$\langle\bar{W}_{\omega,\mathbf{n}}\rangle^{\mathrm{Ch}} = \lim_{T\to\infty}\frac{r^2}{T}\left[\frac{1}{\pi}ckK_k|\langle\phi_k\rangle|^2\right], \tag{33}$$

where Eq. (31) is used. By computing, we obtain the expression that is proportional to $\lim_{T\to\infty}(1/T)\delta^2(a)$. Here, $\delta^2(a)$ is the square of the delta function. Following Landau and Lifshitz[21], we can rewrite this expression as $\delta^2(a)=\delta(a)\times(2\pi)^{-1}\lim_{T\to\infty}\int_{-T/2}^{+T/2}dte^{iat}$ by decomposing one of the delta functions into the Fourier integral. Because of the presence of the delta function, the argument in the exponential can be written as zero; i.e., the exponential becomes replaced by 1. Thus, $\lim_{T\to\infty}(1/T)\delta^2(a)=\lim_{T\to\infty}(1/T)(T/2\pi)\delta(a)$, and the infinite time disappears. This result has the simple physical meaning: if a body travels infinitely long, it radiates the infinite amount of energy, but the energy radiated per unit of time (power) is obviously finite and physically meaningful.

We obtain the simple expression

$$\langle\bar{W}_{\omega,\mathbf{n}}\rangle^{\mathrm{Ch}} = \frac{1}{2(2\pi)^2M}kK_k|F_0|^2\delta(-K_k\cos\theta+M^{-1}k)$$

$$= \frac{1}{2(2\pi)^2M}k|F_0|^2\delta\left(\cos\theta-\frac{k}{MK_k}\right) \tag{34}$$

if it is remembered that K_k is given by Eq. (29) and $\delta^2(s)=\lim_{L\to\infty}(2\pi)^{-1}L\delta(s)$, with $L=VT=McT$. Integrated with respect to the solid angle $d\Omega=d\varphi d\theta\sin\theta$, this expression gives

$$\langle\bar{W}_\omega\rangle^{\mathrm{Ch}} = \frac{1}{4\pi M}k|F_0|^2H\left(M-\frac{k}{K_k}\right). \tag{35}$$

V. TRANSITION SCATTERING

To find the angular-spectral power of scattering radiation, it is necessary to derive an expression for the fluctuating component of the field. Using Eqs. (13)–(15), we obtain that at great distances from the source

$$\phi'_k(\mathbf{x}) = \hat{\mathcal{F}}_k[F(t)G'(\mathbf{x},\mathbf{x}_0(t))]$$

$$\simeq \int d\mathbf{y}G_0(\mathbf{x}-\mathbf{y})[-k^2\epsilon(\mathbf{y})]\hat{\mathcal{F}}_k[F(t)\langle G(\mathbf{y},\mathbf{x}_0(t))\rangle]$$

$$\simeq \frac{k^2}{4\pi r}e^{ikr}\int d\mathbf{y}e^{-ik\mathbf{n}\cdot\mathbf{y}}\epsilon(\mathbf{y})\hat{\mathcal{F}}_k[F(t)\langle G(\mathbf{y},\mathbf{x}_0(t))\rangle]. \tag{36}$$

Here, $\mathbf{n}=\mathbf{x}/|\mathbf{x}|=\mathbf{n}_\parallel\cos\theta+\mathbf{n}_\perp\sin\theta$, where $\mathbf{n}_\parallel$ is a unit vector in direction $\mathbf{V}$.

Consider a source with constant productivity ($F(t)=F_0$) moving with constant velocity ($\mathbf{x}_0(t)=\mathbf{V}t$). In this case,

$$\hat{\mathcal{F}}_k[F(t)\langle G(\mathbf{y},\mathbf{x}_0(t))\rangle] = -\frac{F_0}{4\pi}\int dt\frac{e^{+i\omega t+i\Gamma|\mathbf{y}-\mathbf{x}_0(t)}}{|\mathbf{y}-\mathbf{x}_0(t)|}$$

$$= -\frac{F_0}{4\pi V}e^{ikM^{-1}y_\parallel}$$

$$\times\int_{-\infty}^{+\infty}ds\frac{e^{-ikM^{-1}s+i\Gamma\sqrt{y_\perp^2+s^2}}}{\sqrt{y_\perp^2+s^2}}. \tag{37}$$

By replacing $s\to|y_\perp|\sinh s$, we first reduce the integral to the form

$$\hat{\mathcal{F}}_k[F(t)\langle G(\mathbf{y},\mathbf{x}_0(t))\rangle] = -\frac{F_0}{4\pi V}e^{ikM^{-1}y_\parallel}\times\int_{-\infty}^{+\infty}ds\exp[-ikM^{-1}$$

$$\times|y_\perp|\sinh s+i\Gamma|y_\perp|\cosh s]. \tag{38}$$

Next, by introducing $\Gamma|y_\perp|=\cosh\alpha, kM^{-1}|y_\perp|=\sinh\alpha$, we reduce the integral to the expression that is proportional to the Sommerfeld integral[22] $Z_0(z)$,

$$\hat{\mathcal{F}}_k[F(t)\langle G(\mathbf{y},\mathbf{x}_0(t))\rangle]$$

$$= -i\frac{F_0}{4V}e^{ikM^{-1}y_\parallel}$$

$$\times\frac{1}{i\pi}\int_{-\infty}^{+\infty}ds\exp[i|\mathbf{y}_\perp|\sqrt{\Gamma^2-k^2M^{-2}}\cosh s]$$

$$= -i\frac{F_0}{4V}e^{ikM^{-1}y_\parallel}Z_0(\beta|\mathbf{y}_\perp|). \tag{39}$$

Integral $Z_0(z)$ with a complex argument z describes the field around the moving source. It coincides with the Hankel function $H^{(1)}(z)$ if the contour of integration is chosen in a special manner (to ensure correct asymptotic conditions for radia-

tion). The Hankel function can obviously be expressed in terms of the MacDonald function.[23] Parameter $\beta = \sqrt{\Gamma^2 - k^2 M^{-2}}$ is a complex parameter, where $\Gamma = K_k + i\gamma_k$ (see Eq. (24)). After all calculations, we find that at far distances ($r \gg L$, where L is the distance traveled by the source)

$$\phi'_k(\mathbf{x}) \simeq \frac{e^{ikr}}{r}\left[-i\frac{cF_0}{4M}\frac{k^2}{4\pi}\right] \times \int d\mathbf{y}_\perp dy_\parallel \epsilon(\mathbf{y}_\perp, y_\parallel) e^{+ik(M^{-1}-\cos\theta)y_\parallel} \times e^{-ik\sin\theta \mathbf{n}_\perp \cdot \mathbf{y}_\perp} Z_0(\beta|\mathbf{y}_\perp|). \quad (40)$$

This expression has the form of a divergent spherical wave $[\phi'_k(\mathbf{x}) \sim r^{-1}e^{ikr}]$, i.e., it truly describes propagating radiation. However, the question of what exactly is the origin of this radiation—the moving source or the fluctuation of the medium—is not an accurate one. Both are the necessary components to form the radiation. If any of them is removed ($F_0, M \to 0$ or $\epsilon \to 0$), the radiation disappears.

By substituting Eq. (40) into Eq. (9) and taking into consideration that $\partial_r \phi'_k \simeq ik\phi'_k$ in the domain $K_k^{-1} \ll r \ll \gamma^{-1}$, the angular-spectral power of the radiation can be calculated from

$$\langle \bar{W}_{\omega,\mathbf{n}} \rangle^{\mathrm{tr}} = \lim_{T\to\infty} \frac{r^2}{T}\left[\frac{1}{\pi} ck^2 \langle \phi'_k \phi'^*_k \rangle\right]. \quad (41)$$

To analytically evaluate the expression, we can consider the simplest case of Gaussian distribution: $\langle \epsilon(\mathbf{y}')\epsilon(\mathbf{y}'')\rangle = \langle \epsilon^2 \rangle \exp[l^{-2}(y'_\parallel - y''_\parallel)^2 + l^{-2}(\mathbf{y}'_\perp - \mathbf{y}''_\perp)^2]$. By changing variables $\mathbf{y}', \mathbf{y}'' \to \mathbf{y} = \mathbf{y}' - \mathbf{y}''$, $\mathbf{Y} = \frac{1}{2}(\mathbf{y}' + \mathbf{y}'')$ to calculate the double integral with respect to $y'_\parallel$ and $y''_\parallel$. We obtain

$$\begin{aligned}\lim_{T\to\infty}\frac{r^2}{T}\langle|\phi'_k|^2\rangle &= \lim_{T\to\infty}\frac{1}{T}\left[-\frac{F_0}{4V}\frac{k^2}{4\pi}\right]^2 \langle\epsilon^2\rangle \int dy'_\parallel dy''_\parallel \exp\left[\frac{(y'_\parallel - y''_\parallel)^2}{l^2} + ik\left(\frac{1}{M} - \cos\theta\right)(y'_\parallel - y''_\parallel)\right] \\ &\times \int d\mathbf{y}'_\perp d\mathbf{y}''_\perp e^{-ik\sin\theta \mathbf{n}_\perp\cdot(\mathbf{y}'_\perp - \mathbf{y}''_\perp) - l^{-2}(\mathbf{y}'_\perp - \mathbf{y}''_\perp)^2} Z_0(\beta|\mathbf{y}'_\perp|) Z_0^*(\beta|\mathbf{y}''_\perp|) \\ &= \lim_{T\to\infty}\frac{L}{VT}\left[-\frac{F_0}{4}\frac{k^2}{4\pi}\right]^2 \langle\epsilon^2\rangle l\sqrt{\pi}\exp\left[-\frac{1}{4}k^2l^2\left(\frac{1}{M} - \cos\theta\right)^2\right] \int d\mathbf{y}_\perp e^{-ik\sin\theta\mathbf{n}_\perp\cdot\mathbf{y}_\perp - l^{-2}(\mathbf{y}_\perp)^2} \\ &\times \int d\mathbf{Y}_\perp Z_0\left(\beta\left|\mathbf{Y}_\perp + \frac{1}{2}\mathbf{y}_\perp\right|\right) Z_0^*\left(\beta\left|\mathbf{Y}_\perp - \frac{1}{2}\mathbf{y}_\perp\right|\right). \end{aligned} \quad (42)$$

In terms of the modified variables, differentials $d\mathbf{y}'d\mathbf{y}'' = d\mathbf{y}d\mathbf{Y}$. By definition of quantities L and T, the ratio $\lim_{T\to\infty} L/VT = 1$. In Eq. (42) we calculate the integral with respect to $\mathbf{Y}_\perp$ using the theorem of cylindrical function composition.[22] Let us define the angles for integration from the fixed vector $\mathbf{y}_\perp$. The integral is then written as

$$\begin{aligned}\mathrm{G}(y_\perp) &\equiv \int d\mathbf{Y}_\perp Z_0\left(\beta\left|\mathbf{Y}_\perp + \frac{1}{2}\mathbf{y}_\perp\right|\right) Z_0^*\left(\beta\left|\mathbf{Y}_\perp - \frac{1}{2}\mathbf{y}_\perp\right|\right) \\ &= \int_0^\infty dY_\perp Y_\perp \int_0^{2\pi} d\phi \left[H(Y_\perp - y_\perp)\left(\sum_{m=-\infty}^{+\infty} Z_m(\beta Y_\perp) J_m\left(\frac{1}{2}\beta y_\perp\right) e^{im\phi} \sum_{n=-\infty}^{+\infty} Z_n^*(\beta Y_\perp) J_n^*\left(\frac{1}{2}\beta y_\perp\right) e^{-in(\phi+\pi)}\right)\right. \\ &\left. + H(y_\perp - Y_\perp)\left(\sum_{m=-\infty}^{+\infty} J_m(\beta Y_\perp) Z_m\left(\frac{1}{2}\beta y_\perp\right) e^{im\phi} \sum_{n=-\infty}^{+\infty} J_n^*(\beta Y_\perp) Z_n^*\left(\frac{1}{2}\beta y_\perp\right) e^{-in(\phi+\pi)}\right)\right]. \end{aligned}$$

Here, the Heaviside step function $H(z)$ is used, $Z_m(x)$ and $Z_n(x)$ are cylindrical functions,[22] and $J_m(x)$ and $J_n(x)$ are the Bessel functions. After integrating with respect to angle ϕ, only diagonal terms remain in the double sum, and we obtain

$$\begin{aligned}\mathrm{G}(y_\perp) &= 2\pi \sum_{m=-\infty}^{+\infty} (-1)^m \int_0^\infty dY_\perp Y_\perp \\ &\times \left[H(Y_\perp - y_\perp)|Z_m(\beta Y_\perp)|^2 \left|J_m\left(\frac{1}{2}\beta y_\perp\right)\right|^2 \right. \\ &\left. + H(y_\perp - Y_\perp)|J_m(\beta Y_\perp)|^2 \left|Z_m\left(\frac{1}{2}\beta y_\perp\right)\right|^2\right]. \end{aligned} \quad (43)$$

It is convenient to rewrite this expression in the form

$$\begin{aligned}\mathrm{G}(y_\perp) &= \frac{2\pi}{|\beta|^2} \sum_{m=-\infty}^{+\infty} (-1)^m \\ &\times \left[\left|J_m\left(\frac{1}{2}\beta y_\perp\right)\right|^2 \int_{|\beta|y_\perp}^\infty ds s \left|\beta Z_m\left(\frac{\beta}{|\beta|}s\right)\right|^2 \right. \\ &\left. + \left|Z_m\left(\frac{1}{2}\beta y_\perp\right)\right|^2 \int_0^{|\beta|y_\perp} ds s \left|\beta J_m\left(\frac{\beta}{|\beta|}s\right)\right|^2\right]. \end{aligned} \quad (44)$$

The principal contribution in the integral with respect to $\mathbf{y}_\perp$ [see Eq. (42)] comes from domain $|\mathbf{y}_\perp| \leqslant l$. We can simplify Eq. (44) for a special case of source motion when Mach number $M \sim 1$. For $M \sim 1$, the effect of transition radiation is very distinctly expressed. Since in this case $|\beta|^{-1} \gg l$, the corresponding integral becomes equal to $G(y_\perp) \simeq 2\pi C|\beta|^{-2}$. Here, C is a coefficient independent of the source velocity. In fact, the integral with respect to s for $|\beta| y_\perp \to 0$ does not contain any parameters and, for this reason, produces a numerical value of order unity. We can write that

$$\int d\mathbf{y}'_\perp d\mathbf{y}''_\perp e^{-ik \sin\theta \mathbf{n}_\perp \cdot (\mathbf{y}'_\perp - \mathbf{y}''_\perp) - l^{-2}(\mathbf{y}'_\perp - \mathbf{y}''_\perp)^2} \times Z_0(\beta|\mathbf{y}'_\perp|) Z_0^*(\beta|\mathbf{y}''_\perp|) \simeq 2\pi^2 C l^2 e^{-(1/4)k^2 l^2 \sin^2\theta} |\beta|^{-2}, \tag{45}$$

where we assume that the source velocity is comparable to the sound speed; i.e., β is small. Under this condition, $\beta \simeq ikM^{-1}\sqrt{1-M^2-2\kappa_k - i2\delta_k} \simeq ik\sqrt{2}\sqrt{(1-M-\kappa_k) - i\delta_k}$.

By combining the obtained results, Eqs. (41)–(45), and taking into account that $\langle \epsilon^2 \rangle < 1$, we find the following expression for the angular-spectral density of transition radiation:

$$\langle \bar{W}_{\omega,\mathbf{n}} \rangle^{\mathrm{tr}} \simeq \frac{2\pi\sqrt{\pi} C}{(8\pi)^2} \frac{c|F_0|^2 \langle \epsilon^2 \rangle l^3 k^4 M^2}{\sqrt{(1-M^2-2\kappa_k)^2 + 4\gamma_k^2}} \times \exp\left[-\frac{1}{4}(kl)^2 \left[\left(\frac{1}{M} - \cos\theta \right)^2 + \sin^2\theta \right] \right]. \tag{46}$$

This expression produces an estimate of the scattering energy output for a subsonic, $M \leqslant 1$, motion in a nonhomogeneous medium. If $M \geqslant 1$, an additional (Cherenkov) channel [Eq. (34)] opens.

VI. CONCLUSION

When a physical object without its own eigenfrequency moves through an acoustically *homogeneous* medium, the only possible form of acoustic radiation is the emission of Mach shock waves, which appear when source velocity surpasses the speed of sound, i.e., when $M_* \equiv M \cos\theta \geqslant 1$ ($M = V/c$ is the Mach number.) In *inhomogeneous* media, in nonstationary media, or in the neighborhood of such media, the motion of the source is accompanied by the so-called transition radiation (scattering and diffraction), which takes place even when the source moves with subsonic velocity.[6]

In the considered case of a strongly fluctuating medium, modeled by Eq. (1), the conditions for Cherenkov radiation can change drastically. In fact, the condition $(kl)^2 = (a_2/a_1) \times (M_*^{-1} - 1)/(1 + a_2\langle \epsilon^2 \rangle - M_*^{-1}) > 0$ [following from Eq. (32)] shows that in such a fluctuating medium the radiation channel opens for the subsonic Mach numbers, $M < 1$. This type of radiation is possible in the framework of our simple model when $(1 + a_2\langle \epsilon^2 \rangle)^{-1} < M \cos\theta_k < 1$.

The shock wave with a sharp front does not form in this case because different spectral components are radiated under different angles. In fact, the relationship between the radiated short and long wave angles is described by a simple formula [Eq. (32)] that makes experimental verification possible.

The characteristics of the transition scattering energy follow from Eq. (46). The power of the radiation tends to zero for both small and large values of wave numbers. The direction of transition radiation depends strongly on the source velocity. As expected, the transition radiation effect disappears when the source velocity approaches zero.

When the source moves at a subsonic speed, the characteristic space scale of the domain (where the source moves and from where the radiation is emitted) must be sufficiently large. This is very important when conducting experimental observations. In fact, the attached field reorganization does not occur instantaneously and takes some time to develop because the relief of stress always occurs with a finite (sonic) velocity. However, during this time the source travels additional distance (L). Therefore, it is necessary that this characteristic scale L is greater than the characteristic distance L_{ph} at which the radiation is formed, $L \gg L_{\mathrm{ph}} \sim \omega^{-1}(V^{-1} - c^{-1})^{-1}$. In fact, the resulting field contains two components: the first ("attached field") describes the intrinsic field of the source, $\phi_k^i \sim \exp(-ikM^{-1}x)$, while the second ("radiated field") is the result of the interaction with inhomogeneities, $\phi_k^r \sim \exp(-i\mathbf{k}\cdot\mathbf{x})$. Thus, the expression for the spectral energy (being a quadratical functional of fields) contains an interference term proportional to the product of these components. This (spatially oscillating) interference term vanishes after the inevitable space, temporary or statistical averaging. This is important to remember when defining the full energy or performing experimental measurements. Therefore, at large distances ($L \gg L_{\mathrm{ph}}$), when the intrinsic field of the source and the free field separate, the radiation can be observed and registered.

The presented analysis of wave propagation and radiation by moving sources in the fluctuating media was made with the assumption that the fluctuation level is not too high: $(kl)^4\langle \epsilon^2 \rangle < 1$. For this reason, we were able to limit our consideration only to the first term in Eq. (B13). However, for liquids, in the vicinity of the critical point[20] where acoustical parameters of the medium exhibit large fluctuations, the subsequent expansion terms may be needed in Eq. (B13). Obviously, in general, it is not possible to calculate all diagrams, and, therefore, some diagrams have to be omitted. Under certain conditions, however, it is feasible to take all higher-order contributions into account without much extra effort. For example, if we consider kernel $\Sigma(r) \simeq k^4 G_0(r) B(r) \to \Sigma(r) \simeq k^4 \langle G(r) \rangle B(r)$, we can obtain the so-called *self-consistent* approximation. In this context, only the last diagram in the series for Σ of Fig. 3(b) is needed. Physically the self-consistent approximation is very natural: it describes the signal propagation from one fluctuation (scatterer) to another, which happens not in empty space but in space filled with other scatterers.

We considered the basic features of the effect of transition scattering in the framework of the simplest scalar (acoustical) model, Eq. (1). However, the transition radiation and transition scattering are universal physical phenomena.

They might exist not only for scalar (acoustical) fields[6,12,24] but also for other types of waves of arbitrary physical nature.[7]

APPENDIX A: ACOUSTIC PARAMETER FLUCTUATIONS IN MEDIA

As noted above, transition scattering radiation arises when a source moves through a medium whose properties are such that the speed of sound fluctuates due to a variety of natural phenomena. Such hydrodynamical systems are described by equations of the type of Eq. (13)–(15). Below we consider two simplest examples.

The first example is the system with inhomogeneous distribution of density. The mass and momentum conservation equations are $D_t\rho+\rho\ \mathrm{div}\ \mathbf{v}=0$ and $\rho D_t\mathbf{v}+\nabla p=\mathbf{f}$. The full derivative with respect to time is described by the differential operator $D_t=\partial_t+(\mathbf{v}\cdot\nabla)$. Other notations are standard: ρ is density, p is pressure, etc. The system of equations must be completed by the equation of state, $p=p(\rho,\ldots)$ or by the equation of the energy conservation. Consider the second case. The first law of thermodynamics $dU=-pd(\rho^{-1})+\delta Q$ can be rewritten in the form

$$\frac{dU}{dt}=-p\frac{d}{dt}\frac{1}{\rho}+\dot{Q}. \tag{A1}$$

Here, U is the internal energy per unit mass, and $\dot{Q}$ is the heat quantity introduced into the unit mass per unit time. For $U=U(p,\rho)$, Eq. (A1) becomes

$$\left(\frac{\partial U}{\partial p}\right)_\rho\frac{dp}{dt}+\left(\frac{\partial U}{\partial \rho}\right)_p\frac{d\rho}{dt}-\frac{p}{\rho^2}\frac{d\rho}{dt}=\dot{Q}. \tag{A2}$$

It follows from here that

$$\left(\frac{\partial U}{\partial \rho}\right)_p-\frac{p}{\rho^2}=\left(\frac{\partial}{\partial\rho}\left[U+\frac{p}{\rho}\right]\right)_p=\left(\frac{\partial H}{\partial\rho}\right)_p. \tag{A3}$$

Here H is the heat function (enthalpy). We obtain thus

$$\left(\frac{\partial U}{\partial p}\right)_\rho\frac{dp}{dt}+\left(\frac{\partial H}{\partial \rho}\right)_p\frac{d\rho}{dt}=\dot{Q}. \tag{A4}$$

In terms of Jacobians, this equation is written as

$$\frac{\partial(U,\rho)}{\partial(p,\rho)}\frac{dp}{dt}+\frac{\partial(H,p)}{\partial(\rho,p)}\frac{d\rho}{dt}=\dot{Q}. \tag{A5}$$

Using Jacobian's properties,[20,25] we obtain

$$\frac{\partial(U,\rho)}{\partial(p,\rho)}\frac{\partial(\rho,p)}{\partial(H,p)}\frac{dp}{dt}+\frac{d\rho}{dt}=\frac{\partial(\rho,p)}{\partial(H,p)}\dot{Q}. \tag{A6}$$

Calculation of the Jacobians gives

$$\frac{\partial(\rho,p)}{\partial(H,p)}=\frac{\partial(\rho,p)}{\partial(T,p)}\frac{\partial(T,p)}{\partial(H,p)}=-\rho\left(-\frac{1}{\rho}\frac{\partial\rho}{\partial T}\right)_p\left[\left(\frac{\partial H}{\partial T}\right)_p\right]^{-1}$$

$$=-\rho\frac{\alpha_p}{c_p}. \tag{A7}$$

Similarly,

$$\frac{\partial(U,\rho)}{\partial(p,\rho)}\frac{\partial(\rho,p)}{\partial(H,p)}=\frac{\partial(U,\rho)}{\partial(T,\rho)}\frac{\partial(T,\rho)}{\partial(T,p)}\left[\frac{\partial(T,p)}{\partial(H,p)}\right]^{-1}$$

$$=-\frac{c_v}{c_p}\left(\frac{\partial\rho}{\partial p}\right)_T=-\frac{1}{c^2}. \tag{A8}$$

Here, $c^2=(c_p/c_v)(\partial p/\partial\rho)_T$ is the square of adiabatic sound speed, T is temperature, c_p and c_v are specific heat at constant pressure and volume, respectively, and $\alpha_p=-\rho^{-1}(\partial\rho/\partial T)_p$.

Equation (A6) is transformed to

$$-\frac{1}{c^2}\frac{dp}{dt}+\frac{d\rho}{dt}=-\frac{\alpha_p}{c_p}\rho\dot{Q} \tag{A9}$$

or, in combination with a continuity equation, is written as

$$\frac{1}{\rho c^2}\frac{dp}{dt}+\mathrm{div}\ \mathbf{v}=\frac{\alpha_p}{c_p}\dot{Q}. \tag{A10}$$

The equilibrium state is characterized by the constant value of pressure p_0 and the absence of fluid motion, $\mathbf{v}=0$. For a perturbed state, $p=p_0+p_1$ and $\mathbf{v}=\mathbf{v}_1$. In linear approximation, we obtain the set of equations

$$\partial_t\mathbf{v}_1+\frac{1}{\rho_0}\nabla p_1=\frac{\mathbf{f}}{\rho_0},$$

$$\frac{1}{\rho_0c_0^2}\partial_t p_1+\mathrm{div}\ \mathbf{v}_1=\left[\frac{\alpha_p}{c_p}\right]_0\dot{Q}. \tag{A11}$$

To derive the wave equation when force and heat sources are present, we take derivatives of the expressions with respect to time and coordinates and linearly combine them. The terms with velocity derivatives cancel each other to produce the following second-order equation:

$$\rho_0\ \mathrm{div}\left[\frac{1}{\rho_0}\nabla p_1\right]-\frac{1}{c_0^2}\partial_{tt}p_1=\rho_0\ \mathrm{div}\ \frac{\mathbf{f}}{\rho_0}-\partial_t\left[\frac{\alpha_p\rho_0}{c_p}\right]_0\dot{Q}. \tag{A12}$$

Wave equation (A12) (without the right part and with assumed coordinate-dependent density ρ_0) was formulated by Bergmann.[26] In this context, we have to note that if the model of barotropic fluid is chosen, $\rho=\rho(p)$, i.e., density is a function of pressure only, then the density of the equilibrium state (besides the gravitational field) can only be constant (because $p_0=$const in the equilibrium state) and cannot depend on coordinates. If density is a function of several thermodynamical arguments [e.g., $\rho=\rho(p,T)$ or $\rho(p,s)$], then even if $p_0=$const, equilibrium density can be coordinate dependent, $\rho(\mathbf{x})$, if the second argument is coordinate dependent. In such case, stationary fluctuations of density can take place when some mechanisms of energy input is present in the medium. By introducing a new field variable, namely, $p_1=\sqrt{\rho_0}P$, we can eliminate the term containing the first spatial derivative of p_1 and obtain the equation structurally close to Eq. (1).

The second example of the medium, for which the pressure evolution equation has the form of Eq. (1), is a liquid with distributed gas bubbles (cavitated liquid,[27] bubble chamber,[28] upper oceanic layer,[29] jet wake, swirls, etc.).

We assume that the characteristic length λ of the acoustical wave is very large relative to the average distance d between the bubbles and to their radii R, which are small: $\lambda \gg d \gg R$. In this case, the homogeneous approximation is valid: the liquid with gas babbles can be considered as a homogeneous (on average) medium with some effective density, pressure, and other quantities (Ackeret[30]). The density of the mixture is $\rho=\rho_l(1-X)+\rho_g X$, where ρ_l and ρ_g are, respectively, the densities of liquid and gas $(\rho_g \ll \rho_l)$. Quantity X is the volume fraction of gas in liquid. We assume that $X \ll 1$, which follows from $R \ll d$. Density perturbation is then $\rho'=\rho-\rho_0 \simeq (1-X_0)\rho_l' - \rho_{0g}X' \simeq c_l^{-2}p' - \rho_{0g}X' - (\rho_0 \alpha_p T_0/c_p)s'$. Here, the equilibrium density is $\rho_0=\rho_{0l}(1-X_0)+\rho_{0g}X_0 \simeq \rho_{0l}=\text{const}$ and c_l is the sound speed in pure liquid. The set of linearized (with respect to field perturbations) equations, which describe the evolution of the gas-liquid mixture, is thus

$$\rho_0 \partial_t \mathbf{v}' + \nabla p' = \mathbf{f},$$

$$\partial_t \rho' + \rho_0 \operatorname{div} \mathbf{v}' = 0,$$

$$\rho_0 T_0 \partial_t s' = Q,$$

$$\rho' = \frac{1}{c_l^2} p' - \rho_{0g} X' - \frac{\rho_0 \alpha_p T_0}{c_p} s',$$

$$X'(\mathbf{x}) = \int_0^\infty dR_0 n(R_0,\mathbf{x}) V'(R_0),$$

$$\ddot{V}' + \omega_0^2 V' + \hat{D}V' = -\frac{4\pi R_0}{\rho_0} p'. \tag{A13}$$

Here, the first two equations are the mass and momentum conservation equations (in linear approximation) for the mixture, the second is the equation of thermal conduction in liquid when conduction and viscosity have little effect on the efficiency of the sound-generating mechanism (conditions for such possibility were discussed in Ref. 6), and the fourth expression for density perturbation is written in linear approximation, too. Here, p', ρ', and s' are variations of pressure, density, and entropy (per mass unit) relative to their equilibrium values, α_p is the termal expansion coefficient, and c_p is the specific heat at constant pressure. Quantities $\mathbf{f}$ and Q characterize the effects of the applied force and thermal sources in the medium. The last equation represents the linearized version of the evolution equation of one gas bubble in the external field. Dots signify the second derivative with respect to time. Bubbles are assumed to form spherical cavities and can only pulsate. The interaction between bubbles is neglected. It means that the distance between bubbles is large, $R \ll d$. The equation of motion of one bubble is governed by Rayleigh's equation.[25,31] The small spherically symmetrical volume perturbation of the gas bubble of initial radius R_0 is $V' \simeq 4\pi R_0^2 R'$, $R'=R(t)-R_0$. The bubbles are distributed with respect to their sizes according to some local distribution function $n(R_0,\mathbf{x})$, which tends to zero when $R_0 \to 0$ and $R_0 \to \infty$. The proper frequency of spherical oscillations is $\omega_0=\sqrt{3\gamma p_0/\rho_0 R_0^2}$. Here, p_0 is the gas pressure inside the bubble. The dissipative part of the equation describing the bubble pulsation can be found from the analysis of specific mechanism of energy losses or given by phenomenological estimates.

For air bubbles in water (at the atmospheric pressure p_0 and $\gamma \simeq 1.4$), one can obtain the rough estimate $\omega_0 R_0 \simeq 20$ m/s; i.e., for the bubble of $R_0 \sim 0.1$ mm, the resonance frequency is ~ 33 kHz. For air bubbles of order of microns, surface tension μ needs to be taken into account: $\omega_0 \to \omega_0 = \sqrt{3\gamma p_0/\rho_0 R_0^2 - 2\mu/\rho_0 R_0^3}$.

In the low-frequency limit, when the characteristic frequency of waves is small, $\omega \ll \omega_0$, in the equation for V' oscillations one can neglect all terms with derivatives with respect to time; i.e., quasistatic expression is valid: $V' \simeq -4\pi R_0^3 p'/3\gamma p_0$. Then by collecting all necessary expression, we find the equation for pressure in the form

$$\Delta p' - \frac{1}{c_l^2}\left[1 + \frac{\rho_0}{\rho_{0g}}\frac{c_l^2}{c_g^2} X_0(\mathbf{x})\right] \partial_t^2 p' = \operatorname{div} \mathbf{f} - \frac{\alpha_p}{c_p}\partial_t Q. \tag{A14}$$

Here, c_l is the speed of sound in pure liquid. The presence of bubbles can radically change the speed of sound in the mixture because the correcting factor $\epsilon(\mathbf{x})=(\rho_0/\rho_{0g}) \times (c_l^2/c_g^2) X_0(\mathbf{x})$ is not small even for small concentrations of bubbles. So, for $1 \gg X_0 \gg X_{cr}=(\rho_{0g}/\rho_0)(c_g^2/c_l^2)$, the second term in brackets is prevalent. For air bubbles in water, $X_{cr} \simeq 6 \times 10^{-5} \ll 1$. If pulsation ω is comparable with ω_0, dispersive and even nonlinear effects must be taken into account.[32]

APPENDIX B: DYSON'S EQUATION IN TERMS OF FUNCTIONAL DERIVATIVES AND GREEN'S FUNCTION

We expose here a simple method of finding the Green's function based on the use of functional derivatives. First of all, we introduce the trial Green's function that satisfies

$$\Delta G_0(\mathbf{x},\mathbf{x}') + k^2 G_0(\mathbf{x},\mathbf{x}') = \delta((\mathbf{x}-\mathbf{x}')). \tag{B1}$$

For infinite space, $G_0(\mathbf{x},\mathbf{x}')=-(4\pi r)^{-1}\exp(ikr)$, where $r=|\mathbf{x}-\mathbf{x}'|$. Then we rewrite Eq. (14) in the integral form

$$G(\mathbf{x},\mathbf{z}) = G_0(\mathbf{x}-\mathbf{z}) + \int d\mathbf{y} G_0(\mathbf{x}-\mathbf{y})[-k^2\epsilon(\mathbf{y})]G(\mathbf{y},\mathbf{z}) \tag{B2}$$

and average it with respect to fluctuations,

$$\langle G(\mathbf{x},\mathbf{z})\rangle = G_0(\mathbf{x}-\mathbf{z}) + \int d\mathbf{y} G_0(\mathbf{x}-\mathbf{y})\langle[-k^2\epsilon(\mathbf{y})]G(\mathbf{y},\mathbf{z})\rangle. \tag{B3}$$

For a Gaussian homogeneous process when $\langle\epsilon\rangle=0$ and $\langle\epsilon(\mathbf{x}_1)\epsilon(\mathbf{x}_2)\rangle=B(\mathbf{x}_1-\mathbf{x}_2)$, we use the expression (see Ref. 2, Chap. 20, Appendix B)

$$\begin{aligned}\langle\epsilon(\mathbf{y})G(\mathbf{y},\mathbf{z})\rangle &= \int d\mathbf{z}_1 \langle\epsilon(\mathbf{y})\epsilon(\mathbf{z}_1)\rangle \left\langle \frac{\delta G(\mathbf{y},\mathbf{z})}{\delta\epsilon(\mathbf{z}_1)} \right\rangle \\ &= \int d\mathbf{z}_1 B(\mathbf{y}-\mathbf{z}_1) \left\langle \frac{\delta G(\mathbf{y},\mathbf{z})}{\delta\epsilon(\mathbf{z}_1)} \right\rangle.\end{aligned} \tag{B4}$$

Here, the functional derivative is defined by the expression $\delta\epsilon(\mathbf{y})/\delta\epsilon(\mathbf{z})=\delta(\mathbf{y}-\mathbf{z})$, where $\delta(\mathbf{y}-\mathbf{z})$ is a 3D-Dirac

function.[33] Substituting Eq. (B4) into Eq. (B3), we obtain

$$\langle G(\mathbf{x},\mathbf{z})\rangle = G_0(\mathbf{x}-\mathbf{z}) + \int d\mathbf{y} G_0(\mathbf{x}-\mathbf{y}) \times \int d\mathbf{z}_1[-k^2]B(\mathbf{y}-\mathbf{z}_1)\left\langle \frac{\delta G(\mathbf{y},\mathbf{z})}{\delta\epsilon(\mathbf{z}_1)}\right\rangle, \quad \text{(B5)}$$

where the sought function $\langle G\rangle$ is found via its functional derivative $\langle \delta G(\mathbf{y},\mathbf{z})/\delta\epsilon(\mathbf{z}_1)\rangle$. To find this derivative, we use Eq. (B2). The functional derivative of Eq. (B2) with respect to $\epsilon(\mathbf{z}_1)$ leads to

$$\frac{\delta G(\mathbf{y},\mathbf{z})}{\delta\epsilon(\mathbf{z}_1)} = [-k^2]G_0(\mathbf{y}-\mathbf{z}_1)G(\mathbf{z}_1,\mathbf{z}) + \int d\mathbf{z}_2 G_0(\mathbf{y}-\mathbf{z}_2)[-k^2\epsilon(\mathbf{z}_2)]\frac{\delta G(\mathbf{z}_2,\mathbf{z})}{\delta\epsilon(\mathbf{z}_1)}. \quad \text{(B6)}$$

By averaging this equation, we find

$$\left\langle \frac{\delta G(\mathbf{y},\mathbf{z})}{\delta\epsilon(\mathbf{z}_1)}\right\rangle = [-k^2]G_0(\mathbf{y}-\mathbf{z}_1)\langle G(\mathbf{z}_1,\mathbf{z})\rangle + \int d\mathbf{z}_2 G_0(\mathbf{y}-\mathbf{z}_2)\int d\mathbf{z}_3[-k^2]B(\mathbf{z}_2-\mathbf{z}_3) \times\left\langle \frac{\delta^2 G(\mathbf{z}_2,\mathbf{z})}{\delta\epsilon(\mathbf{z}_3)\delta\epsilon(\mathbf{z}_1)}\right\rangle. \quad \text{(B7)}$$

The first term on the right of Eq. (B7) does not vanish when $\epsilon\to 0$; the second term is of the order $\geqslant\epsilon^2$. The next step is to substitute the derived expression into Eq. (B5). We obtain

$$\langle G(\mathbf{x},\mathbf{z})\rangle = G_0(\mathbf{x}-\mathbf{z}) + \int d\mathbf{y} G_0(\mathbf{x}-\mathbf{y})\int d\mathbf{z}_1[-k^2]B(\mathbf{y}-\mathbf{z}_1) \times\left[[-k^2]G_0(\mathbf{y}-\mathbf{z}_1)\langle G(\mathbf{z}_1,\mathbf{z})\rangle + \int d\mathbf{z}_2 G_0(\mathbf{y}-\mathbf{z}_2)\int d\mathbf{z}_3[-k^2]B(\mathbf{z}_2-\mathbf{z}_3) \times\left\langle \frac{\delta^2 G(\mathbf{z}_2,\mathbf{z})}{\delta\epsilon(\mathbf{z}_3)\delta\epsilon(\mathbf{z}_1)}\right\rangle\right],$$

or in other words,

$$\langle G(\mathbf{x},\mathbf{z})\rangle = G_0(\mathbf{x}-\mathbf{z}) + \int d\mathbf{y}d\mathbf{z}_1 G_0(\mathbf{x}-\mathbf{y})[-k^2]^2 B(\mathbf{y}-\mathbf{z}_1) \times G_0(\mathbf{y}-\mathbf{z}_1)\langle G(\mathbf{z}_1,\mathbf{z})\rangle + \int d\mathbf{y}d\mathbf{z}_1 d\mathbf{z}_2 d\mathbf{z}_3 G_0(\mathbf{x}-\mathbf{y}) \times[-k^2]B(\mathbf{y}-\mathbf{z}_1)\times G_0(\mathbf{y}-\mathbf{z}_2)[-k^2]B(\mathbf{z}_2-\mathbf{z}_3) \times\left\langle \frac{\delta^2 G(\mathbf{z}_2,\mathbf{z})}{\delta\epsilon(\mathbf{z}_3)\delta\epsilon(\mathbf{z}_1)}\right\rangle. \quad \text{(B8)}$$

The following strategy is obvious: we calculate the second derivative of Eq. (B2), average it, substitute the obtained result into Eq. (B8), and repeat the procedure over and over for higher derivatives. It is significant that the magnitude of the first term is of order ϵ^0; i.e., it is independent of fluctuations, the second term is of order ϵ^2, and the third is of order $\geqslant\epsilon^4$. If we wish to limit ourselves to the effects of order ϵ^4, no higher, we have to keep in the expression for the second derivative only the terms that do not depend on ϵ.

The second functional derivative with respect to fluctuations of Eq. (B2) [see also Eq. (B6)] gives us

$$\frac{\delta^2 G(\mathbf{z}_2,\mathbf{z})}{\delta\epsilon(\mathbf{z}_3)\delta\epsilon(\mathbf{z}_1)} = [-k^2]G_0(\mathbf{z}_2-\mathbf{z}_1)\frac{\delta G(\mathbf{z}_1,\mathbf{z})}{\delta\epsilon(\mathbf{z}_3)} + [-k^2]G_0(\mathbf{z}_2-\mathbf{z}_3)\frac{\delta G(\mathbf{z}_3,\mathbf{z})}{\delta\epsilon(\mathbf{z}_1)} + \int d\mathbf{z}_4 G_0(\mathbf{z}_2-\mathbf{z}_4)[-k^2\epsilon(\mathbf{z}_4)] \times\frac{\delta^2 G(\mathbf{z}_4,\mathbf{z})}{\delta\epsilon(\mathbf{z}_3)\delta\epsilon(\mathbf{z}_1)}. \quad \text{(B9)}$$

Averaged Eq. (B9) is written as

$$\left\langle \frac{\delta^2 G(\mathbf{z}_2,\mathbf{z})}{\delta\epsilon(\mathbf{z}_1)\delta\epsilon(\mathbf{z}_3)}\right\rangle = [-k^2]G_0(\mathbf{z}_2-\mathbf{z}_1)\left\langle \frac{\delta G(\mathbf{z}_1,\mathbf{z})}{\delta\epsilon(\mathbf{z}_3)}\right\rangle + [-k^2]G_0(\mathbf{z}_2-\mathbf{z}_3)\left\langle \frac{\delta G(\mathbf{z}_3,\mathbf{z})}{\delta\epsilon(\mathbf{z}_1)}\right\rangle + \cdots. \quad \text{(B10)}$$

Next, we use Eq. (B7) to derive the sought expression for the second derivative, which does not contain terms dependent on fluctuations,

$$\left\langle \frac{\delta^2 G(\mathbf{z}_2,\mathbf{z})}{\delta\epsilon(\mathbf{z}_1)\delta\epsilon(\mathbf{z}_3)}\right\rangle \simeq [-k^2]G_0(\mathbf{z}_2-\mathbf{z}_1)[-k^2]G_0(\mathbf{z}_1-\mathbf{z}_3) \times\langle G(\mathbf{z}_3,\mathbf{z})\rangle + [-k^2]G_0(\mathbf{z}_2-\mathbf{z}_3) \times[-k^2]G_0(\mathbf{z}_3-\mathbf{z}_1)\langle G(\mathbf{z}_1,\mathbf{z})\rangle. \quad \text{(B11)}$$

Using the derived expressions for derivatives, we finally find the sought equation,[16]

$$\langle G(\mathbf{x},\mathbf{z})\rangle = G_0(\mathbf{x},\mathbf{z}) + \int d\mathbf{x}_1 d\mathbf{x}_2 G_0(\mathbf{x},\mathbf{x}_1)\Sigma(\mathbf{x}_1,\mathbf{x}_2) \times\langle G(\mathbf{x}_2,\mathbf{z})\rangle. \quad \text{(B12)}$$

The kernel $\Sigma(\mathbf{x}_1,\mathbf{x}_2)\equiv\Sigma_{12}$ of this integral equation is the set of terms

$$\Sigma_{12} = [-k^2]^2 B(\mathbf{x}_1,\mathbf{x}_2)G_0(\mathbf{x}_1,\mathbf{x}_2) + [-k^2]^4\int d\mathbf{x}_3 d\mathbf{x}_4 B(\mathbf{x}_1,\mathbf{x}_3)G_0(\mathbf{x}_1,\mathbf{x}_4) \times B(\mathbf{x}_4,\mathbf{x}_2)G_0(\mathbf{x}_4,\mathbf{x}_3)G_0(\mathbf{x}_3,\mathbf{x}_2) + [-k^2]^4\int d\mathbf{x}_3 d\mathbf{x}_4 B(\mathbf{x}_1,\mathbf{x}_2)G_0(\mathbf{x}_1,\mathbf{x}_4) \times B(\mathbf{x}_4,\mathbf{x}_3)G_0(\mathbf{x}_4,\mathbf{x}_3)G_0(\mathbf{x}_3,\mathbf{x}_2) + \cdots. \quad \text{(B13)}$$

Here, only terms with orders ϵ^2 and ϵ^4 are included. G_0 denotes a "bare" propagator—that is to say, the propagator in a uniform medium. The sought quantity is $\langle G\rangle$, the Green's function for the stochastic medium, also called the "dressed" propagator. To calculate *all* diagrams would amount to solving the problem exactly, which is usually not possible. We

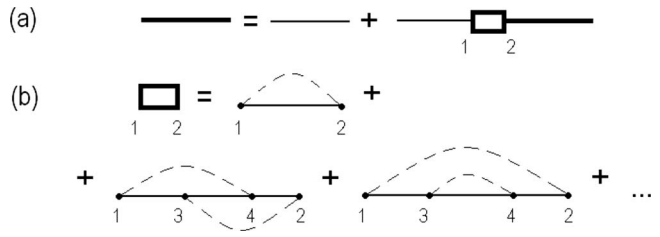


FIG. 3. Graphical (Feynman) representation of (a) Eq. (B12) and (b) Eq. (B13). Averaged Green's function is indicated by a heavy line. The thin line is the bare Green's function $G_0((\mathbf{x},\mathbf{x}_1))$. The dashed line indicates the correlation function $B(\mathbf{x},\mathbf{x}_1)$. Kernel $\Sigma(\mathbf{x},\mathbf{z})$ contains only irreducible diagrams.

therefore assume that $\langle\epsilon^2\rangle$ is small. This allows us to set up a perturbative expansion of Σ.

Equations (B12) and (B13) permit the *à la* Feynman representation (Fig. 3) if the following graphical symbols are introduced: the heavy line represents $\langle G((\mathbf{x},\mathbf{x}_1))\rangle$, the thin line corresponds to $G_0((\mathbf{x},\mathbf{x}_1))$, the thick point illustrates $[-k^2]$, and the dashed line signifies the correlation function $B(\mathbf{x},\mathbf{x}_1)$. The integration is carried out with respect to the inner variables.

Consider a randomly fluctuating in space, but statistically homogeneous medium (see, for example, Ref. 3, p. 342). In this case, statistical characteristics, such as the correlation function, do not change with translation (and rotation) of the framework, i.e., $B(\mathbf{x},\mathbf{x}_1)\equiv B(|\mathbf{x}-\mathbf{x}_1|)$. Equation (B12) becomes

$$\langle G(\mathbf{x}-\mathbf{z})\rangle = G_0(\mathbf{x}-\mathbf{z}) + \int d\mathbf{x}_1 d\mathbf{x}_2 G_0(\mathbf{x}-\mathbf{x}_1)\Sigma(\mathbf{x}_1-\mathbf{x}_2) \times\langle G(\mathbf{x}_2-\mathbf{z})\rangle. \tag{B14}$$

It is resolved via Fourier transformations defined by $Q(\mathbf{s}) = \int d\mathbf{q}(2\pi)^{-3}Q(\mathbf{q})\exp i\mathbf{q}\cdot\mathbf{s}$. After this, we multiply Eq. (B14) by $e^{-i\mathbf{q}\cdot(\mathbf{x}-\mathbf{z})}$ and integrate it with respect to $\mathbf{x}$ (we integrate first with respect to $\mathbf{x}_2$, then with respect to $\mathbf{x}_1$). For Fourier transforms we obtain the following equation:

$$\langle g(\mathbf{q})\rangle = g_0(\mathbf{q}) + \int d\mathbf{x} d\mathbf{x}_1 d\mathbf{x}_2 e^{-i\mathbf{q}\cdot(\mathbf{x}-\mathbf{x}_1+\mathbf{x}_1-\mathbf{x}_2+\mathbf{x}_2-\mathbf{z})} \times G_0(\mathbf{x}-\mathbf{x}_1)\Sigma(\mathbf{x}_1-\mathbf{x}_2)\langle G(\mathbf{x}_2-\mathbf{z})\rangle = g_0(\mathbf{q}) + g_0(\mathbf{q})\Sigma(\mathbf{q})\langle g(\mathbf{q})\rangle, \tag{B15}$$

i.e.,

$$[1-g_0(\mathbf{q})\Sigma(\mathbf{q})]\langle g(\mathbf{q})\rangle = g_0(\mathbf{q}) \rightarrow \langle g(\mathbf{q})\rangle = [g_0^{-1}(\mathbf{q})-\Sigma(\mathbf{q})]^{-1}, \tag{B16}$$

where[34] $g_0(\mathbf{q})=[k^2-q^2]^{-1}$. After a simple algebra, from Eq. (B16), we obtain an equation that determines where the poles are located,

$$-\langle g(\mathbf{q})\rangle^{-1} = -g_0^{-1}(\mathbf{q}) + \Sigma(\mathbf{q}) = 0. \tag{B17}$$

Applying an inverse Fourier transformation to Eq. (B16) we obtain

$$\langle G(|\mathbf{r}-\mathbf{r}_0|)\rangle = \int \frac{d\mathbf{q}}{(2\pi)^3} e^{i\mathbf{q}\cdot|\mathbf{r}-\mathbf{r}_0|} \times\left[k^2-q^2-\int d\mathbf{x}\Sigma(\mathbf{x})\exp(-i\mathbf{q}\cdot\mathbf{x})\right]^{-1} \tag{B18}$$

(see, for example, Ref. 3, p. 358).

APPENDIX C: RADIATION DAMPING AND CHERENKOV EFFECT

Radiation acts on bodies with certain additional force. This force is called radiation damping (or, in electrodynamics,[35] Lorentz frictional force). Let us show that the Cherenkov effect can be expressed via the work of this force.

In our case the work is characterized by the integral [see Eq. (2)]

$$\begin{aligned}\overline{\langle W_f\rangle} &= -\lim_{T\to\infty}\int_{-\infty}^{+\infty} dt\frac{d\omega}{2\pi}(-i\omega)e^{-i\omega t} \times\int d\mathbf{x}\langle\phi_k(\mathbf{x})\rangle F(t)\delta(\mathbf{x}-\mathbf{x}_0(t)) \\ &= -\lim_{T\to\infty}\int_{-\infty}^{+\infty} dt\frac{d\omega}{2\pi}(-i\omega)e^{-i\omega t}\int d\mathbf{x}F(t)\delta(\mathbf{x}-\mathbf{x}_0(t)) \\ &\times\int d\mathbf{x}_1\langle G_k(\mathbf{x},\mathbf{x}_1)\rangle\int_{-\infty}^{+\infty} dt_1 e^{+i\omega t_1}F(t_1)\delta(\mathbf{x}_1-\mathbf{x}_0(t_1)) \\ &= -\int_{-\infty}^{+\infty} d\omega\frac{\omega}{2\pi i}\lim_{T\to\infty}\int_{-\infty}^{+\infty} dt\int_{-\infty}^{+\infty} dt_1 e^{-i\omega(t-t_1)}F(t)F(t_1) \\ &\times\langle G_k(\mathbf{x}_0(t),\mathbf{x}_0(t_1))\rangle.\end{aligned} \tag{C1}$$

Since $G_k=G_{-k}^*$, let us introduce the mixed distribution

$$\langle W(\omega,\ldots)\rangle = -\frac{\omega}{\pi}\lim_{T\to\infty}\int_{-\infty}^{+\infty} dt\int_{-\infty}^{+\infty} dt_1 F(t)F(t_1) \times\Im[e^{-i\omega(t-t_1)}\langle G_k(\mathbf{x}_0(t),\mathbf{x}_0(t_1))\rangle], \tag{C2}$$

which is normalized by the condition

$$\overline{\langle W_t\rangle} = \int_0^\infty d\omega\langle W(\omega,\ldots)\rangle. \tag{C3}$$

Equation (C1) shows that the entire effect is defined by the averaged Green's function $\langle G_k(\mathbf{x},\mathbf{x}_1)\rangle$. Different cases of source motion can be analyzed based on formula (C1).

Consider a source with constant productivity, $F(t)=F_0$, moving with a constant velocity, $\mathbf{x}_0(t)=\mathbf{V}t\equiv Mct$, in a fluctuating, statistically uniform, medium. In this case, the Green's function depends on the difference of arguments and has the following structure: $\langle G_k(\mathbf{x}-\mathbf{x}_1)\rangle=-(4\pi)^{-1}|\mathbf{x}-\mathbf{x}_1|^{-1}\exp+i\Gamma_k|\mathbf{x}-\mathbf{x}_1|$, with $\Gamma=K_k+i\gamma_k$. For low-level fluctuations $\gamma_k\ll K_k\simeq k$. After these assumptions, quantity $\langle W(\omega,\ldots)\rangle$ is

$$\langle W(\omega,\ldots)\rangle = \lim_{T\to\infty}\int_{-\infty}^{+\infty} dtF_0^2\frac{\omega}{4\pi^2}\int_{-\infty}^{+\infty} dt_1 \Im\frac{e^{-i\omega(t-t_1)+i\Gamma_k|\mathbf{V}(t-t_1)|}}{|\mathbf{V}(t-t_1)|}$$

$$= \frac{F_0^2}{4\pi^2 V}\omega\Im\int_{-\infty}^{+\infty}\frac{ds}{|s|}e^{-i(ks-\Gamma_k M|s|)} = \frac{F_0^2}{4\pi^2 M}k$$

$$\times\Im\left[\int_0^{+\infty}\frac{ds}{s}e^{-i(k-\Gamma_k M)s}+\int_0^{+\infty}\frac{ds}{s}e^{+i(k+\Gamma_k M)s}\right]. \tag{C4}$$

It is easy to see that

$$I_{\mp}(M) = \int_0^{+\infty}\frac{ds}{s}e^{(\mp ik+i\Gamma_k M)s}$$

$$= i\Gamma_k\int_{-\infty}^{M}d\mu\int_0^{+\infty}dse^{(\mp ik+i\Gamma_k\mu)s}$$

$$= \int_{-\infty}^{M}d\mu\frac{i\Gamma_k}{\pm ik-i\Gamma_k\mu} = \int_{-\infty}^{M}d\mu\frac{\Gamma_k}{\pm k-\Gamma_k\mu}, \tag{C5}$$

i.e.,

$$I_{\mp}(M) = \int_{-\infty}^{M}\frac{d\mu}{\mu}\frac{\Gamma_k\mu\mp k}{\pm k-\Gamma_k\mu}\pm k\int_{-\infty}^{M}d\mu\frac{1}{\pm k-\Gamma_k\mu}.$$

The first (divergent) term does not contribute to the imaginary part of the expression that interests us. Therefore, the imaginary part of integral of Eq. (C4) can be described as

$$\Im\int_{-\infty}^{+\infty}\frac{ds}{|s|}e^{-i(ks-\Gamma_k M|s|)} = k\int_{-\infty}^{M}d\mu\Im\left[\frac{1}{k-K_k\mu-i\gamma_k\mu}\right.$$

$$\left.+\frac{1}{k+K_k+i\gamma_k\mu}\right]. \tag{C6}$$

For small ε and $\langle\epsilon^2\rangle\ll 1$, expression $\pi^{-1}\varepsilon/(x^2+\varepsilon^2)$ can be replaced by the Dirac function. By integrating with respect to μ, we find for the spectral power of Cherenkov radiation that

$$\langle W(\omega,\ldots)\rangle \simeq \frac{F_0^2}{4\pi}\frac{k^2}{MK_k}\int_{-\infty}^{M}d\mu\delta^{(1)}\left(\mu-\frac{k}{K_k}\right)$$

$$\simeq \frac{F_0^2}{4\pi}\frac{k}{M}H\left(M-\frac{k}{K_k}\right). \tag{C7}$$

This expression is in agreement with Eq. (35) since $K_k\simeq k$.

[1]By the term "a source," we will describe moving objects (beams of charged or neutral particles, bodies in tenuous media, etc.) or localized regions of hydrodynamic stress transported through the medium and produced, for example, by electromagnetic field (through the release of heat or striction), moving vortices, turbulent fluxes, and so on.

[2]A. Ishimaru, *Wave Propagation and Scattering in Random Media* (Academic, New York, 1978).

[3]V. I. Tatarskii, *Scattering of Waves in a Turbulent Atmosphere* (Scientific, Moscow, 1967) (in Russian); V. I. Tatarskii, *Wave Propagation in a Turbulent Medium* (McGraw-Hill, New York, 1961); *The Effects of the Turbulent Atmosphere on Wave Propagation* (Israel Program for Scientific Translations, Jerusalem, 1971).

[4]A. S. Monin and A. M. Yaglom, *Statistical Fluid Mechanics* (MIT, Cambridge, MA, 1975).

[5]V. Pavlov and O. Kharin, "Emission of acoustic waves and formation of heated jet as a fast source moves through a medium with a relativistic equation of state," Zh. Eksp. Teor. Fiz. **98**, 377–386 (1990) [Sov. Phys. JETP **71**, 211–216 (1990)].

[6]V. Pavlov and A. Sukhorukov, "Emission of acoustic transition waves," Sov. Phys. Usp. **28**, 784–804 (1985).

[7]V. L. Ginzburg and V. N. Tsytovich, "Several problems of the theory of transition radiation and transition scattering," Phys. Rep. **49**, 1–89 (1979); *Transition Radiation and Transition Scattering* (Moscow, Nauka, 1984) [Translated into English (Hilger, Bristol, 1990)].

[8]V. L. Ginzburg, "Radiation by uniformly moving sources (Vavilov-Cherenkov effect, transition radiation, and other phenomena)," Phys. Usp. **39**, 973–982 (1996).

[9]For a source moving with subsonic velocity ($M=V/c<1$), the model field equation is $\Delta\phi-c^{-2}\partial_{tt}\phi=\gamma\delta^{(3)}(\mathbf{x}-\mathbf{V}t)$. The Fourier transform of the attached (intrisic) field is given by $\phi_k=(\gamma/2\pi Mc)K_0(k|\mathbf{x}_\perp|\sqrt{M^{-2}-1})\exp ikM^{-1}x_\parallel$, where $k=\omega/c$ and $x_\parallel$ is the coordinate in the source motion direction. The asymptotics of the MacDonald function are $K_0(z)\simeq-\ln z$ for $|z|\ll 1$ and $K_0(z)\simeq\sqrt{\pi/2z}\exp(-z)$ for $|z|\gg 1$; i.e., the attached intrinsic field is localized near the source.

[10]K. Yu. Platonov and G. D. Fleishman, "Transition radiation in media with random inhomogeneities," Phys. Usp. **45**, 235–291 (2002).

[11]V. Pavlov, "Transition radiation of sound in a turbulent medium," Sov. Phys. Acoust. **28**, 55–58 (1982).

[12]V. I. Pavlov and A. I. Sukhorukov, "Transition radiation of sound by a mass source moving over a rough surface," Sov. Phys. Acoust. **29**, 397–399 (1983).

[13]V. D. Lipovskii and V. V. Tamoikin, "Sound emission by moving sources in a nonuniform gaseous medium," Izv. Vyssh. Uchebn. Zaved., Radiofiz. **26**, 183–191 (1983) (in Russian).

[14]All acoustical quantities are considered here in linear approximation and are valid to the second order in ϕ.

[15]This is the standard definition of the force. See, for example, its electrodynamical analogy: V. Ginzburg, "Radiation and radiation friction force in uniformly accelerated motion of a charge," Sov. Phys. Usp. **12**, 565–574 (1970).

[16]A. B. Migdal, *Qualitative Methods in Quantum Theory*, Frontiers in Physics (Addison-Wesley, Reading, MA, 1977), Sec. 3, Chap. 5. We obtained the equation which in quantum electrodynamics is called the *Dyson equation*. Kernel Σ is called the *mass operator*.

[17]In the contemporary literature, such truncation of the series in Eq. (B13) is frequently referred to as Bourret's approximation.

[18]Near the point of phase transition, when thermodynamical fluctuations can be expressed via parameter ϵ, it would be more natural to use the correlation function of Ornstein–Zernike type $B(r)=A\langle\epsilon^2\rangle e^{-r/l}/r$. Here, $A=\kappa T\beta_T/4\pi l^2$, where κ defines the Boltzmann constant, T is the absolute temperature, and $\beta_T=\rho^{-1}(\partial_p\rho)_T$ is defined by the medium state equation $\rho=\rho(p,T)$. However, such choice of the correlation function does not change fundamentally the qualitative understanding of the process considered in the paper.

[19]H. B. Dwighte, *Tables of Integrals* (Macmillan, New York, 1961), Chap. VI: $\int dxx\ln|(1+x)/(1-x)|=2\int dxx\ln(1+x)-\int dxx\ln|1-x^2|=(x^2-1)\ln(x+1)+x-\frac{1}{2}x^2-\frac{1}{2}[(x^2-1)\ln|x^2-1|-x^2]=x+\frac{1}{2}(x^2-1)\ln[(x+1)/(|x-1|)]$.

[20]Yu. Rumer and M. Ryvkin, *Thermodynamics, Statistical Physics and Kinetics* (Nauka, Moscow, 1977). (a) Appendix IV: $\int_0^\infty dxx^{2n}\exp(-\alpha x^2)=(2n-1)!!\sqrt{\pi}2^{-(n+1)}\alpha^{-(2n+1)/2}$. (b) Near the critical point, the fluid is sufficiently hot and compressed that the distinction between the liquid and gaseous phases is almost nonexistent. The density fluctuations in this case are strong. (c) The properties of Jacobiens are exposed in the section "Mathematical applications," p. 536.

[21]L. L. Landau and E. M. Lifshitz, *Electrodynamics of Continuous Media* (Pergamon, New York, 1984).

[22]G. Korn and T. Korn, *Mathematical Handbook for Scientists and Engineers* (McGraw-Hill, New York, 1961), Chap. 6.

[23]The complex integral $Z_0(z)=-i\pi^{-1}\int_C dt\exp(iz\cosh t)$ coincides with the Hankel function $H_0^{(1)}(z)$ for the contour, which begins at $z=-\infty+i0$, ends at $z=+\infty+i(\pi/2)$, and passes through $z=0$ (Ref. 22, 21.8-2). The MacDonald function is defined by $K_0(z)=(\pi/2)iH_0^{(1)}(iz)$. The cylindrical function Z_0 can be expanded as $Z_0(\beta|\mathbf{z}_1-\mathbf{z}_2|)=\Sigma_{k=-\infty}^{+\infty}Z_k(\beta|\mathbf{z}_1|)J_k(\beta|\mathbf{z}_2|)e^{ik(\phi_1-\phi_2)}$, where $Z_k(s)$ are the cylindrical functions of order k, $J_k(s)$ are the Bessel functions of order k, β is an arbitrary complex number, $|\mathbf{z}_1|>|\mathbf{z}_2|$, and ϕ_j are polar angles of $\mathbf{z}_j$ (see Ref. 22, 21.8-13).

[24]E. V. Pavlova and O. A. Kharin, "Acoustic transition radiation from sources crossing a chaotic phase barrier," Sov. Phys. Acoust. **38**, 496–498 (1992).

[25]L. D. Landau and E. M. Lifshitz, *Fluid Dynamics* (Pergamon, London,

1959).
[26]P. Bergmann, "The wave equation in a medium with a variable index of refraction," J. Acoust. Soc. Am. **17**, 329–333 (1946).
[27]G. Flinn, "Physics of acoustic cavitation in liquids," in *Physical Acoustics*, edited by W. Mason (Academic, New York, 1964), Vol. **1B**, p. 7–138.
[28]V. A. Akulichev, *Acoustic Cavitation in Cryogenic and Boiling Liquids* (Springer, The Netherlands, 1982).
[29]L. M. Brekhovskikh and Y. Lysanov, *Fundamentals of Ocean Acoustics* (Springer-Verlag, New York, 2002).
[30]J. Ackeret, "Experimentelle und theoretische Untersuchungen über Hohlraumbildung (Kavitation) im Wasser (Experimental and theoretical investigation on cavities formation (cavitation) in water)," Tech. Mech. und Thermodyn. **1**(2), 63–72 (1930).
[31]O. M. Rayleigh, "On the pressure developed in a liquid during the collapse of the spherical cavity," Philos. Mag. **34**, 94–98 (1917).
[32]K. A. Naugol'nykh and L. A. Ostrovsky, *Nonlinear Wave Processes in Acoustics* (Nauka, Moscow, 1990).
[33]The functional derivative of functional $F[\epsilon(\mathbf{z})]=\int d\mathbf{z} f(\epsilon(\mathbf{z}))$ is calculated using $\delta F[\epsilon(z)]/\delta\epsilon(\mathbf{x})=\int d\mathbf{z}(\partial_\epsilon f(\epsilon))\delta(\epsilon(\mathbf{z}))/\delta\epsilon(\mathbf{x})$.
[34]Integrals with respect to q are calculated via contour integration by taking into account the positions of singularities, which should be contoured correctly. Therefore, one can write $[k^2-q^2]^{-1}\rightarrow[k^2-q^2+i0]^{-1}$. Introduction of symbol $+i0$ reminds us about the rule of pole-contouring. This correction is equivalent to taking into account an effective infinitesimal absorption.
[35]L. D. Landau and E. M. Lifshitz, *The Classical Theory of Fields* (Addison-Wesley, Cambridge, MA, 1951).

Inversion of spinning sound fields

Michael Carley[a)]
Department of Mechanical Engineering, University of Bath, Bath BA2 7AY, England

(Received 30 September 2008; revised 24 November 2008; accepted 24 November 2008)

A method is presented for the reconstruction of rotating monopole source distributions using acoustic pressures measured on a sideline parallel to the source axis. The method requires no *a priori* assumptions about the source other than that its strength at the frequency of interest varies sinusoidally in azimuth on the source disk so that the radiated acoustic field is composed of a single circumferential mode. When multiple azimuthal modes are present, the acoustic field can be decomposed into azimuthal modes and the method applied to each mode in sequence. The method proceeds in two stages, first finding an intermediate line source derived from the source distribution and then inverting this line source to find the radial variation in source strength. A far-field form of the radiation integrals is derived, showing that the far-field pressure is a band-limited Fourier transform of the line source, establishing a limit on the quality of source reconstruction, which can be achieved using far-field measurements. The method is applied to simulated data representing wind-tunnel testing of a ducted rotor system (tip Mach number of 0.74) and to control of noise from an automotive cooling fan (tip Mach number of 0.14), studies which have appeared in the literature of source identification. © *2009 Acoustical Society of America.* [DOI: 10.1121/1.3050311]

PACS number(s): 43.28.We, 43.50.Nm, 43.20.Rz [AH] Pages: 690–697

I. INTRODUCTION

This paper describes a method for determining rotating source distributions from acoustic measurements. This is a problem which has been examined by a number of researchers, with many [1–6] considering the problem of estimating the amplitudes of the acoustic modes at the termination of a circular duct, as in the case of aircraft engines. The motivation for these studies has usually been to determine the source terms in their own right in order to find the source mechanisms responsible for the noise or to improve noise control measures, but a second application has been in developing models which can be used to predict the acoustic field. This prediction model can be used in active noise control[7,8] or in using near-field measurements taken in a wind tunnel to make far-field predictions of noise radiated by aircraft in flight.[1,9] This gives rise to two different, though related, problems: the first is the determination, to within some tolerance, of the acoustic source; the second is the determination of the acoustic source to within a tolerance sufficient to give accurate predictions of the acoustic field at points other than the measurement positions.

This paper considers a model problem for the recovery of a rotating source distribution from a set of measurements along a sideline, a line parallel to the source axis. The question of how to position microphones, and how many to use, features in the analysis of many researchers. Typical microphone configurations have included 3 microphones at 120 angular positions,[1] 91 microphones on a fixed polar array,[6] 18 microphones rotating over 20 positions,[2] and 21 microphones located on a fixed arc,[4] depending on the experimental facilities used and the fidelity of results required. Recent work on engine noise has also included the use of sensor arrays mounted inside or on the engine. Examples are the use of 100 pressure sensors on the surface of the intake[10] and simulations of an array of 150 microphones mounted on the wall of an engine duct.[11] In these cases, the methods used are described as "beamforming" and come from the class of techniques used for source location rather than for source characterization.

In this paper, we present an inversion technique which uses data from a linear arrangement of microphones to recover the details of a distribution of monopoles on a disk. This corresponds to the problem of thickness noise of a propeller or other rotors,[12] sound from a baffled circular piston[13] or to sound radiated by the termination of a circular duct, when the Rayleigh approximation is valid.[14,15] The only assumption made is that the source and the acoustic field have a known sinusoidal variation in azimuth—no assumption is required about the form of the radial variation of the source nor is a far-field approximation needed. The resulting method is applied to simulated data using parameters characteristic of problems to which identification methods have been applied in the past.

The source recovery technique which is developed here is based on the measurement techniques used in wind-tunnel measurement of aerodynamic sources[1–9] but could also be viewed in the more general framework of sound source reconstruction in other areas of acoustics[16] and, in particular, in relation to cylindrical near-field acoustical holography (NAH)[17,18] where measurements are taken on a cylindrical surface surrounding a source region and then forward projected to find the acoustic field elsewhere in space or back-projected to find the acoustic quantities which characterize the source. The method of this paper shares some similarities with NAH but differs in incorporating known information about the source geometry and azimuthal dependence.

[a)]Electronic mail: m.j.carley@bath.ac.uk

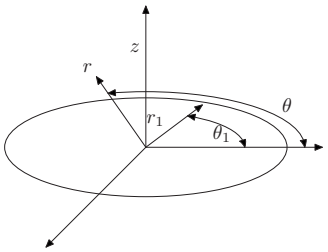


FIG. 1. Coordinate system for radiation prediction.

II. INVERSION OF SPINNING SOUND FIELDS

The acoustic fields to be considered in this paper can all be viewed as being generated by sources distributed over a disk. Figure 1 shows the arrangement of the problem. We use cylindrical coordinates (r,θ,z) with an acoustic monopole source distributed over the disk, $0\leqslant r\leqslant 1$, $z=0$, nondimensionalizing all lengths on a disk radius. At a single frequency ω, the acoustic field p is given by[12]

$$p(r,\theta,z,\omega)=\int_0^1\int_0^{2\pi} s(r_1,\theta_1)\frac{e^{ikR}}{4\pi R}d\theta_1 r_1 dr_1, \tag{1}$$

with wavenumber $k=\omega/c$, c the speed of sound, and source-observer distance

$$R^2=r^2+r_1^2-2rr_1\cos(\theta-\theta_1)+z^2.$$

The source term $s(r_1,\theta_1)$ can be decomposed into a series of azimuthal modes with sinusoidal variation,

$$s(r_1,\theta_1)=\sum_{n=-\infty}^{\infty} s_n(r_1)e^{jn\theta_1},$$

which, upon insertion into Eq. (1) with the transformation $\theta-\theta_1\rightarrow\theta_1$, yields

$$p(r,\theta,z,\omega)=\sum_{n=-\infty}^{\infty} e^{jn\theta}\int_0^1\int_0^{2\pi} s_n(r_1)\frac{e^{j(kR-n\theta_1)}}{4\pi R}d\theta_1 r_1 dr_1, \tag{2}$$

with R being redefined,

$$R^2=r^2+r_1^2-2rr_1\cos\theta_1+z^2.$$

The acoustic field at a frequency ω is thus a sum of azimuthal modes, each of which is directly generated by a corresponding azimuthal mode on the source disk. The aim of the inversion procedure is to recover the source function(s) $s_n(r_1)$ given as input some acoustic pressures measured in the field. The nature of these measurements will depend on the type of source being studied.

There are two main categories of problem which will be considered: rotating sources such as propellers and fans and ducted sources where the duct termination can be considered a disk-shaped source. For a source rotating at angular frequency Ω, the radiated field contains only harmonics of frequency $n\Omega$. Furthermore, if the source strength is steady in the rotating reference frame, there is only one azimuthal mode, of order n, present at each of these frequencies. This means that the acoustic field of Eq. (2) reduces to

$$p(r,\theta,z,n\Omega)=e^{jn\theta}\int_0^1\int_0^{2\pi} s_n(r_1)\frac{e^{j(kR-n\theta_1)}}{4\pi R}d\theta_1 r_1 dr_1.$$

The properties of the acoustic field are largely controlled by the rotor speed and, in particular, the tip Mach number, which, for a source of unit radius, is $M_t=\Omega/c$. When the source rotates supersonically, $M_t>1$, the acoustic field is dominated by the source around the sonic radius $r^*=1/M_t$.[19] When $M_t<1$, the blade tip is the dominant region, and in the far field, its radiation is exponentially stronger than that from inboard regions.[20] This means that the measured field is effectively the field radiated by the tip, and recovering the details of the source at smaller radii will be difficult. On the other hand, if the aim is to accurately compute the acoustic field at a new set of points, it may well be sufficient to capture only the source at the tip.

The structure of the rotating field has been studied using model solutions,[21–23] and the role of the sonic radius has been clarified. The field is made up of a segmented near field, which undergoes a transition around r^*. In "tunneling" across this transition region, the sound field decays exponentially, explaining the relatively weak field radiated by subsonically rotating sources. Supersonic sources have part of the source lying beyond r^* so that they can radiate strongly into the field, without losing energy in tunneling through the transition. This transition region means that source recovery will always be a hard problem if only far-field data are available, a result which will be derived in Sec. II C by considering the bandwidth of the spatial data in the far field.

When the radiating system is a circular duct, the problem can be modeled by taking the noise source to be the duct termination. In that case, the source distribution is composed of the duct modes which have propagated to the end of the duct. The field inside a rigid circular duct is composed of modes of the form $J_n(k_{mn}r)\exp[j(n\theta-k_{zmn}z)]$, where $J_n(\cdot)$ is the Bessel function of the first kind, $J_n'(k_{mn})=0$, and k_{zmn} is an axial wavenumber.[14] When k_{zmn} has an imaginary part, the mode decays exponentially in the duct and does not propagate to the termination. In any case, the source strength at the duct termination can be taken to be the acoustic velocity generated by the modes which do propagate, and the radiated noise can be accurately computed over much of the field using a Rayleigh integral[3,14] or a Kirchhoff integral over a wider range of polar angles.[3,24] The source can again be modeled as a circular disk with an azimuthally varying source term. A number of methods have been developed for the identification of the radiating modes[2–6] and have been found to be accurate and robust, considering the assumptions made in their development.

The one extra difficulty in the duct case compared to the rotor problem is that the source at the duct termination may be composed of modes of more than one azimuthal order. In this case, there are procedures which use measurements at multiple angles to extract the modal amplitudes in the acoustic field. For example, a method has been presented which uses 360 measurements distributed over a semicircular "hoop" to find the amplitudes of the azimuthal modes radi-

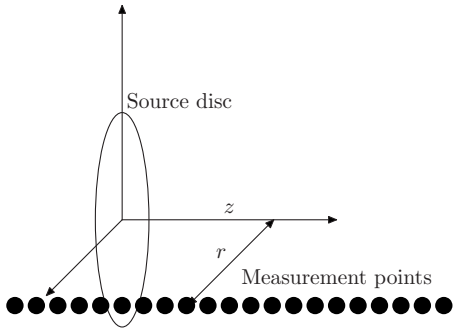


FIG. 2. Arrangement of experimental measurements.

ated from a duct.[2] The hoop of microphones was then moved to find the modal amplitudes as a function of axial displacement z.

From the known properties of rotating acoustic fields and established experimental techniques, it is clear that it is possible to measure and/or extract the complex amplitude of a single azimuthal mode radiated by a disk-shaped source. Indeed, if the source is tonal so that the modal content does not change with time, the measurements could, in principle, be performed with only two microphones, one fixed as a phase reference, and another moving along the sideline.

A. Formulation

Figure 2 shows the basic experimental arrangement. The input to the inversion method is the amplitude of a single azimuthal mode $p(r,z)$, with r fixed. When the sound is generated by a steady rotating source, $p(r,z)$ can be found by measuring the field on one sideline. When modes of different azimuthal order are present, the field must be measured on multiple sidelines of the same radius r, varying θ, and a decomposition procedure applied to find $p(r,z)$, as discussed in the previous section.

However the acoustic field may have been measured and processed, the sound radiated by one source mode of azimuthal order n at frequency ω is found by integration over the source disk,[12]

$$p(r,z) = \int_0^1 f(r_1) \int_0^{2\pi} \frac{e^{j(kR-n\theta_1)}}{4\pi R} d\theta_1 r_1 dr_1, \tag{3}$$

where the observer is positioned at $(r,0,z)$. The aim of the inversion algorithm is to recover the radial source distribution $f(r_1)$ from the field pressures $p(r,z)$.

To begin to recover $f(r_1)$, the first stage is to rewrite Eq. (3) in a transformed coordinate system (r_2,θ_2,z) centered on the measurement sideline (Fig. 3). This transformation has been used in calculations of transient radiation from pistons[13,25] and in studies of propeller noise fields.[21–23,26] Transforming Eq. (3) gives $p(r,z)$ as an integral over a line source $K(r,r_2)$,

$$p(r,z) = \int_{r-1}^{r+1} \frac{e^{jkR}}{R} K(r,r_2) r_2 dr_2, \tag{4}$$

$$R = (r_2^2 + z^2)^{1/2},$$

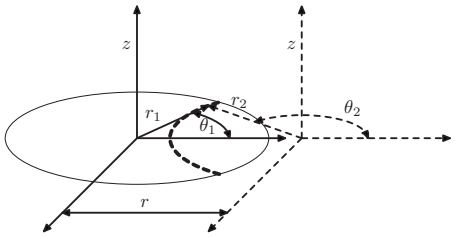


FIG. 3. Coordinate systems (r_1,θ_1,z) and (r_2,θ_2,z). The (r_2,θ_2,z) system is denoted by a dashed line, and the thick line shows the region of integration over θ_2 in the transformed system.

$$K(r,r_2) = \frac{1}{4\pi} \int_{\theta_2^{(0)}}^{2\pi-\theta_2^{(0)}} e^{-jn\theta_1} f(r_1) d\theta_2 \tag{5}$$

for observer positions with $r>1$. The original coordinates (r_1,θ_1) are related to (r_2,θ_2) by

$$r_1^2 = r^2 + r_2^2 + 2rr_2 \cos\theta_2, \tag{6a}$$

$$\theta_1 = \tan^{-1} \frac{r_2 \sin\theta_2}{r + r_2 \cos\theta_2}, \tag{6b}$$

so that the limits of integration in Eq. (5) are given by setting $r_1=1$,

$$\theta_2^{(0)} = \cos^{-1} \frac{1 - r^2 - r_2^2}{2rr_2}. \tag{7}$$

The function $K(r,r_2)$ depends only on the observer lateral displacement and is constant for all points on a sideline parallel to the source axis. The inversion method proposed is to measure $p(r,z)$ at fixed r, invert Eq. (4) to recover $K(r,r_2)$, and then use Eq. (5) to recover $f(r_1)$.

B. Inversion algorithm

The first stage of the inversion procedure is to use measured sideline data to recover the source function $K(r,r_2)$. Noting the behavior of K at its end points, Eq. (A3), we write

$$K(r,r_2) = [(r_2 - (r-1))(r+1-r_2)]^{1/2} K'(r,r_2). \tag{8}$$

The integral of Eq. (4) is discretized to give

$$\sum_{i=1}^{N} \frac{e^{jkR_{ij}}}{R_{ij}} (r + t_i^{(N)}) w_i^{(N)} K_i' = p_j, \tag{9}$$

where

$$R_{ij} = [(r + t_i^{(N)})^2 + z_j^2]^{1/2},$$

where z_j is the axial displacement of the jth measurement point and $(t_i^{(N)}, w_i^{(N)})$ are the nodes and weights of an N-point Gauss–Chebyshev quadrature rule of the second kind.

Equation (9) can be written as a system of equations relating the vector of measured pressures $\mathbf{p}$ to the unknown vector of sources $\mathbf{K}'$,

$$[\mathbf{A}]\mathbf{K}' = \mathbf{p}, \tag{10}$$

$$A_{ji} = \frac{e^{jkR_{ij}}}{R_{ij}}(r + t_i^{(N)})w_i^{(N)}. \tag{11}$$

In practice, the system will be overdetermined, with the number of measured pressures M being greater than N, the number of values of K' to be determined. At this stage, the system is solved for $\mathbf{K}'$, using some suitable method for ill-conditioned problems, with K being recovered from Eq. (8).

The second stage in determining the source distribution is to invert Eq. (5) to recover $f(r_1)$. We proceed by approximating $f(r_1)$ as a sum of Legendre polynomials $P_q(r_1)$,

$$f(r_1) = \sum_{q=0}^{Q} F_q P_q(r_1), \tag{12}$$

so that

$$K(r, r_2) = \frac{1}{4\pi}\sum_{q=0}^{Q} F_q \int_{\theta_2^{(0)}}^{2\pi-\theta_2^{(0)}} e^{-jn\theta_1} P_q(r_1) d\theta_2, \tag{13}$$

giving rise to the system of equations

$$[\mathbf{B}]\mathbf{F} = \mathbf{K}, \tag{14}$$

$$B_{iq} = \frac{1}{4\pi}\int_{\theta_2^{(0)}}^{2\pi-\theta_2^{(0)}} e^{-jn\theta_1} P_q(r_1) d\theta_2,$$

$$r_2 = r + t_i^{(N)}. \tag{15}$$

The integration is performed using a standard Gauss–Legendre quadrature. As before, this system can be solved using a method suitable for ill-conditioned problems and $f(r_1)$ reconstructed from the coefficient vector $\mathbf{F}$.

C. Far-field limitations

The integral of Eq. (4) is identical to the exact integral of Eq. (3). If we make the standard far-field approximations, we can establish some limit on the accuracy of reconstruction possible using far-field results. Expanding R to first order in r_2,

$$R \approx R_0 + \frac{r}{R_0}(r_2 - r), \quad R_0 = [r^2 + z^2]^{1/2},$$

so that

$$p \approx \frac{e^{jk(R_0 - r^2/R_0)}}{R_0}\int_{r-1}^{r+1} e^{jkrr_2/R_0} K(r, r_2) r_2 dr_2, \tag{16}$$

which can be rewritten as

$$p \approx \frac{e^{jk(R_0 - r^2/R_0)}}{R_0}\int_{-\infty}^{\infty} e^{j\alpha r_2} K(r, r_2) r_2 H(r_2 - (r-1)) \times H(r + 1 - r_2) dr_2,$$

$$\alpha = kr/R_0, \tag{17}$$

where $H(\cdot)$ is the Heaviside step function.[27]

In the far field, Eq. (17) shows that the measured pressure on a sideline is proportional to a band-limited Fourier transform of the source term $K(r, r_2)$. A reconstruction algorithm based on far-field measurements can only recover components of $K(r, r_2)$ with spatial frequency $0 \leq \alpha \leq k$, with components outside this frequency band being lost in tunneling across the transition region between the near and far fields.

This result provides a link between NAH and the method of this paper. In NAH, a Fourier transform on the sideline data, i.e., Eq. (17), is used to recover the coefficients of a field expansion in cylindrical wave-functions.[17,18] This leads to difficulties with finite aperture effects due to the periodicity enforced by the finite Fourier transform. In this algorithm, no use is made of the Fourier transform in the reconstruction procedure so that the shortcomings of the finite Fourier transform do not cause the spurious sources which appear in NAH. On the other hand, the discretization introduced by the finite number of samples on the sideline can lead to aliasing as in NAH and any other reconstruction procedure based on spatial sampling of the acoustic field. The implication of Eq. (17) is that in order to use the information, which is present on the sideline, the sampling rate must be such as to capture behavior up to wavenumber k. If the minimum sampling rate is taken to be twice per wavelength, then the sideline measurements should be taken no more than π/k apart.

III. RESULTS

Two test cases have been simulated as a first test of the algorithm of Sec. II B. The first uses parameters characteristic of the counter rotating integrated shrouded propfan (CRISP) ducted rotor tests,[1] and the second models an automotive cooling fan which has been used in tests of noise control.[7,8] In each case, the sound field $p(z)$ is computed by integration of Eq. (3). To simulate measurement errors and background noise, a Gaussian random signal of amplitude $\epsilon \max|p|$ is added to the computed pressures before using them in the inversion scheme.

For the ducted rotor test case, $M_t = 0.74$, $k = 7.4$, $n = 10$, $M = 128$, $N = 64$, $r = 1.125$, and $0 \leq z \leq 4$. The source $f(r_1)$ was synthesized by adding the first four duct modes of circumferential order n with a random phase so that the source was given by

$$f(r_1) = \sum_{m=1}^{4} e^{j\phi_m} J_n(k_{mn} r_1),$$

where ϕ_m is a random phase $0 \leq \phi_m \leq 2\pi$. The number of measurement points M was chosen to be approximately equal to that in the CRISP tests where data were taken at 120 points.[1] Since the aim of the calculation was to assess the ability of the technique to resolve a source with multiple radial modes present, the modal amplitudes were kept equal and the phase randomized to generate an oscillatory source term.

In the cooling fan case, $M_t = 0.14$, $k = 0.84$, $n = 6$, $M = 16$, $N = 16$, $r = 1.25$, and $0 \leq z \leq 8$. This time, the source used was $f(r_1) = (1 - r_1)^{1/2}$, as this gives a reasonable physical behavior near the blade tip.[20] Again, the number of sensor positions

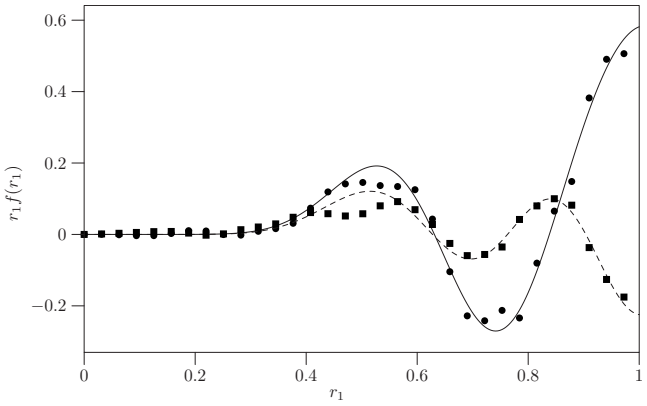


FIG. 4. Ducted fan test case, source term, $\epsilon=0$; solid and dashed lines, $\Re(f)$ and $\Im(f)$; circles and squares, $\Re(g)$ and $\Im(g)$.

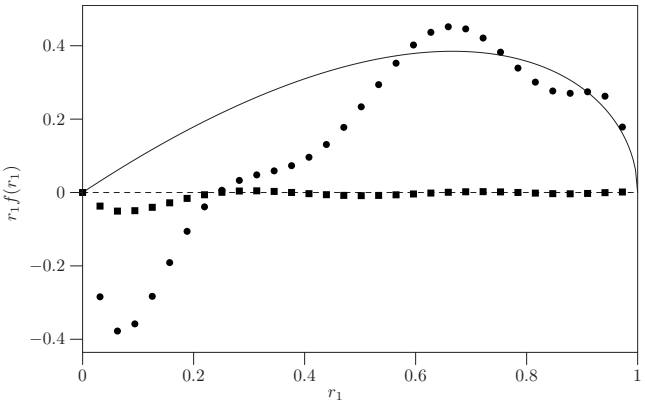


FIG. 6. Cooling fan test case, source term, $\epsilon=0$.

was chosen to be similar to that used in the original work: in this case, 17 microphones were used in the authors' source reconstruction experiments.[7,8]

The two test cases which have been chosen represent two realistic problems with quite different characteristics. The CRISP case is similar to many wind-tunnel tests which aim to extract the acoustic source from in-field measurements: the source is quite high frequency, and the tip Mach number is such that although energy is lost in the transition to the far field, the acoustic field is quite strong and there is sufficient information to allow the source to be determined reasonably accurately. The low-speed cooling fan, however, presents a rather more difficult problem. Due to the low rotor speed, the field decays rapidly inside the sonic radius $r^*=7.14$, and the measured field has lost much of the content useful for source reconstruction.

The inversion method has been implemented using OCTAVE (Ref. 28) and the REGULARIZATION TOOLS package of Hansen.[29,30] Equation (10) is solved using Hansen's implementation of truncated singular value decomposition,[31] with the regularization parameter automatically selected using the L-curve criterion.[32] The same technique is then used to solve Eq. (14) and to find $f(r_1)$.

Two measures are used to assess the accuracy of the method. The first is to compare the recovered source $g(r_1)$ with the input $f(r_1)$. The second is to use $g(r_1)$ to compute the acoustic field $q(r,z)$ at a new set of points and compare this field to $p(r,z)$ computed using $f(r_1)$. This assesses the ability of the algorithm to "project" measured data into the field.

A. Source reconstruction

The inversion algorithm has been run with zero added noise and with $\epsilon=10^{-3}$, equivalent to a maximum signal-to-noise ratio of 60 dB. Figures 4–7 show the reconstructed source for the two test cases, with the source terms weighted on radius r_1, as in the radiation integrals. The source reconstruction in the ducted fan case (Figs. 4 and 5) is quite good in both cases. With zero noise, it accurately reproduces the shape and amplitude of the input source. With added noise, the reconstruction is not quite as good, especially for inboard $r_1 \lessapprox 0.8$, but the details of the source are captured quite well near $r_1=1$, the dominant region for radiation at this wavenumber.

The cooling fan results (Figs. 6 and 7) are not as good, probably because the number of sensors is quite small and because the acoustic field is so much weaker than in the ducted fan case due to the low rotor speed. As discussed in Sec. II, sound from source regions inside the sonic radius decays exponentially as it radiates. Here the whole source lies inside the sonic radius, meaning that the acoustic field is composed largely of evanescent waves, making source reconstruction difficult.

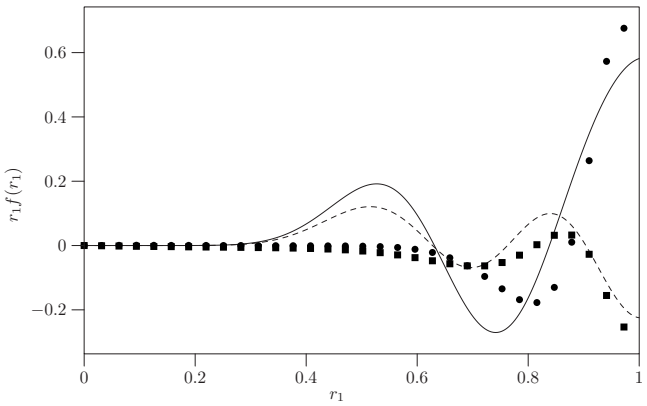


FIG. 5. Ducted fan test case, source term, $\epsilon=10^{-3}$.

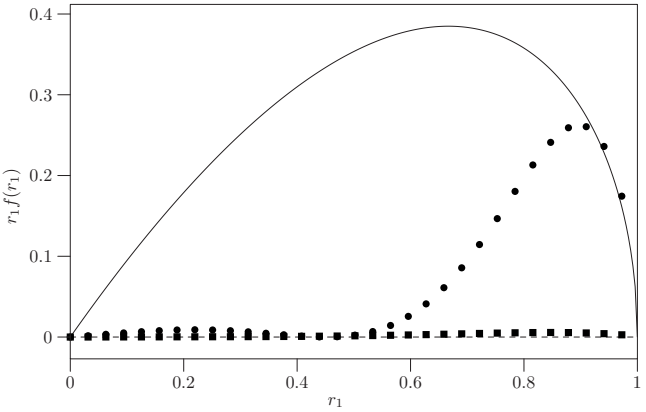


FIG. 7. Cooling fan test case, source term, $\epsilon=10^{-3}$.

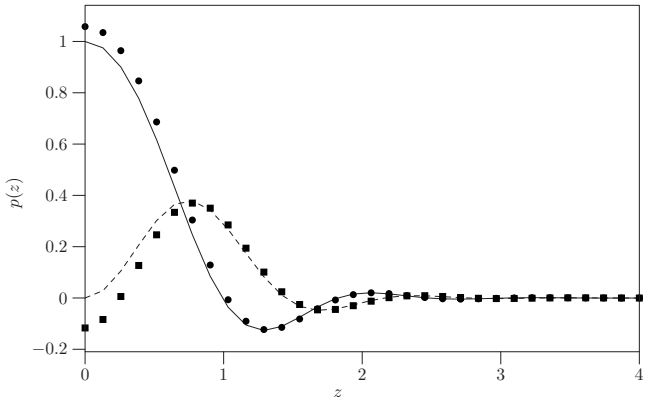


FIG. 8. Ducted fan test case, reconstructed near-field noise, $\epsilon=10^{-3}$: solid and dashed lines, $\Re(p)$ and $\Im(p)$; circles and squares, $\Re(q)$ and $\Im(q)$.

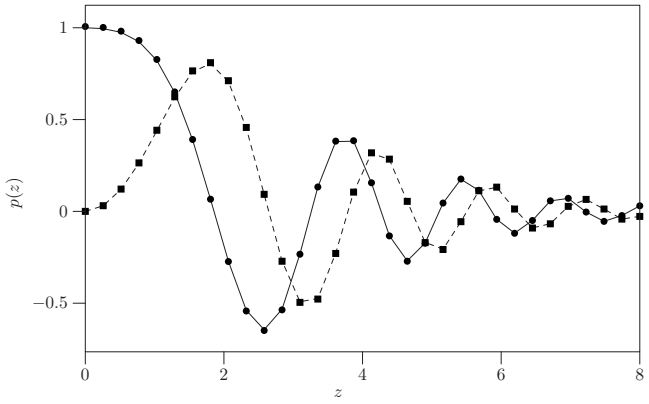


FIG. 10. Ducted fan test case, reconstructed far-field noise, $\epsilon=0$.

In Fig. 6, the reconstructed source oscillates considerably at small radii, but the tip behavior is very well captured. This might be expected: the tip is strongly dominant, meaning that the recovery of the inboard source is very poorly conditioned. With noise added, the reconstructed source is smoother, although the amplitude is not found accurately. The tip behavior, however, is again accurately computed.

B. Field estimation

Figures 8–12 compare the field computed using $g(r_1)$ to the real field $p(r,z)$, near $(r=2)$ and far from $(r=8)$ the source disk for the $\epsilon=10^{-3}$ case. The results have been scaled on $p(r,0)$ to simplify the comparison. Real and imaginary parts are shown separately as a check on the ability of the method to calculate the phase of the field, important in scattering calculations and in control.

The ducted fan results (Figs. 8 and 9) are very good. The phase has been accurately computed in the near and far fields, and the amplitude error is about 10% of the peak amplitude, or 1 dB. The directivity of the source is such that the field does not decay rapidly on the sideline, aiding the reconstruction technique. As a check that the method does converge to a correct result in the absence of noise, the reconstruction method has also been applied to data with $\epsilon=0$. The recomputed far-field pressures are shown in Fig. 10 and, as they should be, are very close to the correct data, indeed practically indistinguishable from them with the amplitude error at $z=0$ being 0.06 dB.

In the cooling fan case (Figs. 11 and 12), the field decays rapidly and is reconstructed quite poorly. The shape and phase are roughly correct, but the amplitude error is about 50% or 4 dB. The error may be due to the form of the field or to the small number of sensors simulated. Note that although the amplitude of the reconstructed source is much less than that of $f(r_1)$, the reconstructed field amplitude is rather larger. This is due to the exponential dominance of the tip region as an acoustic source on subsonic rotors, mentioned in Sec. II: the difference in tip gradient for $g(r_1)$ has made the computed acoustic field stronger than that found using $f(r_1)$.

In any case, given that the phase has been accurately computed, the result might still be useful in control applications where the phase of the control signal is important in canceling the unwanted noise. Again, we present results with no added noise (Fig. 13), and, here, the comparison is not as good as in the ducted rotor case. The shape of the field has been well captured, but the amplitude is overestimated by about 3% or 0.26 dB.

C. Algorithm performance

To assess the performance of the method when used with varying numbers of measurements, the source reconstruction method has been applied to the simulated CRISP

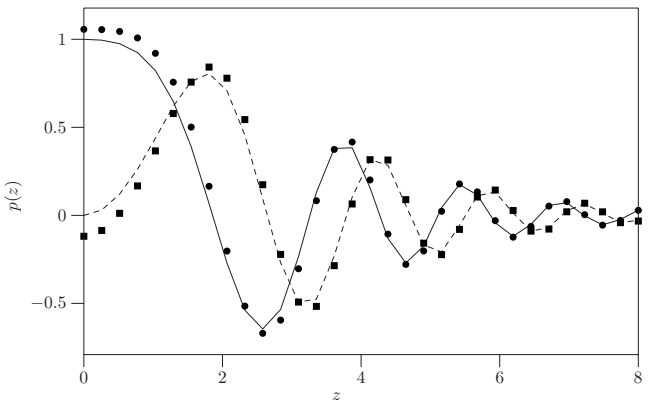


FIG. 9. Ducted fan test case, reconstructed far-field noise, $\epsilon=10^{-3}$.

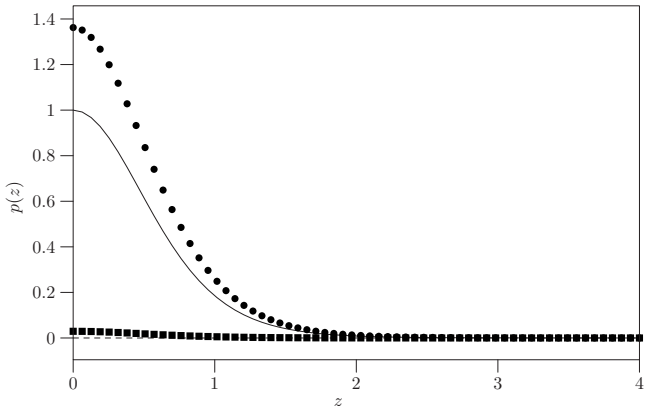


FIG. 11. Cooling fan test case, reconstructed near-field noise, $\epsilon=10^{-3}$.

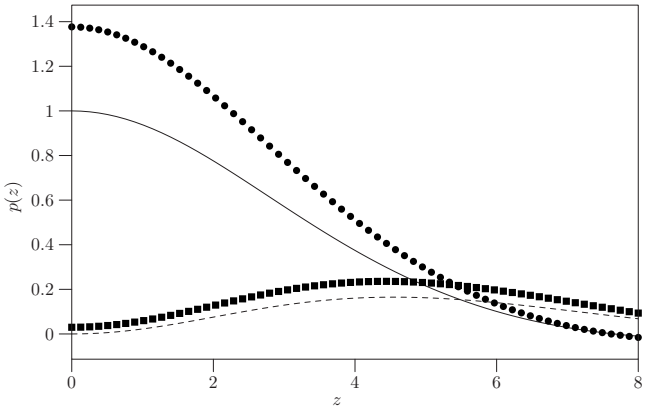


FIG. 12. Cooling fan test case, reconstructed far-field noise, $\epsilon=10^{-3}$.

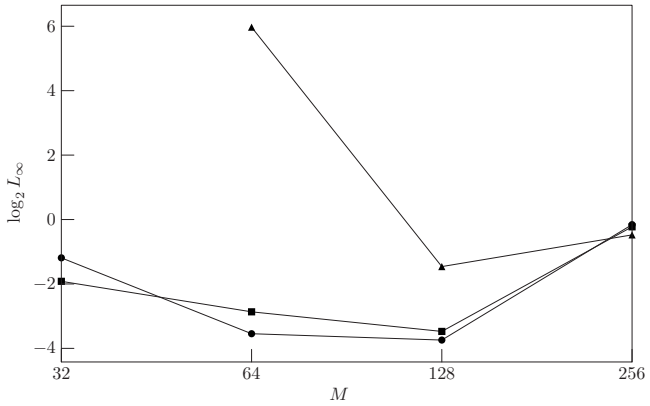


FIG. 14. Error ϵ vs number of sensor positions M and M/N. Squares: $M=N$; circles: $M=2N$; triangles: $M=4N$.

data with $M=32$, 64, 128, 256 and $M/N=1,2,4$. The error measure for the reconstructed source is the L_∞ norm,

$$L_\infty = \frac{\max|r_1 f(r_1) - r_1 g(r_1)|}{\max|r_1(f(r_1)\ |},$$

where the weighting with r_1 has been retained, corresponding to an area weighting of the error. The calculation has been performed with no added noise to check the factors which contribute to the error in the reconstructed quantities.

Figure 14 shows the variation in error with the number of sensors, as a function of the ratio M/N. The first obvious point is that for this set of operating parameters, the error for $M=64$ is very large when $M/N=4$, while the method failed completely at $M=32$. This appears to indicate that the source term cannot be well approximated by only $Q=16$ terms in the expansion of Eq. (12), an unsurprising result.

More interesting is that for $M/N=1,2$, the error decreases steadily as M increases but then increases between $M=128$ and $M=256$. Figure 15 shows the condition number κ of the matrices $[\mathbf{A}]$ and $[\mathbf{B}]$ used in the inversion procedure, as a function of M, with $N=M$. As might be expected, the condition number of both increases with M, as the systems become more poorly conditioned. Machine precision on the computer used for the calculations is approximately $1/2^{52}$. The condition number $\kappa(\mathbf{A})$ of the matrix used to estimate $K(r,r_2)$ is always greater than 2^{60} so that the first part of the inversion scheme is always ill conditioned. The reason for the drop in accuracy past $M=128$ seems to be the higher condition number of the second matrix used $\kappa(\mathbf{B})$. At $M=128$, it rises above 2^{52}, and we conjecture that at this point the loss of precision in calculations is too great for the inversion method and the results begin to worsen. This is a function of the solver used, and it may be that a different choice of regularization scheme would lead to better results.

IV. CONCLUSIONS

A source reconstruction method for the inversion of spinning acoustic fields has been developed and tested on two representative problems. It has been found that the method can work well, even with added noise, depending on the type of source to be identified. The method requires no *a priori* assumptions about the form of the source other than that it be circular and vary sinusoidally in azimuth. This makes it a useful intermediate between near-field acoustical holography, where no information is assumed about the source except its approximate location, and other source identification methods which use assumptions about the location and spatial variation of the source to model its radiation characteristics. In the case of the method of this paper, it may be that when additional information about the source is available, such as its modal structure, users might be able to incorporate this information into the technique to improve

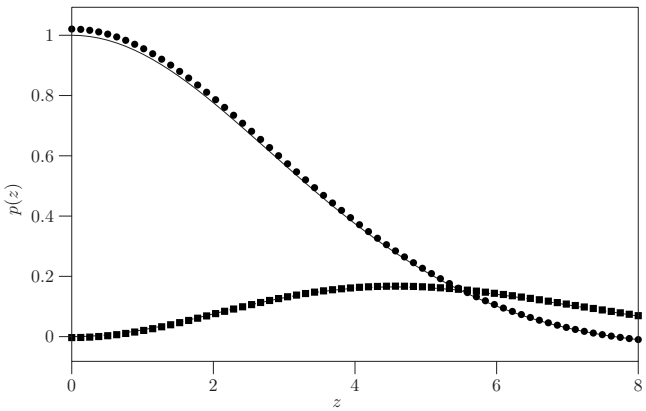


FIG. 13. Cooling fan test case, reconstructed far-field noise, $\epsilon=0$.

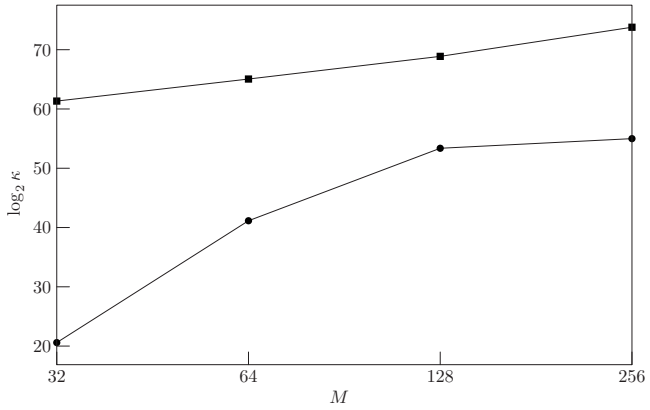


FIG. 15. Condition number κ of reconstruction matrices vs M, $N=M$. Squares: $[\mathbf{A}]$; circles: $[\mathbf{B}]$.

source reconstruction and/or to reduce the number of measurements required.

APPENDIX: END-POINT BEHAVIOR OF $K(r, r_2)$

To establish the behavior of $K(r, r_2)$ near the end points of the integrand in Eq. (4), we note that as $r_2 \to (r-1)$, $\theta_2^{(0)} \to \pi$ and $\theta_1 \to 0$. When $r_2 \to (r+1)$, $\theta_2^{(0)} \to \pi$ and $\theta_1 \to \pi$: We examine the basic integral,

$$K = \frac{1}{4\pi} \int_{\theta_2^{(0)}}^{2\pi - \theta_2^{(0)}} e^{-jn\theta_1} d\theta_2. \tag{A1}$$

For $\theta_2^{(0)} \to \pi$ and resulting small θ_1,

$$K \approx \frac{1}{4\pi} \int_{\theta_2^{(0)}}^{2\pi - \theta_2^{(0)}} 1 d\theta_2. \tag{A2}$$

Integrating,

$$K \approx (2\pi - 2\theta_2^{(0)})/4\pi$$

and using $\cos^{-1} x \to \pi - (1-x^2)^{1/2}$ as $x \to -1$ to give

$$\theta_2^{(0)} \approx \pi - \frac{2^{1/2}(r+1-r_2)^{1/2}(r_2-(r-1))^{1/2}}{(2rr_2)^{1/2}}$$

yields

$$K \approx \frac{(1+r-r_2)^{1/2}(r_2-(r-1))^{1/2}}{2(rr_2)^{1/2}}, \tag{A3}$$

with square root behavior as $r_2 \to (r-1)^+$ and $r_2 \to (r+1)^-$.

[1] F. Holste and W. Neise, "Noise source identification in a propfan model by means of acoustical near field measurements," J. Sound Vib. **203**, 641–665 (1997).

[2] F. Farassat, D. M. Nark, and R. H. Thomas, "The detection of radiated modes from ducted fan engines," in the Seventh AIAA/CEAS Aeroacoustics Conference, American Institute of Aeronautics and Astronautics, Maastricht (2001), Paper No. AIAA 2001-2138.

[3] S. Lewy, "Inverse method predicting spinning modes radiated by a ducted fan from free-field measurements," J. Acoust. Soc. Am. **117**, 744–750 (2005).

[4] S. Lewy, "Numerical inverse method predicting acoustic spinning modes radiated by a ducted fan from free-field test data," J. Acoust. Soc. Am. **124**, 247–256 (2008).

[5] F. O. Castres and P. F. Joseph, "Mode detection in turbofan inlets from near field sensor arrays," J. Acoust. Soc. Am. **121**, 796–807 (2007).

[6] F. O. Castres and P. F. Joseph, "Experimental investigation of an inversion technique for the determination of broadband duct mode amplitudes by the use of near-field sensor arrays," J. Acoust. Soc. Am. **122**, 848–859 (2007).

[7] A. Gérard, A. Berry, and P. Masson, "Control of tonal noise from subsonic axial fan. Part 1: Reconstruction of aeroacoustic sources from far-field sound pressure," J. Sound Vib. **288**, 1049–1075 (2005).

[8] A. Gérard, A. Berry, and P. Masson, "Control of tonal noise from subsonic axial fan. Part 2: Active control simulations and experiments in free field," J. Sound Vib. **288**, 1077–1104 (2005).

[9] N. Peake and W. K. Boyd, "Approximate method for the prediction of propeller noise near-field effects," J. Aircr. **30**, 603–610 (1993).

[10] P. Sijtsma, "Feasibility of noise source location by phased array beamforming in engine ducts," in the 13th AIAA/CEAS Aeroacoustics Conference, American Institute of Aeronautics and Astronautics, Rome (2007), Paper No. AIAA 2007-3696.

[11] C. R. Lowis and P. Joseph, "A focused beamformer technique for separating rotor and stator-based broadband sources," in the 12th AIAA/CEAS Aeroacoustics Conference, American Institute of Aeronautics and Astronautics, Cambridge (2006), Paper No. AIAA-2006-2710.

[12] M. Goldstein, "Unified approach to aerodynamic sound generation in the presence of solid boundaries," J. Acoust. Soc. Am. **56**, 497–509 (1974).

[13] A. D. Pierce, *Acoustics: An Introduction to Its Physical Principles and Applications* (Acoustical Society of America, New York, 1989).

[14] J. M. Tyler and T. G. Sofrin, "Axial flow compressor noise studies," SAE Trans. **70**, 309–332 (1962).

[15] J. W. Posey, M. H. Dunn, and F. Farassat, "Quantification of inlet impedance concept and a study of the Rayleigh formula for noise radiation from ducted fan engines," in the Fourth AIAA/CEAS Aeroacoustics Conference, American Institute of Aeronautics and Astronautics, Toulouse (1998), Paper No. AIAA 98-2248.

[16] M. B. S. Magalhães and R. A. Tenenbaum, "Sound sources reconstruction techniques: A review of their evolution and new trends," Acta Acust. **90**, 199–220 (2004).

[17] E. G. Williams, *Fourier Acoustics: Sound Radiation and Nearfield Acoustical Holography* (Academic, London, 1999).

[18] E. G. Williams, H. D. Dardy, and K. B. Washburn, "Generalized nearfield acoustical holography for cylindrical geometry: Theory and experiment," J. Acoust. Soc. Am. **81**, 389–407 (1987).

[19] D. G. Crighton and A. B. Parry, "Asymptotic theory of propeller noise. Part II: Supersonic single rotation propeller," AIAA J. **29**, 2031–2037 (1991).

[20] A. B. Parry and D. G. Crighton, "Asymptotic theory of propeller noise. Part I: Subsonic single rotation propeller," AIAA J. **27**, 1184–1190 (1989).

[21] C. J. Chapman, "The structure of rotating sound fields," Proc. R. Soc. London, Ser. A **440**, 257–271 (1993).

[22] M. Carley, "Sound radiation from propellers in forward flight," J. Sound Vib. **225**, 353–374 (1999).

[23] M. Carley, "Propeller noise fields," J. Sound Vib. **233**, 255–277 (2000).

[24] S. T. Hocter, "Exact and approximate directivity patterns of the sound radiated from a cylindrical duct," J. Sound Vib. **227**, 397–407 (1999).

[25] F. Oberhettinger, "On transient solutions of the 'baffled piston' problem," J. Res. Natl. Bur. Stand., Sect. B **65**, 1–6 (1961).

[26] M. Carley, "The structure of wobbling sound fields," J. Sound Vib. **244**, 1–19 (2001).

[27] M. J. Lighthill, *An Introduction to Fourier Analysis and Generalised Functions* (Cambridge University Press, Cambridge, 1958).

[28] J. W. Eaton, OCTAVE, http://www.octave.org (Last viewed November 3, 2008).

[29] P. C. Hansen, "Regularization Tools: A Matlab package for analysis and solution of discrete ill-posed problems," Numer. Algorithms **6**, 1–35 (1994).

[30] P. C. Hansen, "Regularization Tools version 4.0 for Matlab 7.3," Numer. Algorithms **46**, 189–194 (2007).

[31] P. C. Hansen, "Regularization Tools: A Matlab package for analysis and solution of discrete ill-posed problems," Technical report, Technical University of Denmark, www.netlib.org/numeralgo (Last viewed November 3, 2008).

[32] P. C. Hansen and D. P. O'Leary, "The use of the L-curve in the regularization of discrete ill-posed problems," SIAM J. Sci. Comput. (USA) **14**, 1487–1503 (1993).

Inferring the acoustic dead-zone volume by split-beam echo sounder with narrow-beam transducer on a noninertial platform

Ruben Patel,[a] Geir Pedersen, and Egil Ona
Institute of Marine Research, P.O. Box 1870 Nordnes, NO-5817 Bergen, Norway

(Received 19 November 2008; accepted 25 November 2008)

Acoustic measurement of near-bottom fish with a directional transducer is generally problematical because the powerful bottom echo interferes with weaker echoes from fish within the main lobe but at greater ranges than that of the bottom. The volume that is obscured is called the dead zone. This has already been estimated for the special case of a flat horizontal bottom when observed by an echo sounder with a stable vertical transducer beam [Ona, E., and Mitson, R. B. (1996). ICES J. Mar. Sci. **53**, 677–690]. The more general case of observation by a split-beam echo sounder with a transducer mounted on a noninertial platform is addressed here. This exploits the capability of a split-beam echo sounder to measure the bottom slope relative to the beam axis and thence to allow the dead-zone volume over a flat but sloping bottom to be estimated analytically. The method is established for the Simrad EK60 scientific echo sounder, with split-beam transducers operating at 18, 38, 70, 120, and 200 kHz. It is validated by comparing their estimates of seafloor slope near the Lofoten Islands, N67-70, with simultaneous measurements made by two hydrographic multibeam sonars, the Simrad EM1002/95 kHz and EM300/30 kHz systems working in tandem.
 [DOI: 10.1121/1.3050325]

PACS number(s): 43.30.Ft, 43.30.Gv, 43.30.Sf [ADP] Pages: 698–705

I. INTRODUCTION

Difficulties in applying acoustic survey methods to fish species distributed close to the bottom are well known (Cushing, 1968, 1983; Simmonds and MacLennan, 2005; Godø and Wespestad, 1993; Stanley *et al.*, 2000; Krieger *et al.*, 2001). A major difficulty has been the failure of proper detection of fish in the so-called "acoustic dead zone" (Ona and Mitson, 1996; Lawson and Rose, 1999). This acoustic dead zone is characterized by overlapping echoes of fish and bottom at ranges beyond that of the first contact of the acoustic pulse with the bottom. If the bottom is sloping and/or the transducer platform is pitching and rolling, the dead-zone volume generally increases.

As most demersal species exhibit diel vertical migration, the proportion of fish within the acoustic dead zone may introduce a large and variable bias, equally present in concurrent bottom trawl survey methods (Godø and Wespestad, 1993; Michalsen *et al.*, 1996; Casey and Myers, 1998; Aglen *et al.*, 1999; Hjellvik *et al.*, 2002, 2003). Combined trawl and acoustic survey methods, as suggested earlier by Godø and Wespestad (1993), Aglen (1996), Everson *et al.* (1996), and Krieger *et al.* (2001), may offer a better solution for a total combined estimate of stock size and composition.

The examples of acoustic dead-zone volume corrections in Ona and Mitson (1996) make use of supplementary data on fish density outside but in the vicinity of the particular acoustic dead zone. This was done to extrapolate the density into the acoustic dead zone, but it considered only the horizontal flat-bottom case when estimating the effective height of the acoustic dead-zone volume. However, on a sloping bottom, the acoustic dead-zone volume is greater, and it increases rapidly with increasing bottom slope. This is similarly true for the more general case of an accelerating transducer platform without stabilization or motion compensation. For transducers that are mounted, for example, on the hull of a fishery research vessel, mechanical or electronic stabilization that fully compensates for the movement of the vessel platform is rare. New acoustic technology may offer solutions to the acoustic dead-zone problem, but a better estimate of the acoustic dead-zone volume of current echo sounder systems is still very useful.

In this paper the acoustic dead-zone volume is computed for the general case of an accelerating noninertial transducer platform without motion compensation. As in the cited acoustic dead-zone paper by Ona and Mitson (1996), the shape of the acoustic beam is incorporated explicitly in the theory. Application of the new algorithm depends on knowledge of the slope of the bottom. Its measurement by a split-beam scientific echo sounder is verified through a comparison with measurements by two hydrographic multibeam sonars working in tandem. Given echo sounder determination of the bottom slope, the effective acoustic dead-zone volume is then computed, enabling the fish density estimate to be refined. This work is part of an ongoing effort to improve methods for estimating the density of targets close to the bottom.

II. THEORY

For the general case of a noninertial transducer platform, the acoustic beam at any arbitrary time will be oriented obliquely to the bottom, whether or not this is horizontal. For directional beams and ranges for which the bottom roughness is small over the extent of the central beam cross sec-

[a]Electronic mail: ruben@imr.no

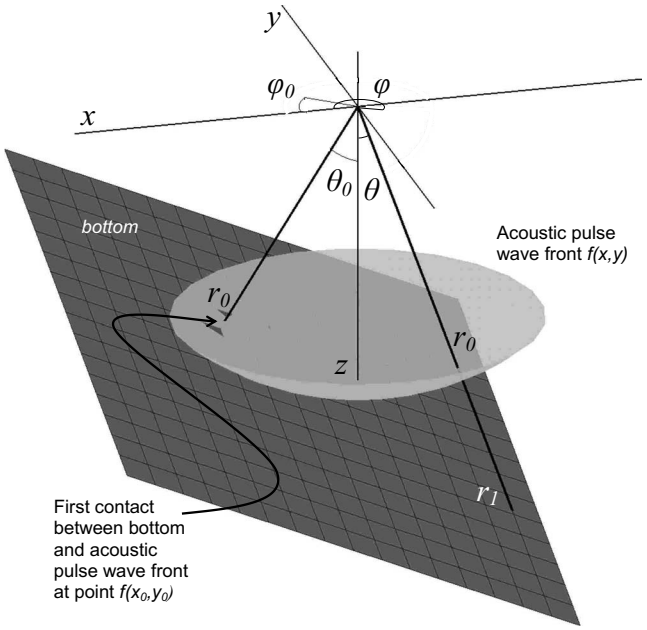


FIG. 1. Schematic diagram of the three-dimensional geometry of an acoustic wave front hitting a sloping bottom. The angles θ_0 and φ_0 define the slope of the plane corresponding to the flat bottom. The integration variables in the dead-zone volume calculation are $r=[r_0,r_1]$, $\theta=[0,\pi/2]$ and $\varphi=[0,2\pi]$. The acoustic wave front is modeled by a spherical plane $f(x,y)$ with radius r_0. The bottom is plane z, which is tangential to the spherical surface $f(x,y)$, with the common point $[x_0,y_0,f(x_0,y_0)]$. This is also where the acoustic pulse first hits the bottom. In spherical coordinates the position of the point where the bottom intersects the acoustic pulse is given by (r_0,φ_0,θ_0).

tion, the bottom may be considered flat. The problem of estimating the dead zone by a moving beam on a noninertial platform is thus tantamount to estimating the dead-zone volume by a vertical beam over a flat but sloping bottom. This insonification geometry is illustrated in Fig. 1.

By inspection and by analogy with the earlier expression given for the dead-zone volume associated with insonification of a flat horizontal bottom by a stable vertical beam (Ona and Mitson, 1996), the dead-zone volume can be expressed thus as

$$V=\int_0^{\pi/2}\int_0^{2\pi}\int_{r_0}^{r_1} r^2 b^2(\theta,\varphi)\sin\theta\, dr d\varphi d\theta, \tag{1}$$

where r, θ, and φ are spherical coordinates and $b(\theta,\varphi)$ is the transducer beam pattern. The beam pattern effectively weights the integration volume under assumption of random target positions. Use of this two-dimensional beam pattern generalizes the earlier expression, which considered a circularly symmetrical beam pattern that depended only on the polar angle. The two angles θ and φ specify the direction of r over which the integration is performed. The lower limit r_0 is the nearest range to the bottom, in direction (θ_0,φ_0). The upper limit r_1 is the range to the bottom in the general direction (θ,φ). In a single direction (θ_0,φ_0), $r_1=r_0$; otherwise $r_1>r_0$.

According to the model, the bottom is a flat surface tangential to the spherical wave front at the point (r_0,θ_0,φ_0) where the bottom is first detected. For the spherical function

$$f(x,y)=\sqrt{r_0^2-x^2-y^2}, \tag{2}$$

the tangent plane at (x_0,y_0) is

$$z=f(x_0,y_0)+\frac{\partial f(x_0,y_0)}{\partial x}(x-x_0)+\frac{\partial f(x_0,y_0)}{\partial y}(y-y_0),$$

$$z=m-\frac{x_0(x-x_0)}{m}-\frac{y_0(y-y_0)}{m}, \tag{3}$$

where

$$m=\sqrt{r_0^2-x_0^2-y_0^2}.$$

The range r_1 to the tangent plane is used as the limit in Eq. (1). In spherical coordinates,

$$r_1=\frac{r_0}{\frac{m}{r_0}\cos\theta+\sin\theta_0\sin\theta\cos(\varphi_0-\varphi)}. \tag{4}$$

Substituting in the result of integrating over r in Eq. (1),

$$V=\frac{r_0^3}{3}\int_0^{\pi/2}\int_0^{2\pi}\{[S(\theta,\phi)]^{-3}-1\}b^2(\theta,\varphi)\sin\theta d\varphi d\theta, \tag{5}$$

where $S(\theta,\phi)=(m/r_0)\cos\theta+\sin\theta_0\sin\theta\cos(\varphi_0-\varphi)$.

It is now possible to relate the acoustic dead-zone volume V of a sloping flat bottom to that of a horizontal flat bottom,

$$V_0=\frac{r_0^3}{3}\int_0^{\pi/2}\int_0^{2\pi}\left\{\left[\frac{m}{r_0}\cos\theta\right]^{-3}-1\right\}\times b^2(\theta,\phi)\sin\theta d\varphi d\theta, \tag{6}$$

derived by substituting the parameter value $\theta_0=0$ in the expression for $S(\theta,\varphi)$ in Eq. (5).

III. MATERIALS

The fisheries research vessel RV "G.O. Sars" was used as the instrument platform. The data were collected during the annual survey of spawning cod, covering the shelf sea from 500 to about 50 m depth on the outside and inside of the Lofoten islands, from 67°N to 70°N, lasting from 17 During the survey the ship's conductivity, salinity and depth probe (CTD) was used to measure these parameters for sound speed calculations.

The calibrated raw data from the Simrad EK60 scientific echo sounder (Andersen, 2001), operating at 18, 38, 70, 120, and 200 kHz, were logged simultaneously with the Simrad EM1002 (95 kHz) and EM300 (30 kHz) multibeam bottom mapping systems. All EK60 transducers were mounted in one of the instrument drop-keels of the vessel in a maximum packing arrangement (Fig. 2). These transducers have nominal half-power beam widths of 7°, except for the 18 kHz transducer, which has a nominal half-power beam width of 11°. The EK60 transducers in the drop-keel are tilted forward in the alongship direction at an angle of 1.5°. Some small misalignments in athwartship and alongship directions are expected due to mounting difficulties.

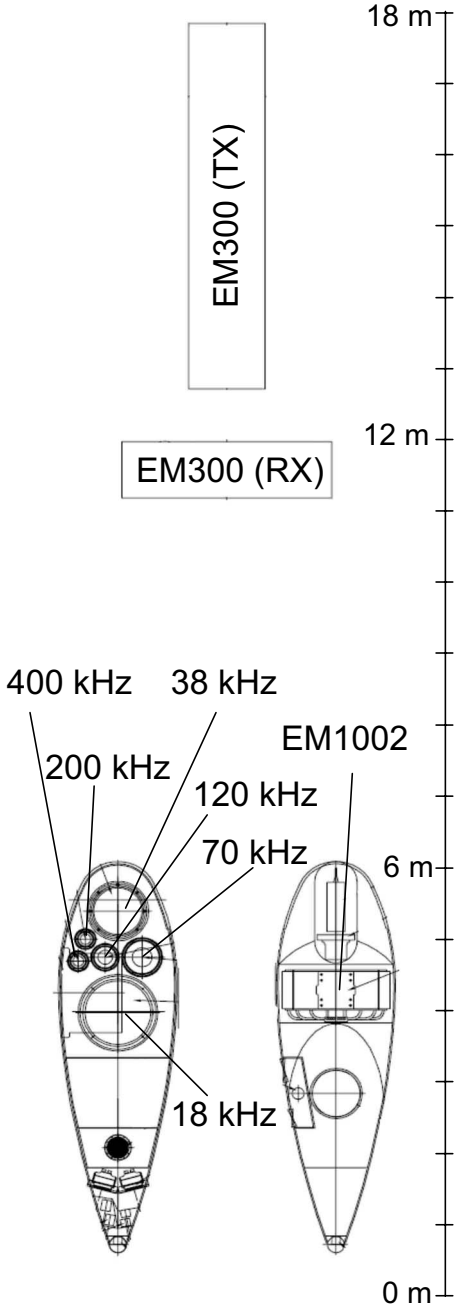


FIG. 2. Schematic representation of the packing arrangement of the transducers mounted on RV G.O. Sars in this experiment. The port drop-keel supports the 18, 38, 70, 120, 200, and 400 kHz transducers for the EK60 echo sounder. The starboard drop-keel supports the EM1002 transducer. The EM300 transducers are configured in a Mills Cross array, and these are mounted forward of the drop-keels.

Prior to the survey, the echo sounders were calibrated using standardized methods (Foote *et al.*, 1987). The standard targets used were the 64 and 60 mm diameter copper targets for 18 and 38 kHz, respectively, while a 38.1 mm tungsten carbide sphere was used for the 70, 120, and 200 kHz echo sounders (Foote, 1981, 1982, 2001; Foote *et al.*, 2005).

The echo sounders were triggered in parallel at the maximum pulse repetition frequency, transmitting the next pulse soon after the bottom echo was received. For comparability, the transmitted pulse duration was identical on all frequencies, 1.024 ms. All raw data from the echo sounders were stored and processed further in MATLAB® Version 6.0.0.88 Release 12 (MathWorks, Inc., Natick, MA).

The Seatex MRU 5 is a motion reference unit, which is mounted on RV G.O. Sars. The unit is specially designed for marine applications. It has three acceleration sensors, and outputs heave, roll, pitch, and yaw. The resolution on all axes is 0.01°, and the output data rate is 100 Hz. This is fed to the Simrad EM1002 and Simrad EM300 systems in real time, where it is used for beamforming, and added to the ping data file in the EK60 echo sounder.

Data from the bottom mapping systems, the Simrad EM1002 and Simrad EM300, were logged in parallel with data from the echo sounder, using the EM systems as master and EK60 as slave. The 30 kHz EM300 forms 135 beams per ping, each with a beam width of $1° \times 2°$, depth range of 10–5000 m, and depth resolution of 4–30 cm. The 95 kHz EM1002 forms 111 beams per ping, each with a beam width of $2° \times 2°$, depth range of 2–1000 m, and depth resolution of 2.4–8 cm.

IV. METHODS

A. Bottom slope from EM1002 and EM300

The EM1002 and EM300 multibeam sonars were operated simultaneously with each other as well as with the EK60. Data from the two sonars were processed and combined by the NEPTUN B postprocessing software (Kongsberg Maritime, 2006), with compensation for vessel motion, tidal state, and other influencing factors. This software is routinely used by the Geological Survey of Norway and also by the Norwegian Hydrographic Service (Sjøkartverket) to specify the bathymetry for navigational charts. For the depth ranges of interest, 50–500 m, the nominal accuracy of data derived from the EM1002 and EM300 sonars is 3 and 5 cm, respectively. The accuracy of the combined data is similar or better.

Data on position and bottom depth were expressed in ASCII files. These data were used to calculate the bottom slope at each position where data were collected from the EK60. This position is defined as the intersection of the average axis of the EK60 beams, without allowance for transducer offset with the bottom. The effect of transducer offset is reckoned to be negligible for the stated depth range of interest.

The bottom slope was calculated from the combined multibeam sonar data in a 20×20 m^2 square centered at each position in the following way. Each square was subdivided into nine equal square cells, and the depth of each cell was calculated by averaging over all the bottom values from the multibeam system inside the respective cell. This yielded nine depth values for each EK60 ping. Using a bicubic fit (Keys, 1981) on these nine values, the surface normal of the center of the 20×20 m^2 square was calculated. This gave the bottom slope from the multibeam sonars at the same position as data from each ping by the EK60.

B. Bottom slope from EK60

From the EK60, the phase angles of the received echoes are available at the digital sampling resolution. The hypoth-

esis is that, when carefully analyzed, both the bottom slope in the fore-aft direction and in the port-starboard direction may be extracted from the bottom echo.

In order to avoid confusion, the phase angle referred to in this paper is the mechanical angle estimated by the split-beam echo sounder system using electrical phase differences in the received signal from paired transducer halves (Ehrenberg, 1979). This configuration was originally designed to measure the position of a detected single target within the acoustic beam, with compensation of the echo amplitude for two-way beam directivity (Brede *et al.*, 1990).

Phase angles and backscattering values were read and processed from the raw echogram files using algorithms specially written for the purpose. The bottom slope from the EK60 data was calculated using the sub-bottom phase angles from each ping. In order to restrict the search for the correct sub-bottom samples, bottom depth was first calculated from the backscattering samples using all five frequencies. For each frequency the bottom was detected using the bottom detection function of the EK60. Correct bottom depths were defined as the respective median value of the detected bottom depths from all frequencies for each ping.

From this point in the analysis, only data from the 38 kHz echo sounder were investigated, and further calculations are performed relative to the bottom. This means that all 38 kHz data are related to the bottom depth rather than to the transducer. In this restricted data set, consisting of 1404 pings, the bottom would appear as a horizontal line. The variance of the phase angles in each sample depth was computed. This quantifies the phase variance as a function of depth. It was expected that the bottom as a target would give low phase variance. A variance threshold was then used to select the samples containing the most correct bottom angles, and the bottom slope was calculated by averaging the selected samples for each ping.

C. Comparing slopes from the multibeam sonar and the EK60

The bottom slope calculated from the multibeam sonar data included compensation for vessel motion and is regarded as being correct. The slope calculated directly from the EK60 data, however, depends on the bottom slope, vessel roll and pitch, and the angular offset due to the actual mounting of the transducer on the drop-keel. Further, there is usually small angular offset between the actual acoustic axis of the split-beam transducer and the electrical center of the transducer, as measured by the electrical phase angles between the transducer quadrants. The vessel roll and pitch were first removed from the EK60 data. The angle of the surface normal to the bottom, as established by the multibeam sonars at each corresponding EK60 ping position, was transformed to the EK60 transducer coordinate system. A linear regression line was fitted to the slope angle, as calculated from the multibeam sonar and EK60 bottom slope, to determine whether the EK60 data contained accurate information about the bottom slope.

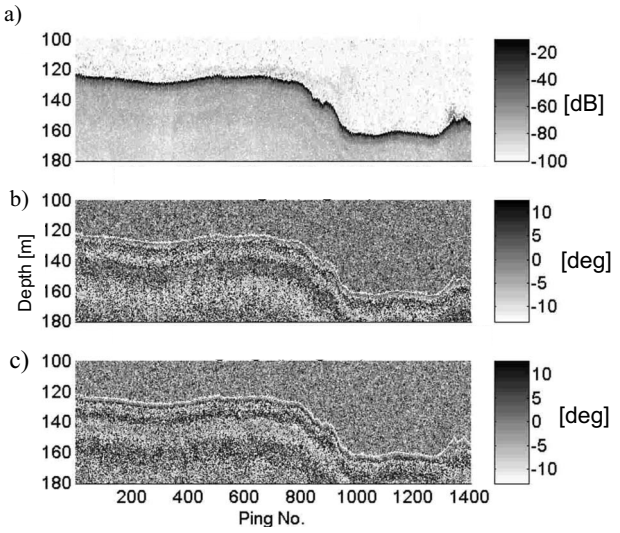


FIG. 3. The echogram and phase angles in the transducer coordinate system of the EK60/38 kHz scientific echo sounder data set used in calculating (a) volume backscattering values in decibels relative to 1 m, (b) athwartship phase angles in degrees, and (c) alongship phase angles in degrees.

D. Calculating the acoustic dead-zone volume

The acoustic dead-zone volume was calculated according to Eq. (5) for the 38 kHz transducer, assuming a circular symmetry and an ideal beam pattern (Kinsler *et al.*, 2000),

$$b(\theta) = \left(\frac{2J_1(ka \sin \theta)}{ka \sin \theta} \right)^2, \tag{7}$$

where a is the transducer radius, k is the acoustic wave number, J_1 is the Bessel function of first order and first kind, and θ is the off-axis beam angle. The nominal beamwidth is 7°; hence ka equals $1.615/[\sin(7\pi/360)]$ to a good approximation.

The integral in Eq. (5) was computed out to a maximum polar angle θ of 20° rather than $\pi/2$. This corresponds to approximately 99.9 % of the backscattered echo energy of the entire acoustic beam. Theoretical calculations are performed first for varying depths and varying bottom slopes and then for a real data set recorded during the survey. The effect of bottom depth and slope on the volume is demonstrated for a transect selected for its variability with respect to both of these parameters.

V. RESULTS

The sound speed was estimated to be 1475 m/s from the CTD data. Sampled data from the selected transect are shown in Fig. 3. A conventional echogram displaying values of volume backscattering strength is shown in Fig. 3(a). The bottom appears as the darkest line throughout the echogram. Below this line the sub-bottom echo values can be seen as gray colors, and above, the volume backscattering from the water column is seen. Single fish echoes appear as darker dots scattered in the echogram. The bottom starts on a plateau at a depth of 122 m; it then drops to a lower basin at a depth of 160 m. At the end of the echogram, between pings 1300 and 1400, a typical fish aggregation can be observed around a sharp peak of 5 m in height. Athwartship and alongship phase angles are shown in Figs. 3(b) and 3(c),

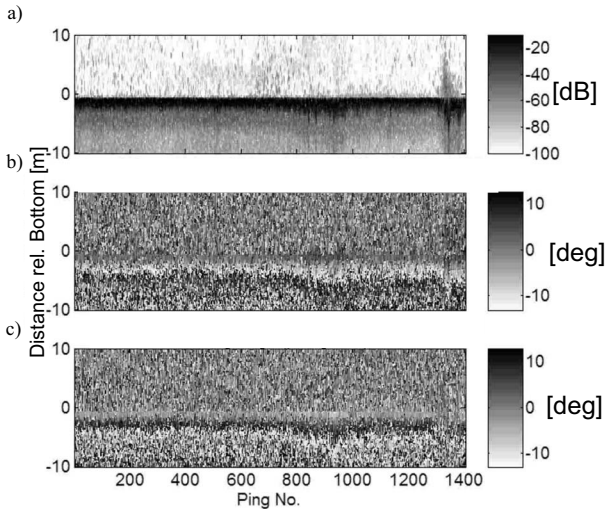


FIG. 4. The echogram and phase angles in the transducer coordinate system from the EK60/38 kHz scientific echo sounder relative to the bottom, which is shown as a straight line. (a) Volume backscattering in decibels relative to 1 m. (b) Athwartship phase angles in degrees. (c) Alongship phase angles in degrees.

respectively. The detected bottom is indicated by a white line. Corresponding bottom structures can be seen in both backscattering strengths and phase angles.

The same data are presented relative to the detected bottom, shown as a straight line, in Fig. 4. In the modified echogram in Fig. 4(a), the fish aggregation can be seen at the end of the echogram as a darker area above the bottom between pings 1300 and 1400. Athwartship and alongship phase angles are shown in Figs. 4(b) and 4(c), respectively. For sloping bottoms, the duration of the bottom echo is seen to be increased. This increase is apparent between pings 800 and 1000 and between pings 1300 and 1400, which agrees with the data in Fig. 3.

The straightened bottom data in Fig. 4 were used to calculate mean volume backscattering values and variances of the phase angles as a function of vertical distance from the bottom in Fig. 5. The bottom appears as a distinct peak in Fig. 5(a). This also corresponds to a minimum in the athwartship angle in Fig. 5(b) and in the alongship angle in Fig. 5(c). The threshold for detecting phase angles corresponding to the bottom slope is shown as a vertical line in each of Figs. 5(b) and 5(c). The variance minimum is followed by a maximum that occurs at a depth where the backscattering flattens out from its maximum. The variance in phase angles is also less before the bottom is detected than afterward.

The calculated athwartship and alongship angles from the multibeam system and the EK60 are presented in Fig. 6. The athwartship and alongship angles from both systems are displayed with respect to the ping numbers in Figs. 6(a) and 6(c), respectively. Scatter diagrams of the bottom slope as measured with the respective EM and EK systems are shown in Figs. 6(b) and 6(d) for the two angles. A linear regression analysis has been performed for each data set. In addition to the central mean regression curve, the 95 % confidence intervals of data and predictions are both shown, with the outermost pair of lines referring to the data and the innermost pair referring to the predictions. Corresponding regression

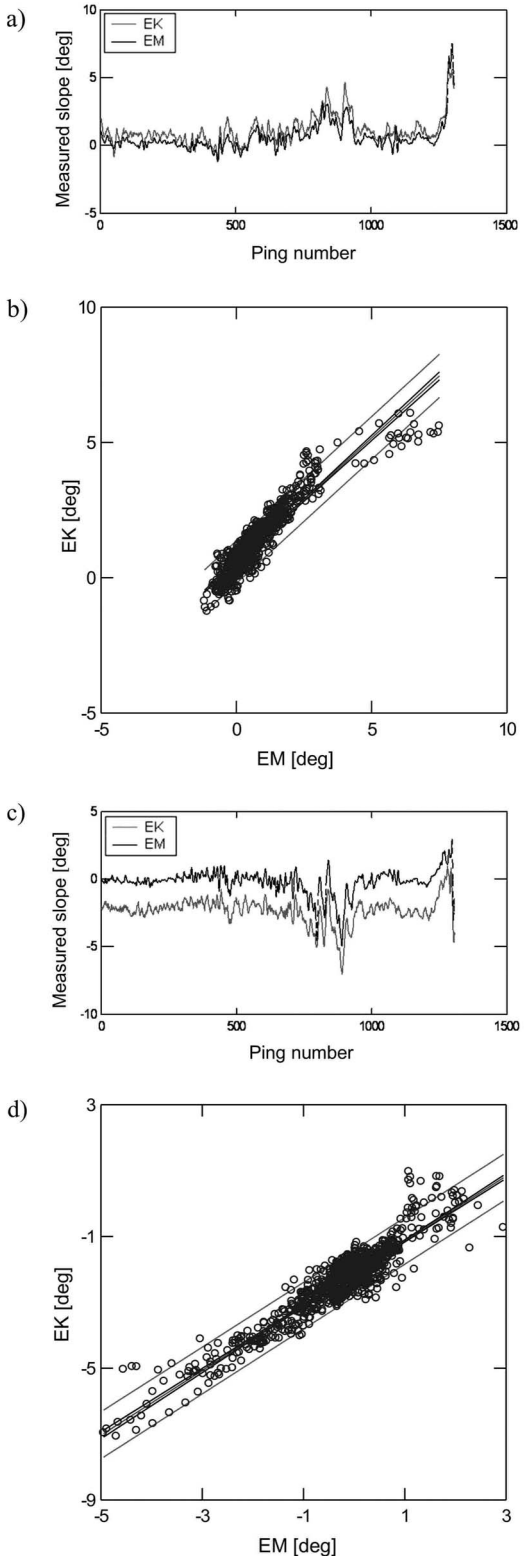


FIG. 6. Comparison of the bottom slope as measured from the EM system and the EK60/38 kHz scientific echo sounder. (a) Athwartship angle bottom slope as calculated from the EM and the EK60 data. (b) Scatter diagram of the athwartship angle with regression lines. (c) Alongship angle bottom slope as calculated from the EM and EK60 data. (d) Scatter diagram of the alongship angle with regression lines.

results with standard errors of parameters and regression are shown in Table I. The increased scatter between EM and EK data with increasing angle is noted.

The regression analysis shows that there are offsets for the EK60-measured athwartship and alongship angles. The

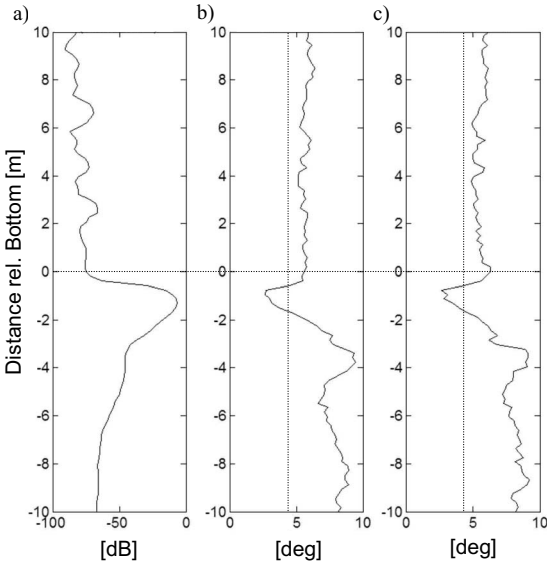


FIG. 5. Average and variance values as a function of depth for backscattering and phase values from the EK60/38 kHz scientific echo sounder. The horizontal line at depth zero indicates the bottom, and the vertical line is the threshold used for detecting relevant phase angles. (a) Mean volume backscattering in decibels relative to 1 m. (b) Standard deviation of athwartship angles in degrees. (c) Standard deviation of alongship angles in degrees.

mean offsets are 0.6° and 2.1° for the respective angles. The cumulative mean transducer offset is 2.2° relative to a flat horizontal bottom under stable sailing conditions. The corresponding acoustic dead-zone volume is 1.7 times that of the dead zone in the ideal case of vanishing offsets. The slope of the regression lines is 0.92 for the athwartship angle and 0.98 for the alongship angle, indicating a second source of error requiring compensation in the dead-zone volume.

Theoretical acoustic dead-zone volumes are presented in Fig. 7 for the particular conditions of the bottom as observed with the EK60. In Fig. 7(a), the acoustic dead-zone volume of a flat horizontal bottom is calculated as a function of depth. The dependences of the dead-zone volume on depth and bottom slope are given in Figs. 7(b) and 7(d); the relative dead-zone volume is given in Fig. 7(c). At an angle of 3°, for instance, the dead-zone volume is double that of a flat horizontal bottom. The curve in Fig. 7(c) with parameter 0° is the same as the curve in Fig. 7(a).

The bottom slope is shown in Fig. 8(a), and the bottom depth as measured with the EK60 is shown in Fig. 8(b). The bottom slope and bottom depth are used as parameters for calculating the acoustic dead-zone volume in Figs. 8(c) and 8(d).

VI. DISCUSSION

Estimation of the bottom slope from EK60 split-beam data using phase angle variance works well for relatively

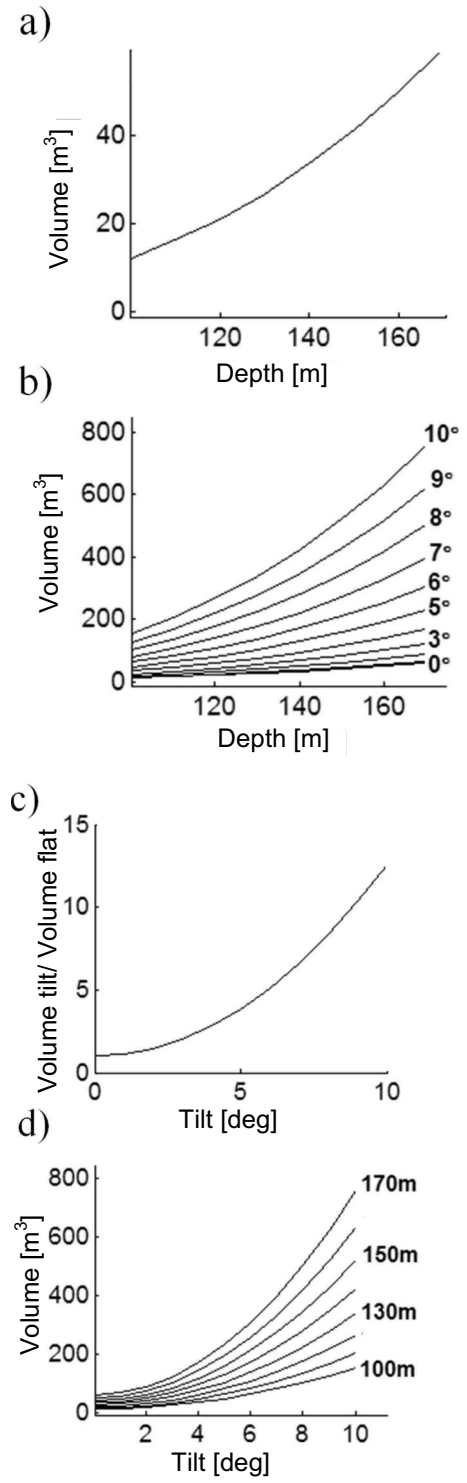


FIG. 7. Theoretical calculations of the dead-zone volume. (a) Dead-zone volume of a flat bottom as a function of depth. (b) Dead-zone volume as a function of depth for different bottom slope angles. (c) Dead-zone volume of a sloping bottom relative to a flat bottom as a function of tilt angle. (d) Dead-zone volume as a function of slope angle for different depths.

TABLE I. Results of the linear regression of measured slope on the EK60 scientific split-beam echo sounder on the corresponding slope measured on the EM multibeam systems according to the equation $\Phi_{EK}=\alpha\Phi_{EM}+\beta$, including the parameter standard errors and the standard error, SE, of the regression.

Direction	N	α	SE (α)	β	SE (β)	SE
Athwarthship	1306	0.917	0.011	0.576	0.013	0.401
Alongship	1306	0.984	0.011	−2.120	0.010	0.360

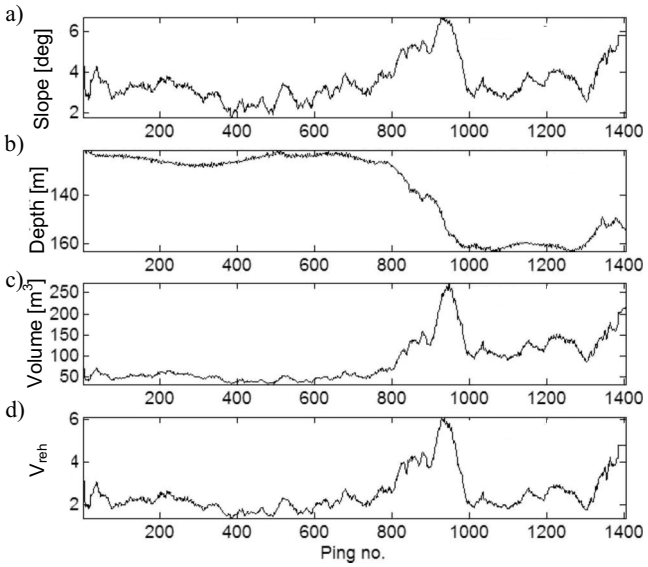


FIG. 8. Calculations based on data from the EK60/38 kHz scientific echo sounder. (a) Bottom slope θ_0. (b) Bottom depth r_0. (c) Acoustic dead-zone volume V. (d) Relative acoustic dead-zone volume (V_{rel}).

short transects. For longer transects with variable bottom slopes, the bottom data should be analyzed in shorter sections to maintain a low variance.

The phase variance function has a distinct signature, which can be used to identify samples containing the bottom slope. There is potential advantage to using phase measurements to improve bottom detection algorithms. Currently, such algorithms only make use of the bottom backscattering amplitude and measures of its stability.

Echo sounder transducers are typically mounted with an alongship tilt, which is measured by the method developed in this paper. Mounting difficulties and changing vessel-ballasting will also contribute to offsets in the measured alongship and athwartship tilts. From the regression curves in Fig. 6, it is evident that there is a good fit between the multibeam sonar data and EK60 data. Within the angular range of the split-beam detectors, 6°–7° off the acoustic axis, the regression works well. Angles outside this range will deviate from the true bottom slope and will be underestimated. For larger slopes alternative methods must be added for estimating the acoustic dead-zone volume.

Transducers are usually mounted with a tilt of 1.5° in the alongship direction, which corresponds to a relative acoustic dead-zone volume of 1.3 under stable transducer-platform operating conditions. In the examined case, other misalignments produce a mean effective tilt of 2.2°, hence with acoustic dead-zone volume that is 1.7 times that of a flat horizontal bottom. This shows that the transducer misalignment contributes significantly to the estimated volume, which should be taken into account when analyzing data from bottom fish surveys.

The theoretical curves for dead-zone volume calculations in Fig. 7 show that there is a simple relationship among bottom depth, bottom slope angle, and dead-zone volume. It is clear that calculating the relative acoustic dead-zone volume, instead of the physical acoustic dead-zone volume, for all depths and slope angles could speed up the calculations for applications at sea. It is noted that tilt angles of 3° will double the acoustic dead-zone volume. Typical slopes of the sea bottom in the geographical distribution area of the North East Arctic cod stock are between 0° and 4°. These angles are also characteristic of rolls of a large research vessel when conducting surveys on this stock.

It is possible to determine the acoustic dead-zone volume associated with observation by a directional split-beam transducer mounted on a noninertial platform. The split-beam phase information is sufficient to determine the bottom slope. Scientific echo sounder systems currently used for fish-stock abundance estimation can be used for a direct measurement of the bottom slope, enabling the acoustic dead-zone volume to be estimated. This information can be used further for extrapolating fish density from areas near the acoustic dead zone into the dead-zone volume to reduce a recognized variable bias in acoustic abundance estimation of demersal fish.

ACKNOWLEDGMENTS

The authors wish to thank Dr. Ole Christensen for preprocessing the multibeam sonar data, and we also want to thank to Dr. Ken Foote for valuable input to and help with the manuscript.

Aglen, A. (**1996**). "Impact of fish distribution and species composition on the relationship between acoustic and swept-area estimates of fish density," ICES J. Mar. Sci. **53**, 501–505.

Aglen, A., Engås, A., Huse, I., Michalsen, K., and Stensholt, B. K. (**1999**). "How vertical fish distribution may affect survey results," ICES J. Mar. Sci. **56**, 345–360.

Andersen, L. N. (**2001**). "The new Simrad EK60 scientific echo sounder system," 141st Meeting Acoustical Society of America, Chicago, Illinoise, 4–8 June 2001, J. Acoust. Soc. Am. **109**(5), 2336.

Brede, R., Kristensen, F. H., Solli, H., and Ona, E. (**1990**). "Target tracking with a split-beam echo sounder," Rapp. P.-V. Reun.-Cons. Int. Explor. Mer **189**, 254–263.

Casey, J. M., and Myers, R. A. (**1998**). "Diel variation in trawl catchability: Is it as clear as day and night?" Can. J. Fish. Aquat. Sci. **55**, 2329–2340.

Cushing, D. H. (**1968**). "Direct estimation of a fish population acoustically," J. Fish. Res. Board Can. **25**, 2349–2364.

Cushing, D. H. (**1983**). "The outlook for fisheries research in the next ten years," in *Global Fisheries: Perspectives for the 1980s*, edited by B. J. Rothschild (Springer-Verlag, New York), pp. 263–277.

Ehrenberg, J. E. (**1979**). "A comparative analysis of *in situ* methods for directly measuring the acoustic target strength of fish," IEEE J. Ocean. Eng. **OE-4**, 141–152.

Everson, I., Bravington, M., and Goss, C. (**1996**). "A combined acoustic and trawl survey for efficiently estimating fish abundance," Fish. Res. **26**, 75–91.

Foote, K. G. (**1981**). "Optimizing copper spheres for precision calibration of hydroacoustic equipment," J. Acoust. Soc. Am. **71**, 742–746.

Foote, K. G. (**1982**). "Maintaining precision calibrations with optimal copper spheres," J. Acoust. Soc. Am. **73**, 1054–1063.

Foote, K. G. (**2001**). "Calibrating a narrowband 18-kHz sonar," *Proceedings of the OCEANS 2001 MTS/IEEE Conference and Exhibition*, Vol. **4**, pp. 2503–2505.

Foote, K. G., Chu, D., Hammar, T. R., Baldwin, K. C., Mayer, L. A., Hufnagle, L. C., Jr., and Jech, J. M. (**2005**). "Protocols for calibrating multibeam sonar," J. Acoust. Soc. Am. **117**, 2013–2027.

Foote, K. G., Knudsen, H. P., Vestnes, G., MacLennan, D. N., and Simmonds, E. J. (**1987**). "Calibration of acoustic instruments for fish density estimation: A practical guide," International Council for the Exploration of the Sea Cooperative Research Report No. 144, Copenhagen, Denmark, pp. 1–69.

Godø, O. R., and Wespestad, V. G. (**1993**). "Monitoring changes in abundance of gadoids with varying availability to trawl and acoustic surveys,"

ICES J. Mar. Sci. **50**, 39–51.
Hjellvik, V., Godø, O. R., and Tjøstheim, D. (**2002**). "Diurnal variation in bottom trawl survey catches: Does it pay to adjust?," Can. J. Fish. Aquat. Sci. **59**, 33–48.
Hjellvik, V., Michalsen, K., Aglen, A., and Nakken, O. (**2003**). "An attempt at estimating the effective fishing height of the bottom trawl using acoustic survey recordings," ICES J. Mar. Sci. **60**, 967–979.
Keys, R. G. (**1981**). "Cubic convolution interpolation for digital image processing," IEEE Trans. Acoust., Speech, Signal Process. **29**, 1153–1160.
Kinsler, L. E., Frey, A. R., Coppens, A. B., and Sanders, J. V. (**2000**). "Radiation and reception of acoustic waves," *Fundamentals of Acoustics*, 4th ed. (Wiley, New York).
Kongsberg Maritime. (**2006**). "Neptune B: Post-processing system for bathymetric data," 855-164114/ Rev. B, Product description from Kongsberg Maritime AS, Strandpromenaden 50, P.O. Box 111, N-3191 Horten, Norway.
Krieger, K., Heifetz, J., and Ito, D. (**2001**). "Rockfish assessed acoustically and compared to bottom-trawl catch rates," Alaska Fishery Research Bulletin **8**, 71–77.
Lawson, G. L., and Rose, G. A. (**1999**). "The importance of detectability to acoustic surveys of semi-demersal fish," ICES J. Mar. Sci. **56**, 370–380.
Michalsen, K., Godø, O. R., and Fernö, A. (**1996**). "Diel variation in the catchability of gadoids and its influence on the reliability of abundance indices," ICES J. Mar. Sci. **53**, 389–395.
Ona, E., and Mitson, R. B. (**1996**). "Acoustic sampling and signal processing near the seabed: The deadzone revisited," ICES J. Mar. Sci. **53**, 677–690.
Simmonds, E. J. and MacLennan, D. N. (**2005**). "Observation and measurement of fish," *Fisheries Acoustics*, 2nd ed. (Blackwell, London), pp. 163–215.
Stanley, R. D., Kieser, R., Cooke, K. D., Surry, A. M., and Mose, B. (**2000**). "Estimation of widow rockfish (*Sebastes entomelas*) shoal off British Columbia, Canada as a joint exercise between stock assessment staff and the fishing industry," ICES J. Mar. Sci. **57**, 1035–1049.

Model selection and Bayesian inference for high-resolution seabed reflection inversion

Jan Dettmer[a] and Stan E. Dosso
School of Earth and Ocean Sciences, University of Victoria, Victoria BC V8W 3P6, Canada

Charles W. Holland
Applied Research Laboratory, State College, The Pennsylvania State University, State College, Pennsylvania 16804-0030

(Received 6 August 2008; revised 26 November 2008; accepted 3 December 2008)

This paper applies Bayesian inference, including model selection and posterior parameter inference, to inversion of seabed reflection data to resolve sediment structure at a spatial scale below the pulse length of the acoustic source. A practical approach to model selection is used, employing the Bayesian information criterion to decide on the number of sediment layers needed to sufficiently fit the data while satisfying parsimony to avoid overparametrization. Posterior parameter inference is carried out using an efficient Metropolis–Hastings algorithm for high-dimensional models, and results are presented as marginal-probability depth distributions for sound velocity, density, and attenuation. The approach is applied to plane-wave reflection-coefficient inversion of single-bounce data collected on the Malta Plateau, Mediterranean Sea, which indicate complex fine structure close to the water-sediment interface. This fine structure is resolved in the geoacoustic inversion results in terms of four layers within the upper meter of sediments. The inversion results are in good agreement with parameter estimates from a gravity core taken at the experiment site.
 [DOI: 10.1121/1.3056553]

PACS number(s): 43.30.Pc, 43.30.Ma, 43.60.Pt [AIT] Pages: 706–716

I. INTRODUCTION

Knowledge of seabed sediment geoacoustic properties is important for a variety of acoustic/sonar applications in shallow water environments, and inferring information about sediment properties from acoustic data has received wide attention.[1–16] In recent years, Bayesian inference has been applied increasingly to yield optimal parameter estimates and quantify parameter uncertainties and interrelationships using Markov chain Monte Carlo (MCMC) methods to estimate the posterior probability density (PPD).[8–16] This paper considers Bayesian inference for high-resolution seabed reflection-coefficient data to infer fine-scale local sediment structure with the ultimate goal of quantifying spatial variability in sediments. The data are obtained using an experimental procedure developed by Holland and Osler[6] using a bottom-moored hydrophone and a ship-towed impulsive source (e.g., seismic boomer). Due to the local scale (~100 m seabed footprint), the effects of spatial and temporal variabilities in the water column and seabed are greatly reduced compared to long-range acoustic measurements. Further, impulsive sources with a large bandwidth to high frequencies have potential to resolve fine structure in the uppermost sediment. The inversion is carried out using a plane-wave reflection-coefficient forward model.[10]

Seismoacoustic reflection data can be inverted in the time and/or frequency domains to yield seabed sound-velocity, density, and attenuation profiles.[9–11] Dettmer *et al.*[11] selected the model parametrization by estimating the number of sediment layers as the number of distinct reflected arrivals in a sequential Bayesian inversion of time- and frequency-domain data. However, high-quality single-bounce reflection-coefficient data can contain information about sediment layers with thicknesses much less than the typical boomer pulse length (~0.5 m at 1500 m/s) in the experiment. In particular, high-quality broadband data that extend to high frequencies (several kilohertz) can contain information about layers of order of centimeters. In such cases, time-domain data are not sufficient to address model selection (e.g., model parametrization) because distinct reflected arrivals for each layer cannot be identified.

Model selection is a common aspect of geoscientific inverse problems, where complex unknown environments often result in nonuniqueness and unknown theory error. A model is considered to be any particular choice of physical theory, its appropriate parametrization, and a statistical representation for the data errors that are used to explain the observed physical system. The goal of the selection is to determine the simplest model that sufficiently explains the data by applying a parsimony criterion. In this paper, parametrization of the model is considered in terms of the number of layers. Model comparison is also a central problem for studying spatial sediment variability. Meaningful comparison of sediment structure between sites requires a scheme to quantify uncertainty at each site, including uncertainty due to the model parametrization selected. Model selection is particularly important for high-quality data, since poor model selection will result in systematic errors, potentially causing biased results. It is important to note that while underparametrizing a model can be desirable in terms of simplicity, it

[a]Electronic mail: jand@uvic.ca

can also lead to unrealistically small parameter uncertainty estimates and large theory error. A simple model will reach high likelihood levels only for specific parameter values; it can only access a limited part of the data space and therefore will indicate small uncertainties for the predicted parameters, which can be misleading. Furthermore, the theory error introduced by underparametrization can lead to biases in the parameter estimates. Hence, it is important to select an appropriate parametrization based on an objective criterion. Which and how many models are considered in a study depend largely on subjective choices such as the intended use of the model and prior knowledge of the environment.

Several approaches exist to examine models by quantifying their likelihood using point estimates [e.g., evaluated at the maximum *a posteriori* (MAP) model vector].[17,18] For example, assuming Gaussian-distributed data errors, χ^2 probabilities can be used to evaluate the likelihood of a model at single points in the model spaces considered. An F-test can be used to calculate the allowable change in the χ^2 misfit for given levels of significance.[19] The main problem with the maximum-likelihood method is that it is biased toward unjustifiably complex models,[18] which is in conflict with the generally accepted concept of *Ockham's razor*[20,21] that empirically demands preference of simple models. To avoid this bias, asymptotic methods such as the Akaike information criterion (AIC) (Ref. 22) have been introduced. However, the AIC is still biased toward complex models for large data sets.[23] The Bayesian information criterion (BIC)[23,24] is related to Bayesian factors and eliminates the bias by accounting for the number of data. The BIC is used for model selection in this paper.

The class of models considered here are multilayered sediment models defined by layer thickness, sound velocity, density, and attenuation, and the model selection problem is to determine the optimum number of sediment layers required to sufficiently match the observed data while satisfying parsimony. High-dimensional model spaces (up to 35 parameters) and strong parameter correlations are addressed by an efficient MCMC sampling scheme using the Metropolis–Hastings (MH) algorithm in principal-component space with a Cauchy proposal distribution.[25] The scaling of the Cauchy distribution is initially based on a linear approximation of the PPD around a likely model, and progressively transformed into a nonlinear estimate during the burn-in phase.[25] A massively parallel algorithm implementation is developed to allow feasible application to complicated and computationally demanding problems.[10]

To summarize, this paper illustrates how layered profiles can be resolved from seabed reflection data collected with an acoustic-source pulse length that is large compared to the layering structure. The inherent nonuniqueness of the problem is practicably addressed by applying the BIC to ensure parsimony of inversion results.

The remainder of the paper is structured as follows. In Sec. II, Bayesian inference is reviewed with an emphasis on model comparison. The Bayesian inference is then applied to data collected on the Malta Plateau, Mediterranean Sea, in Sec. III. In Sec. III C the model selection is carried out applying the BIC to the MAP model parameters from the inversion. Finally, Sec. III D presents the results in terms of geoacoustic profiles for the selected model, and compares the results to cores taken at the experiment site.

II. BAYESIAN INFERENCE

This section gives a brief overview of the Bayesian formulation of inverse problems; more complete treatments can be found elsewhere.[21,26–30] Let $\mathbf{d} \in \mathbb{R}^N$ be a random variable of N observed data containing information about a physical system. Further, let $\mathcal{I}$ denote the model specifying a particular choice of physical theory, model parametrization, and error statistics to explain that physical system. Let $\mathbf{m} \in \mathbb{R}^M$ be a random variable with M free parameters representing one realization of the model $\mathcal{I}$. Bayes' rule can then be written as

$$P(\mathbf{m}|\mathbf{d},\mathcal{I}) = \frac{P(\mathbf{d}|\mathbf{m},\mathcal{I})P(\mathbf{m}|\mathcal{I})}{P(\mathbf{d}|\mathcal{I})}, \tag{1}$$

where the conditional probability $P(\mathbf{m}|\mathbf{d},\mathcal{I})$ represents the PPD of the unknown model parameters given the observed data, prior information, and choice of model $\mathcal{I}$. The conditional probability $P(\mathbf{d}|\mathbf{m},\mathcal{I})$ describes the data-error statistics. Since data errors include measurement and theory errors (which cannot generally be separated), the specific form of this distribution is often not known. To interpret Eq. (1) quantitatively, some particular form that describes the data-error statistics reasonably well must be assumed for this distribution. In practice, mathematically simple distributions, such as multivariate Gaussian distributions, are commonly used; the validity of such assumptions should be checked using statistical tests.[11,12] (Other distributions can be used as long as an appropriate likelihood function can be formulated; e.g., double exponential distributions are commonly applied for data sets that include large outliers.[30]) The general multivariate Gaussian distribution for real data is given by

$$P(\mathbf{d}|\mathbf{m},\mathcal{I}) = \frac{1}{(2\pi)^{N/2}|\mathbf{C}_d|^{1/2}} \times \exp\left(-\frac{1}{2}(\mathbf{d}-\mathbf{d}(\mathbf{m}))^{\mathrm{T}}\mathbf{C}_d^{-1}(\mathbf{d}-\mathbf{d}(\mathbf{m}))\right), \tag{2}$$

where $\mathbf{C}_d$ is the data covariance matrix and $\mathbf{d}(\mathbf{m})$ is the modeled data. The covariance matrix $\mathbf{C}_d$ is often unknown since the source of errors may be poorly understood. In some cases, data-error statistics can be parametrized (e.g., as variances or as a covariance matrix based on an assumed form such as an autoregressive moving average[19]) and included in the inversion, either implicitly[31] or explicitly as unknown hyperparameters with assigned priors.[31–33] Data-error covariance matrices can also be estimated nonparametrically from data residuals (i.e., the difference between modeled and measured data, considered later). In inverse theory, $P(\mathbf{d}|\mathbf{m},\mathcal{I})$ is interpreted as a likelihood function $\mathcal{L}(\mathbf{m}|\mathcal{I})$ of $\mathbf{m}$ for fixed (observed) data $\mathbf{d}$. Note that given a Gaussian data-error distribution, the likelihood function is not Gaussian distributed for nonlinear inverse problems. The term $P(\mathbf{m}|\mathcal{I})$ in Eq. (1) gives the model prior distribution. In this paper, prior distributions are considered to be bounded, uniform distributions of the form

$$P(\mathbf{m}|\mathcal{I}) = \begin{cases} \prod_{i=1}^{M(\mathcal{I})} (m_i^+ - m_i^-)^{-1} & \text{if } m_i^- \le m_i \le m_i^+, \quad i = 1, M(\mathcal{I}) \\ 0 & \text{else.} \end{cases} \tag{3}$$

The conditional probability $P(\mathbf{d}|\mathcal{I})$ is commonly referred to as the *evidence* or *marginal likelihood* of $\mathcal{I}$. It describes how likely a certain parametrization $\mathcal{I}$ is given the observed data and prior. Since the evidence $P(\mathbf{d}|\mathcal{I})$ normalizes Eq. (1), it can be written as

$$\mathcal{Z}(\mathcal{I}) = P(\mathbf{d}|\mathcal{I}) = \int_{\mathcal{M}} P(\mathbf{d}|\mathbf{m},\mathcal{I}) P(\mathbf{m}|\mathcal{I}) d\mathbf{m}. \tag{4}$$

A. Estimating the PPD

To estimate the PPD for a fixed choice of model $\mathcal{I}$, MCMC sampling methods are usually applied.[21,28,29,34,35] The PPD can then be used to obtain model parameter and uncertainty estimates. For a fixed parametrization, Eq. (1) can be written as

$$P(\mathbf{m}|\mathbf{d},\mathcal{I}) \propto \mathcal{L}(\mathbf{m}|\mathcal{I}) P(\mathbf{m}|\mathcal{I}), \tag{5}$$

where $P(\mathbf{m}|\mathbf{d},\mathcal{I})$ quantifies the state of information about the model parameters given the data, prior information, and parametrization. This paper considers the likelihood function to be of the form given by Eq. (2). For inference it is common to work with the log-likelihood to avoid floating-point underflow due to the exponential dependence. The multidimensional PPD is generally interpreted in terms of properties such as the MAP model vector estimate $\hat{\mathbf{m}}$, the *a posteriori* mean model vector $\bar{\mathbf{m}}$, the model covariance matrix $\mathbf{C}_m$, and marginal-probability distributions $P(m_i|\mathbf{d})$, defined as

$$\hat{\mathbf{m}} = \arg_{\max} P(\mathbf{m}|\mathbf{d},\mathcal{I}), \tag{6}$$

$$\bar{\mathbf{m}} = \int_{\mathcal{M}} \mathbf{m}' P(\mathbf{m}'|\mathbf{d},\mathcal{I}) d\mathbf{m}', \tag{7}$$

$$\mathbf{C}_m = \int_{\mathcal{M}} (\mathbf{m}' - \bar{\mathbf{m}})(\mathbf{m}' - \bar{\mathbf{m}})^T P(\mathbf{m}'|\mathbf{d},\mathcal{I}) d\mathbf{m}', \tag{8}$$

$$P(m_i|\mathbf{d}) = \int_{\mathcal{M}} \delta(m_i' - m_i) P(\mathbf{m}'|\mathbf{d},\mathcal{I}) d\mathbf{m}', \tag{9}$$

where δ denotes the Dirac delta function. Higher-dimensional marginal distributions can be defined similar to Eq. (9). While analytic solutions to Eqs. (6)–(9) exist for linear inverse problems, nonlinear problems such as geoacoustic inversion must be solved numerically.

MAP estimates [Eq. (6)] can be found by numerical optimization methods, such as adaptive simplex simulated annealing (ASSA), which combines the local downhill simplex method within a very fast simulated annealing global search.[5] For the inversion considered in this paper, the optimization problem is particularly challenging with large numbers of parameters, and a parallel implementation of the ASSA optimization was developed employing message passing.[36] Parallel ASSA uses as many simplexes as the number of available central processing units (CPUs). On each CPU, a single simplex is optimized for a certain number of steps using the usual scheme.[5] The models of all simplexes are then randomly regrouped into new simplexes and ASSA optimization is again performed for a certain number of steps before the regrouping process is carried out again, and so on until convergence. This optimization searches the parameter space more thoroughly than an algorithm with one simplex, and makes optimization of complicated problems feasible. Detailed performance studies to examine the scaling of the parallel ASSA algorithm with the number of CPUs were not carried out.

The integrals of Eqs. (8) and (9) are computed here using MCMC sampling that applies an adaptive MH algorithm.[34,35] The implementation follows Dosso and Wilmut.[25] Initially, the algorithm computes a linearized PPD approximation around a reasonably good starting model. This yields an approximate linear model covariance matrix that can be used for a principal-component rotation of the parameter space where the coordinate axes align with the dominant correlation directions around the starting model. During a burn-in phase, the linear model covariance matrix is progressively replaced with a fully nonlinear covariance estimate. To further ensure efficient sampling of the posterior, perturbations in the Markov chain are drawn from a Cauchy proposal distribution that is scaled by the eigenvalues of the covariance matrix, which represent variances of the principal components. During sampling, chain thinning[26] is applied to keep reasonably small sample sizes for models with large numbers of parameters. The algorithm is massively parallel for feasibility.[11]

B. Model selection

Bayesian evidence (also referred to as the normalizing constant or free energy), Eq. (4), and Bayesian factors (the ratio of evidences for two models) are the basis for model selection. Evidence brings a natural parsimony to the model selection problem, which is also referred to as the Bayesian razor. In contrast to Ockham's razor, Bayesian model selection is not based on a qualitative preference for model parsimony based on aesthetic or empirical reasons, but rather favors parsimony intrinsically and quantitatively.[21] Estimating evidence is challenging due to the requirement to integrate the likelihood with respect to the prior,[37] and finding robust and accurate estimators for the evidence integral has seen the attention of much research. Note that evidence estimates should ideally be *future proof*, allowing future researchers to compare results obtained for the same data by different techniques.[38]

Several methods to approximate evidence by predictive distributions and by analytical examination of asymptotic behavior can be found in literature.[17,18] Further, numerous attempts in the statistics community have been made to find MCMC sampling estimators that are based on the posterior.[18,37,39–41] Several estimators (dependent on the in-

verse likelihood) have been shown to be unstable.[37] Others, such as importance sampling based on the posterior,[18] have been applied[40] including an application to geoacoustic inversion.[15] However, significant problems exist with this approach since evidence depends on the likelihood's relationship to the prior.[42,43] This can lead to unstable and inaccurate results due to only sampling from the posterior. Results are also sensitive to the choice of importance sampling function.[28,37,42,43]

Evidence can also be addressed by treating the inversion as a transdimensional problem and applying reversible-jump Markov chains[44] that can jump between spaces of different dimensionalities. Although interesting for future applications, transdimensional inversion is challenging and implementation for general problems is difficult.[34,45] Reversible-jump Markov chains also require specifying all models to be considered *a priori*. Other Monte Carlo based methods that also/only sample from the prior such as thermodynamic integration,[40,46] annealed importance sampling,[47] and nested sampling[38] exist that can give unbiased estimates of the evidence for general problems. However, the associated computational cost of these methods is high.

Due to the high computational demands of the forward and inverse problems considered in this paper, an asymptotic point estimate (for the maximum-likelihood model vector $\mathbf{m}^{\mathrm{ML}}$) is used to carry out model selection. The BIC[23,24] is an asymptotic approximation derived for diffuse multivariate normal prior distributions (with mean at the maximum-likelihood estimate and variance of the Kullback–Leibler expected information).[48] The BIC is given by

$$\mathrm{BIC}(\mathcal{I}) = -2\,\log_e \mathcal{L}(\mathbf{m}^{\mathrm{ML}}|\mathcal{I}) + M\,\log_e N. \tag{10}$$

Since the BIC is based on the negative log likelihood, the model with the smallest BIC is selected as the preferred model. The value of the BIC cannot be directly associated with a probability and cannot yield the significance of the selection. In comparison to the also commonly used AIC,[22]

$$\mathrm{AIC}(\mathcal{I}) = -2\,\log_e \mathcal{L}(\mathbf{m}^{\mathrm{ML}}|\mathcal{I}) + 2M, \tag{11}$$

the BIC corrects for the number of data and favors simpler models than the AIC for $N > 8$.

The AIC has been shown to yield excessively complex models, particularly for large numbers of data.[23] Since this study addresses large data sets, the BIC is used here for model selection.

III. INVERSION RESULTS

A. Experiment and data

This section applies the inversion and model selection to seismoacoustic reflection data collected April 6, 2002, during the Boundary02 experiment at 36° 24.515′ N, 14° 38.142′ E on the Malta Plateau, Mediterranean Sea. The acoustic data were generated with an electromechanical impulsive source (GeoAcoustics 5813B Geopulse boomer) with a short pulse length (<1 ms) and a broad bandwidth (0.5–10 kHz). Data were recorded at a single receiver that was part of a vertical line array of four hydrophones, and sampled at 48 kHz. The hydrophone used in this data set was at 124 m depth and the water depth was 144 m. The sound-velocity profile is shown in Fig. 1 and was fairly constant with the sound velocity varying less than 5 m/s over the water column. The source was towed at 0.3 m depth.

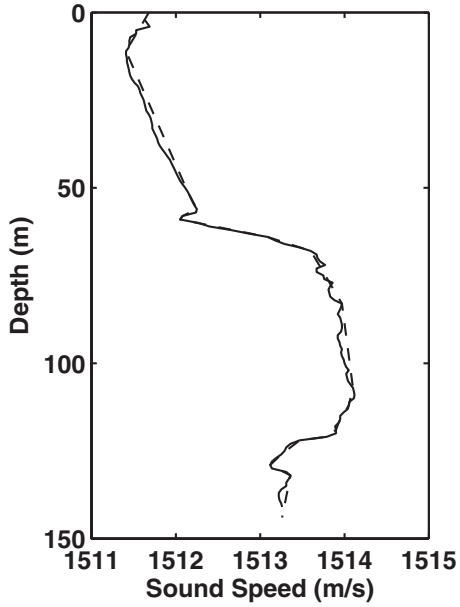


FIG. 1. Original conductivity temperature depth profile (solid line) and approximated sound-velocity profile (dashed line, used for the inversion) collected at site 13 on the Malta Plateau.

Figure 2 shows part of the seismoacoustic traces (in reduced time) with the lines across traces indicating the direct arrival and the part of the bottom response used to compute the reflection coefficients. Figure 2 also illustrates the need for an objective model selection criterion. Reflected energy is most concentrated around the water-sediment interface reflection at 0.109 s and at a later event at about 0.114 s (times are given here for the shortest-range trace). In both instances, the events are spread out in time and the model parametriza-

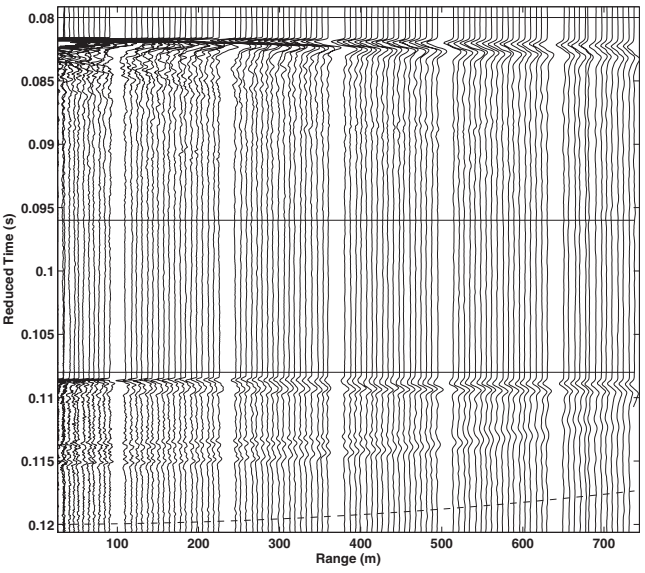


FIG. 2. Seismoacoustic traces (in reduced time, with 1512 m/s reducing velocity) collected at site 13 on the Malta Plateau.

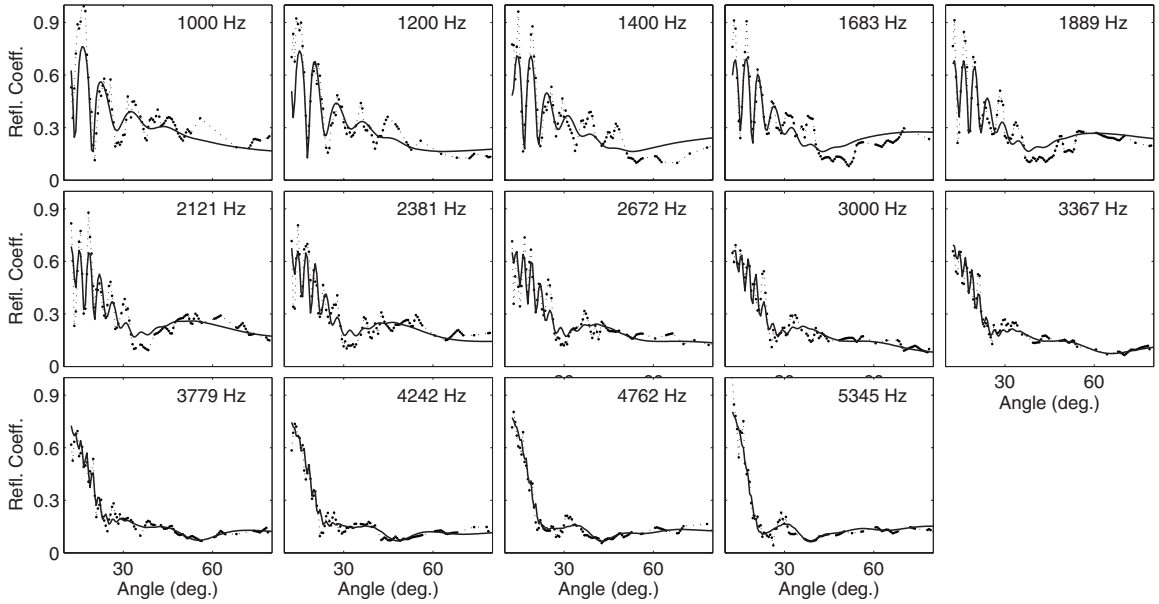


FIG. 3. Reflection-coefficient data as a function of grazing angle and frequency (band centers indicated). The solid line indicates the best fit obtained from the MAP parameters of the model selected by the BIC (five layers).

tion is not obvious: Both zones could contain reflections from one or more layers as individual reflectors cannot be clearly identified.

Reflection-coefficient data as a function of grazing angle and frequency were computed from time-windowed direct and bottom-reflected arrivals using the method of Holland,[7] and are shown in Fig. 3. In this case, the bottom response is time windowed to approximately 6 m depth below the seafloor for ranges between 25 and 740 m, as indicated in Fig. 2. The data are averaged into 14 frequency bands from 1000 to 5300 Hz using a Gaussian frequency average[10,49] with a fractional bandwidth of 1/10 of the center frequency, resulting in bandwidths from 100 to 530 Hz. Figure 4 shows reflection-coefficient data that are averaged over 5 Hz bands and compares these to the reflection-coefficient data that are used in the inversion. The fractional bandwidth of 1/10 was found to retain structure in the reflection-coefficient data while reducing noise and resulting in reflection-coefficient data that are computationally feasible in the inversion. The data are interpolated onto a uniform spacing in angle; points with a signal to noise ratio of less than 6 dB were excluded. Further, interpolated data that fall into recording gaps (due to experiment design) are excluded from the inversion. This results in approximately 90 data at each frequency with an angular range from 12° to 81°.

B. Forward model

The forward model consists of a plane-wave reflection-coefficient model that approximates the seabed as a layered lossy fluid.[11] The replica reflection-coefficient data for each frequency band are computed using the same frequency av-

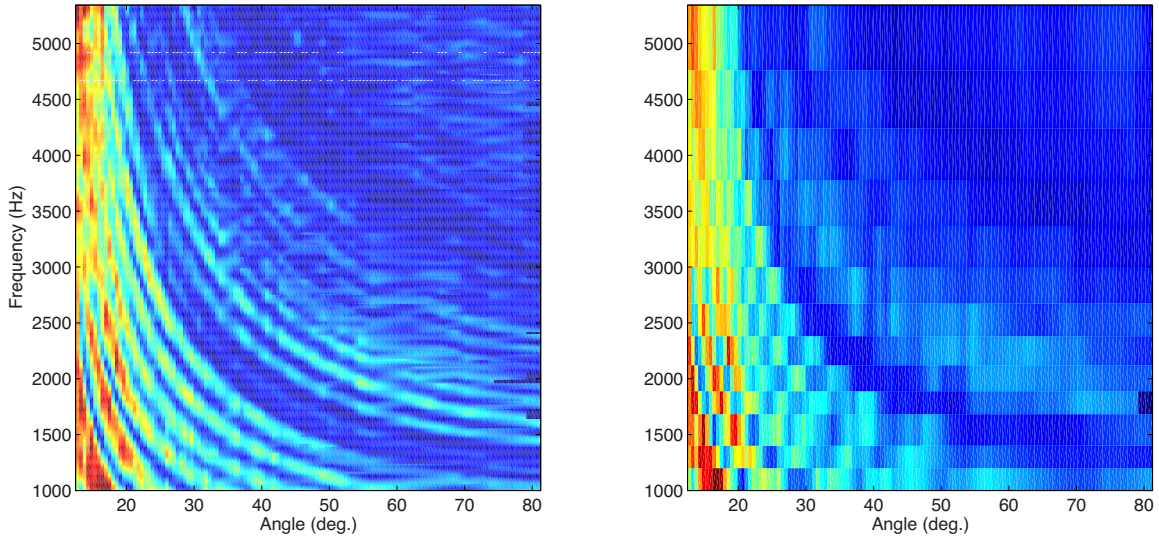


FIG. 4. (Color online) Reflection-coefficient data (clipped at 1.0) as a function of grazing angle and frequency. The left panel shows reflection-coefficient data for a narrow, constant 5 Hz band average and the right panel shows the data used in the inversion, averaged with a fractional bandwidth of 1/10.

eraging as used for the measured data. However, to address limited computational time, the number of frequencies in the average is limited to 12 frequencies per band. A forward modeling study was carried out to ensure that full-wave field effects are negligible and that plane-wave modeling is sufficient.[10]

C. Model selection

The model selection study was carried out using two groups of models. Group A assumes an increasing number of layers for the uppermost part of the sediment (corresponding to the reflected arrivals beginning at 0.109 s in Fig. 2) and a single reflector at depth (corresponding to the deeper reflected arrivals at 0.114 s). Group B considers the same numbers of layers for the uppermost part of the sediment but contains two reflectors (i.e., a layer) at depth. This way both zones that show reflections in the time series are addressed by the model selection. Group A includes five models with three to seven layers and group B includes seven models with two to eight layers. For each model the total number of parameters in the inversion increases by 4 with each additional layer.

To obtain inversion and model selection results, data-error statistics were estimated from the reflection-coefficient processing[7] and optimization was carried out using plane-wave reflection-coefficient inversion.[12,14] Prior bounds were chosen to be uniform over wide intervals. However, the thin topmost layers were differentiated by choosing prior bounds for layer thickness from 5 to 50 cm. This seems justified since the uppermost events in the time series (Fig. 2) extend over about 1.5 m (at 1500 m/s). These priors also allow the more complex models to exactly represent the simpler models. The resulting likelihood values were used to calculate the BIC for all models and results are shown in Fig. 5. All values are plotted on a $\log_e$ scale since the range of values is large (due to the fact that the two-and three-layer models are very

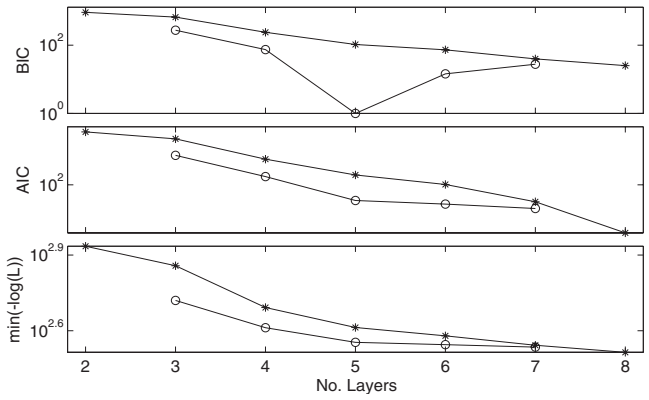


FIG. 5. BIC, AIC, and negative log likelihood for group A (open circles) and group B (asterisks) on a $\log_e$ scale. Note that for presentation purposes BIC and AIC values have been shifted so that the minimum value of each is unity.

unlikely with high BIC and AIC values). This figure shows that, based on the minimum value of the BIC, the five-layer model from group A is selected. The BIC values for the models in group B do not reach values as low as those for group A, but consistently decrease with more complex parametrizations. This result indicates that the data show high sensitivity to the presence of several layers close to the sediment-water interface. In addition, the simpler models of group A are preferred over the models of group B which contain more structure at depth. Figure 5 also shows the negative log likelihood (i.e., the data misfit) for all parametrizations. It is important to note that these values consistently decrease with an increasing number of layers. The BIC depends strongly on the likelihood values, and a better fit to the data for more complex models is important for the BIC to yield meaningful results. For comparison, the figure also shows the AIC values, with the minimum for both groups occurring at the models with most layers. This bias toward too-complex models for large data sets is commonly observed in other problems.[23]

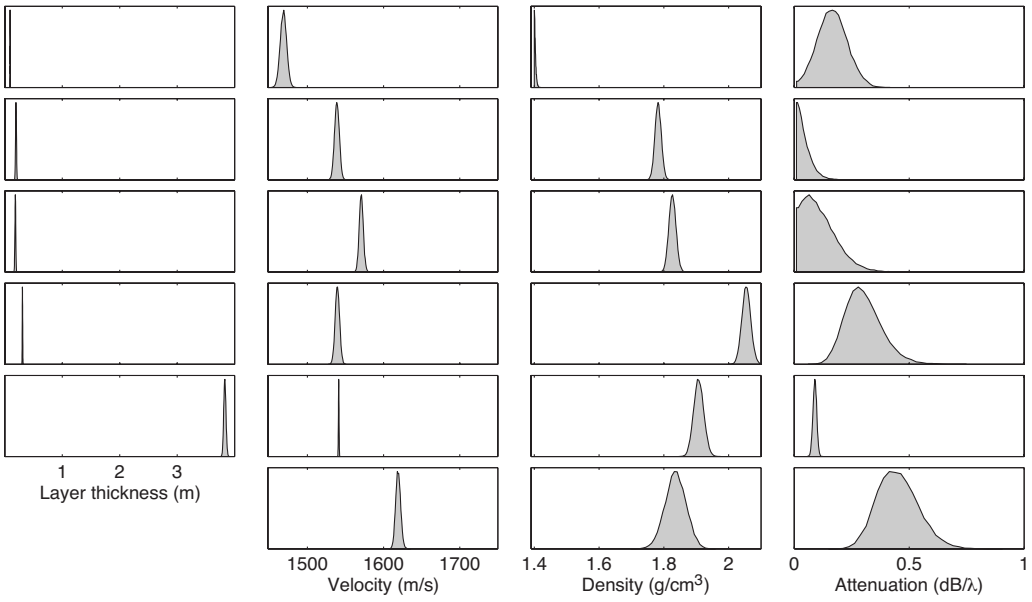


FIG. 6. Marginal-probability distributions for the five-layer model from group A. Note that for presentation purposes all marginals are scaled to the same height, not to unit area. Units for attenuation are given in terms of decibels per wavelength (λ).

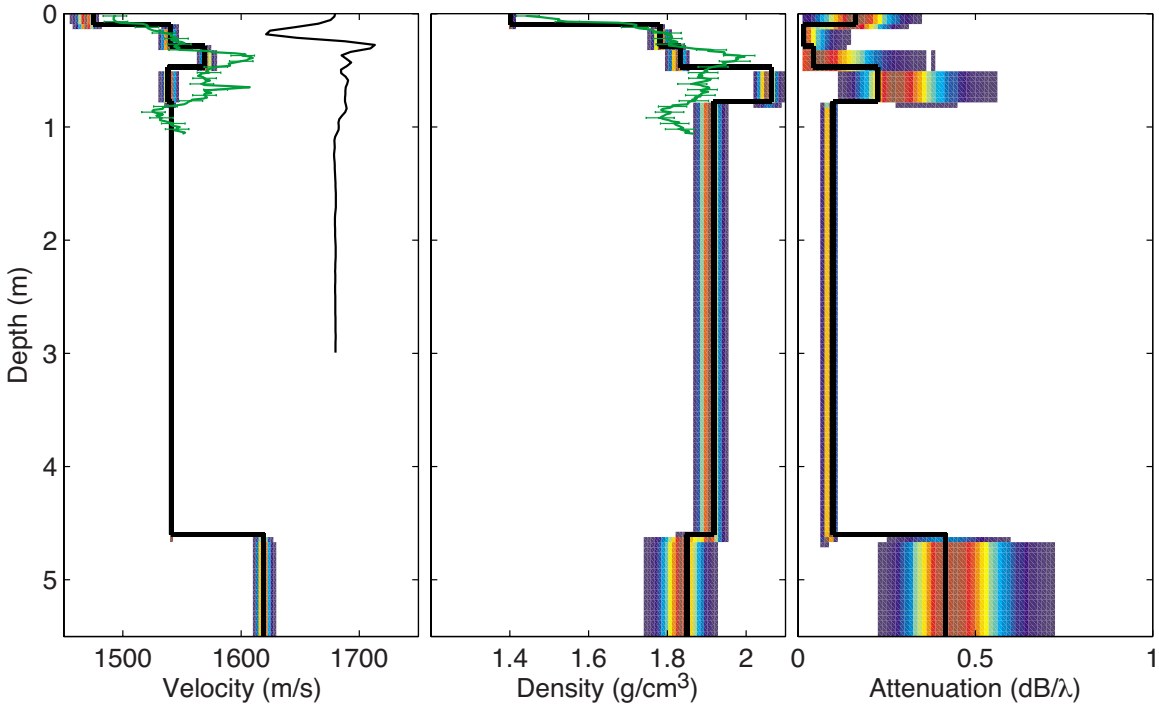


FIG. 7. (Color online) Marginal-probability depth distributions and MAP sediment profiles (solid line) for the five-layer model of group A (selected by BIC). A core (solid line with error bars) taken on site is shown for comparison, with error bars shown for every fifth datum. The left panel also shows the acoustic-source pulse (at 1500 m/s).

D. Posterior parameter inference

Once the preferred model parametrization was identified using the BIC, the data residuals for this model were used to compute a nonparametric estimate of the data-error covariance matrix at each frequency.[11] Posterior statistical tests were carried out for raw residuals $\mathbf{d}\text{-}\mathbf{d}(\hat{\mathbf{m}})$ and for standardized residuals $\mathbf{C}_d^{-1/2}[\mathbf{d}\text{-}\mathbf{d}(\hat{\mathbf{m}})]$ to ensure that these estimates and the assumptions of random Gaussian errors were reasonable. The run test for randomness failed at all 14 frequencies (at the 0.05 level) for raw residuals; however, standardized residuals passed the test at 13 out of 14 frequencies. The Kolmogorov–Smirnov test for Gaussianity was passed (at the 0.05 level) four times for the raw residuals and nine times for the standardized residuals. Overall, the results of the statistical tests suggest the data-error statistics are reasonably well quantified, providing confidence in the inversion results. The estimated error covariance matrices were then used in the integration of the PPD by MH sampling.

The integration was carried out for three models from group A, the five-layer model that was picked due to the BIC and the models with three and seven layers to observe the variability in results with model selection. The fit to the measured data for the five-layer model is shown in Fig. 3, and marginal distributions are shown in Fig. 6. The marginal distributions indicate generally well-resolved parameters with distinct changes in sound velocity and density between layers. Attenuation shows low resolution in thin layers but high resolution within the fourth layer of about 4 m thickness.

Figure 7 shows the MAP sediment profiles and associated uncertainties in terms of marginal-probability depth distributions for the five-layer model. The uncertainties are obtained from a large random subset of the PPD (4×10^5 models). The inversion results indicate four layers within the upper meter of the sediments. Figure 7 also shows the acoustic-source pulse (at 1500 m/s) for comparison with the inversion results: Note that the pulse length is large compared to the layered structure resolved in the upper sediments. The appearance and location of the thick layer are consistent with what would be expected from the time-domain data (Fig. 2) where no significant reflections occur between 0.11 and 0.113 s. The inversion result also shows an interface at about 4.5 m depth. Confidence for the half-space parameter values, particularly density, is low, likely because the half-space lacks a lower reflector and hence data information is available only from the reflection off the upper interface.

Figure 7 also shows the sound-velocity and density estimates from a shallow gravity core taken at the site. The core error bars, shown for every fifth datum, represent measurement errors associated with the time-of-flight and gamma-ray attenuation density estimates for a perfectly calibrated system, but do not include errors due to sediment disturbance from sampling, retrieval, and storage. Note that the core represents a highly localized sample (10 cm diameter compared to the ~100 m experiment footprint) and that spikes in the core values could be due to anomalies such as seashells. The core indicates a complicated structure in the upper part of the sediment. The inversion results show a similar fine structure to the core. In particular, the density profile estimate from the inversion matches the core profile estimate well: The location of interfaces in the inversion result coincides with interfaces in the core. Below about 0.5 m depth, the inversion results appear to represent an average value in the sound velocity estimated by the core.

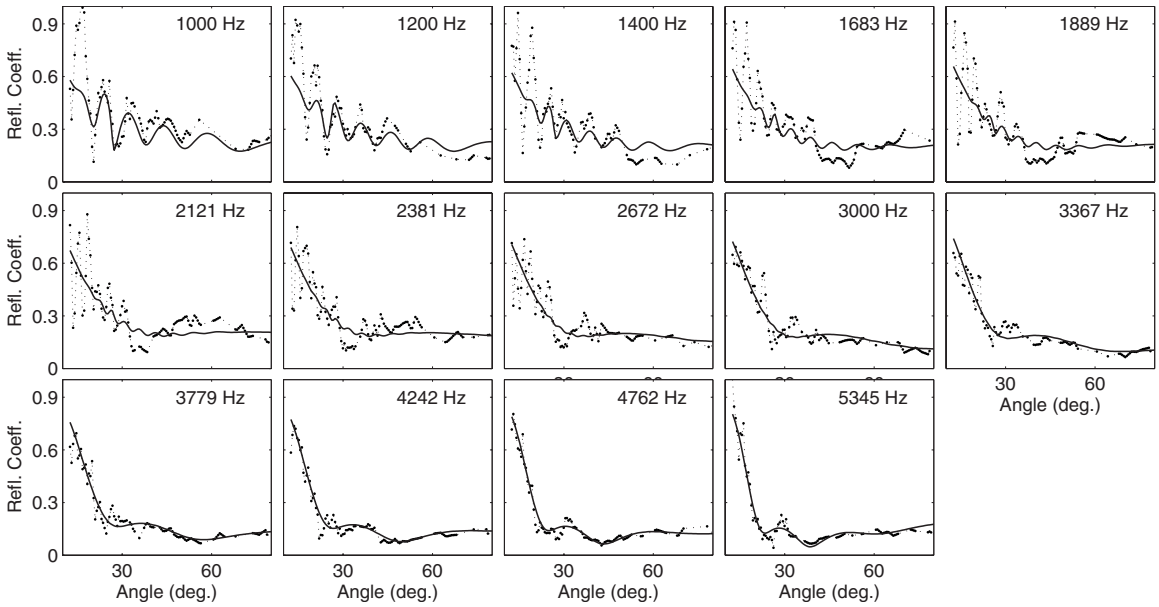


FIG. 8. The best fit obtained from the MAP parameters of the three-layer model of group A.

Figures 8 and 9 show inversion results for the three-layer model from group A. The fit to the data (Fig. 8) is considerably worse than that for the five-layer model (Fig. 3), and the resulting low log-likelihood value dominates the BIC, resulting in the rejection of the three-layer model. However, high frequencies show a much better fit than low frequencies, suggesting that the model is more appropriate for the shallowest structure than for the deep structure. This is also evident in the marginal distributions in Fig. 9, which show that the first layer of about 10 cm thickness is resolved. However, the sound velocity for the uppermost layer is likely biased and appears high compared to the water sound velocity of 1513 m/s (see Fig. 1). The reason for this effect in sound velocity is likely the underparametrization of the three-layer model, which causes the surficial layer to account for not only the uppermost sound velocity but also somewhat the deeper sound-velocity structure. Below that, the sound-velocity results seem to approximately average over the core structure shown in Fig. 7 in a similar fashion. The density inversion results appear to be consistently too low compared to the five-layer model and the core.

Figures 10 and 11 show inversion results for the seven-layer model from group A. The fit to the data (Fig. 10) is slightly improved compared to Fig. 3, but according to the

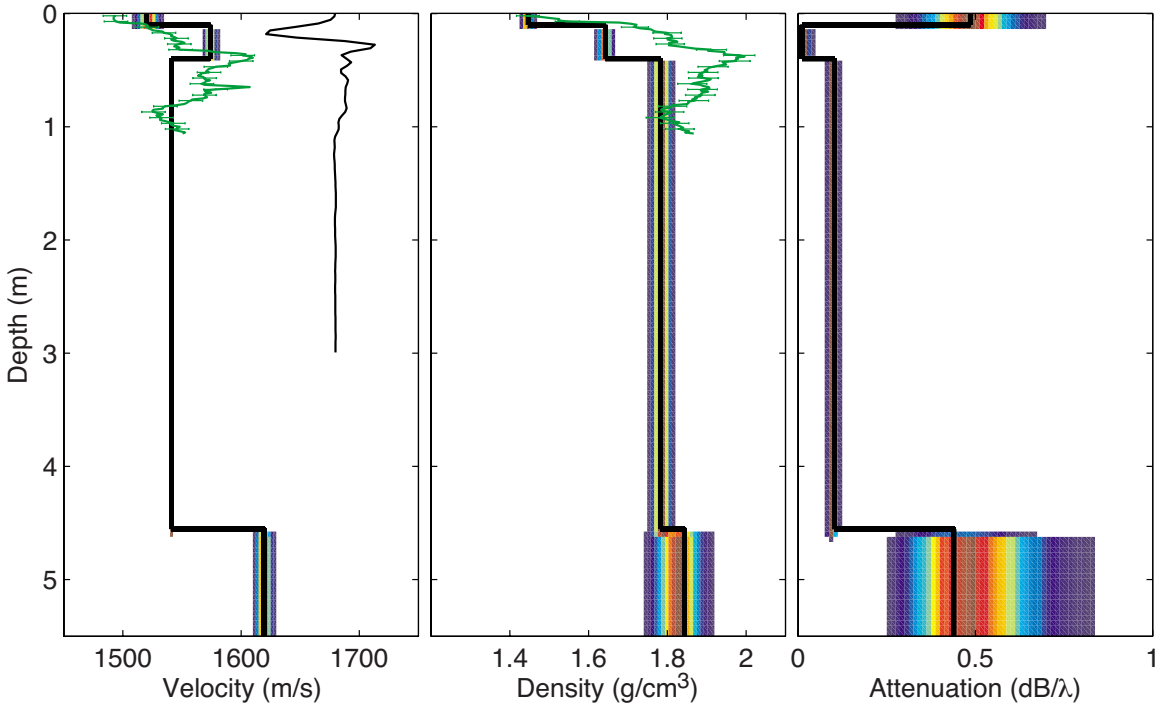


FIG. 9. (Color online) Marginal-probability depth distributions and MAP sediment profiles (solid line) for the three-layer model of group A. A core (solid line with error bars) taken on site is shown for comparison. Core error bars are shown for every fifth datum on the core.

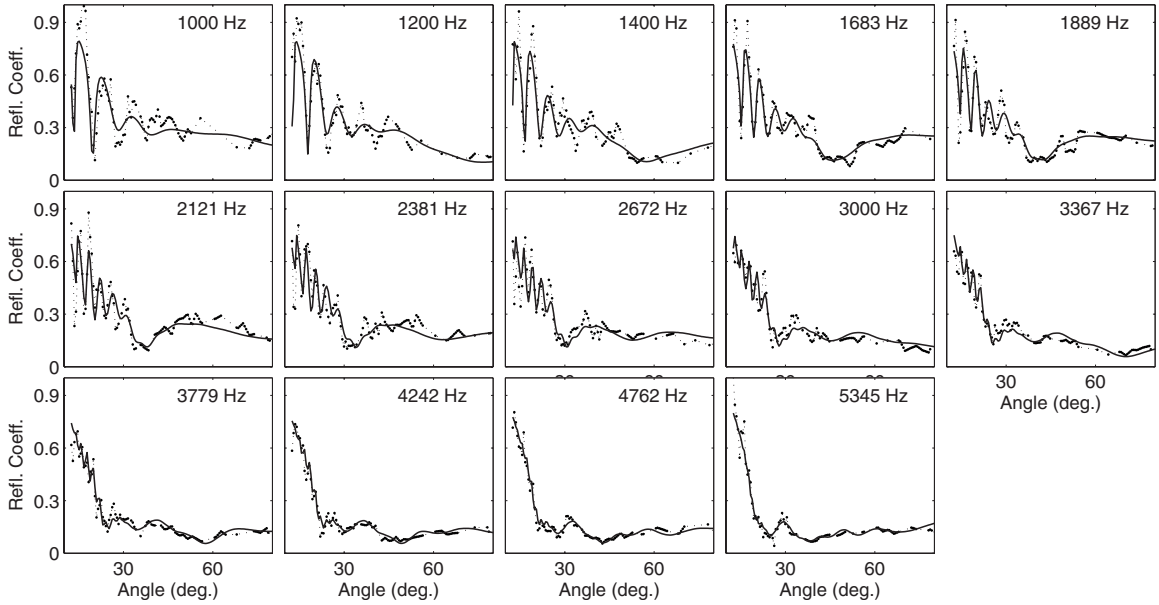


FIG. 10. The best fit obtained from the MAP parameters of the seven-layer model of group A.

BIC, this improvement does not justify the additional layers. The model shows a similar structure throughout the first part of the seabed when compared to Fig. 7, but two more layers were found just above the thickest layer. In general, parameter uncertainties are larger for this model, as expected, because the more complex model allows the inversion to access more of the data space. Comparing this result to the core indicates good agreement of the two estimates. In particular, a feature in the core at about 1 m depth that was not included in the five-layer model appears in the seven-layer model. Between 0.5 m and 1 m depths, the sound velocity of the third layer appears to be lower than the core estimate. Both the five- and seven-layer models are fairly similar and indicate similar features. The model selection based on the BIC selected the simpler model of the two in compliance with Ockham's razor.

IV. SUMMARY

This paper illustrates a practical approach to Bayesian model selection for geoacoustic inversion where the acoustic pulse length is large compared to the layered sediment structure. Model selection is particularly important for geoacoustic inversion when the information content of the data is high

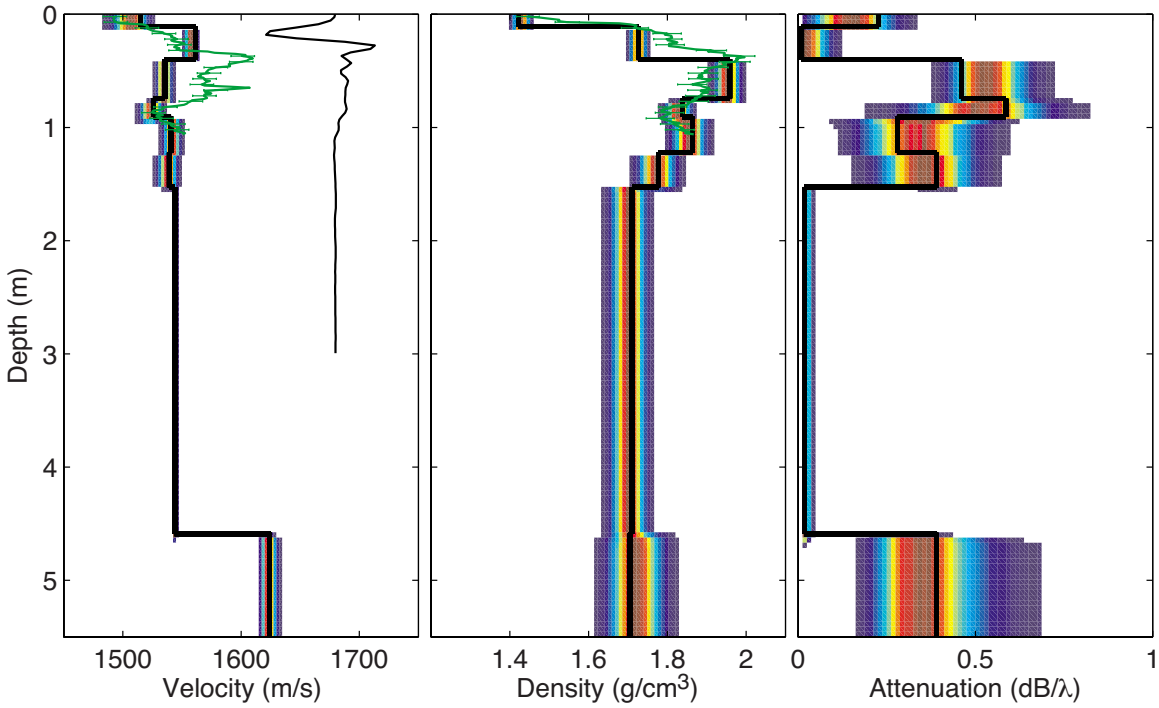


FIG. 11. (Color online) Marginal-probability depth distributions and MAP sediment profiles (solid line) for the seven-layer model of group A. A core (solid line with error bars) taken on site is shown for comparison. Core error bars are shown for every fifth datum on the core.

and nonuniqueness causes difficulty in selecting appropriate models. Further, quantitative model selection is crucial to quantify spatial variability of geoacoustic properties; only quantitative selection of appropriate model parametrizations allows for quantitative comparison between different experiment sites.

The practical approach consists of applying the BIC to MAP parameter estimates, focusing on selecting the most likely number of sediment layers required to sufficiently fit the data. The BIC provides an asymptotic approximation to Bayesian evidence and therefore introduces a parsimony criterion that avoids overparametrization of the model. At the same time, excessive underparametrization is avoided, which is important for reasonable geoacoustic uncertainty estimates, since too few parameters cause excessively small uncertainties.

Once model selection is completed, posterior parameter inference is carried out by integrating the PPD for the most likely model, employing a MH algorithm. Data errors are estimated from the data residuals in terms of a nonparametric data covariance matrix. The validity of assuming Gaussian data errors was examined *a posteriori*.

The approach is successfully applied to data collected on the Malta Plateau. The seismoacoustic time series show complicated arrivals since the acoustic-source pulse length is large compared to the layering structure. As a result, the appropriate amount of structure needed to parametrize the sediment model is not obvious. The Bayesian inference results (including model selection and posterior parameter estimates) illustrate high resolution, well below the length of the source pulse. The results also show generally good agreement with the sound-velocity and density estimates from a gravity core taken at the same site.

ACKNOWLEDGMENTS

The authors gratefully acknowledge the support of the Office of Naval Research postdoctoral fellowship (Grant No. N000140710540) and the Ocean Acoustics Program (ONR OA Code 321). The data were collected under the Boundary Characterization Joint Research Project including the NATO Undersea Research Centre (NURC), Pennsylvania State University—ARL-PSU (State College, PA), Defence Research and Development Canada—DRDC-A (Canada), and the Naval Research Laboratory—NRL (Washington, DC).

[1]M. D. Collins, W. A. Kuperman, and H. Schmidt, "Nonlinear inversion for ocean-bottom properties," J. Acoust. Soc. Am. **93**, 2770–2783 (1992).
[2]C. E. Lindsay and N. R. Chapman, "Matched field inversion for geoacoustic model parameters using adaptive simulated annealing," IEEE J. Ocean. Eng. **18**, 224–231 (1993).
[3]P. Gerstoft and C. F. Mecklenbräuker, "Ocean acoustic inversion with estimation of a posteriori probability distribution," J. Acoust. Soc. Am. **104**, 808–819 (1998).
[4]C. F. Mecklenbräuker and P. Gerstoft, "Objective functions for ocean acoustic inversion derived by likelihood methods," J. Comput. Acoust. **8**, 259–270 (2000).
[5]S. E. Dosso, M. J. Wilmut, and A.-L. S. Lapinski, "An adaptive-hybrid algorithm for geoacoustic inversion," IEEE J. Ocean. Eng. **26**, 324–336 (2001).
[6]C. W. Holland and J. Osler, "High-resolution geoacoustic inversion in shallow water: A joint time- and frequency-domain technique," J. Acoust. Soc. Am. **107**, 1263–1279 (2000).
[7]C. W. Holland, "Seabed reflection measurement uncertainty," J. Acoust. Soc. Am. **114**, 1861–1873 (2003).
[8]S. E. Dosso, "Quantifying uncertainty in geoacoustic inversion. I. A fast Gibbs sampler approach," J. Acoust. Soc. Am. **111**, 129–142 (2002).
[9]J. Dettmer, S. E. Dosso, and C. W. Holland, "Uncertainty estimation in seismo-acoustic reflection travel-time inversion," J. Acoust. Soc. Am. **122**, 161–176 (2007).
[10]J. Dettmer, S. E. Dosso, and C. W. Holland, "Full wave-field reflection coefficient inversion," J. Acoust. Soc. Am. **122**, 3327–3337 (2007).
[11]J. Dettmer, S. E. Dosso, and C. W. Holland, "Joint time/frequency-domain inversion of reflection data for seabed geoacoustic profiles," J. Acoust. Soc. Am. **123**, 1306–1317 (2008).
[12]C. W. Holland, J. Dettmer, and S. E. Dosso, "Remote sensing of sediment density and velocity gradients in the transition layer," J. Acoust. Soc. Am. **118**, 163–177 (2005).
[13]S. E. Dosso, P. L. Nielsen, and M. J. Wilmut, "Data error covariance in matched-field geoacoustic inversion," J. Acoust. Soc. Am. **119**, 208–219 (2006).
[14]S. E. Dosso and C. W. Holland, "Geoacoustic uncertainties from viscoelastic inversion of seabed reflection data," IEEE J. Ocean. Eng. **31**, 657–671 (2006).
[15]D. J. Battle, P. Gerstoft, W. S. Hodgkiss, W. A. Kuperman, and P. L. Nielsen, "Bayesian model selection applied to self-noise geoacoustic inversion," J. Acoust. Soc. Am. **116**, 2043–2056 (2004).
[16]Y. Jiang, N. R. Chapman, and H. A. DeFerrari, "Geoacoustic inversion of broadband data by matched beam processing," J. Acoust. Soc. Am. **119**, 3707–3716 (2006).
[17]A. E. Gelfand, D. K. Dey, and H. Chang, *Bayesian Statistics 4* (Oxford University Press, Oxford, 1992), pp. 147–167.
[18]A. E. Gelfand and D. K. Dey, "Bayesian model choice: Asymptotics and exact calculations," J. R. Stat. Soc. **56**, 501–514 (1994).
[19]D. C. Montgomery and E. A. Peck, *Introduction to Linear Regression Analysis* (Wiley, New York, 1992).
[20]R. L. Parker, *Geophysical Inverse Theory* (Princeton University Press, Princeton, NJ, 1994).
[21]D. J. C. MacKay, *Information Theory, Inference, and Learning Algorithms* (Cambridge University Press, Cambridge, 2003).
[22]H. Akaike, *Proceedings of the Second International Symposium in Information Theory* (Akademiai Kiado, Budapest, 1973), pp. 267–281.
[23]R. E. Kass and A. E. Raftery, "Bayes factors," J. Am. Stat. Assoc. **90**, 773–795 (1995).
[24]G. Schwartz, "Estimating the dimension of a model," Ann. Stat. **6**, 461–464 (1978).
[25]S. E. Dosso and M. J. Wilmut, "Uncertainty estimation in simulataneous Bayesian tracking and environmental inversion," J. Acoust. Soc. Am. **124**, 82–97 (2008).
[26]A. F. M. Smith, "Bayesian computational methods," Philos. Trans. R. Soc. London, Ser. A **337**, 369–386 (1991).
[27]A. F. M. Smith and G. O. Roberts, "Bayesian computation via the Gibbs sampler and related Markov Chain Monte Carlo methods," J. R. Stat. Soc. Ser. B (Methodol.) **55**, 3–23 (1993).
[28]*Markov Chain Monte Carlo in Practice*, Interdisciplinary Statistics, edited by W. R. Gilks, S. Richardson, and D. J. Spiegelhalter (Chapman and Hall, London/CRC, Boca Raton, FL, 1996).
[29]M. Sambridge and K. Mosegaard, "Monte Carlo methods in geophysical inverse problems," Rev. Geophys. **40**, 3-1–3-29 (2002).
[30]A. Tarantola, *Inverse Problem Theory and Methods for Model Parameter Estimation* (Siam, Philadelphia, PA, 2005).
[31]S. E. Dosso and M. J. Wilmut, "Data uncertainty estimation in matched-field geoacoustic inversion," IEEE J. Ocean. Eng. **31**, 470–479 (2005).
[32]A. Malinverno and V. A. Briggs, "Expanded uncertainty quantification in inverse problems: Hierarchichal Bayes and empirical Bayes," Geophysics **69**, 1005–1016 (2004).
[33]M. Sambridge, K. Gallagher, A. Jckson, and P. Rickwood, "Trans-dimensional inverse problems, model comparison and the evidence," Geophys. J. Int. **167**, 528–542 (2006).
[34]N. Metropolis, A. Rosenbluth, M. Rosenbluth, and A. T. A. E. Teller, "Equations of state calculations by fast computing machines," J. Chem. Phys. **21**, 1087–1092 (1953).
[35]W. K. Hastings, "Monte Carlo sampling methods using markov chains and their applications," Biometrika **57**, 97–109 (1970).
[36]W. Gropp, E. Lusk, and A. Skjellum, *Using MPI, Portable Parallel Programming With the Message-Passing Interface* (MIT, Cambridge, MA, 1999).

[37]S. Chib, "Marginal likelihood from the Gibbs output," J. Am. Stat. Assoc. **90**, 1313–1321 (1995).
[38]J. Skilling, *Bayesian Statistics 8* (Oxford University Press, Oxford, 2007), pp. 491–524.
[39]M. A. Newton and A. E. Raftery, "Approximate Bayesian inference with the weighted likelihood bootstrap (with discussions)," J. R. Stat. Soc. **56**, 3–48 (1994).
[40]J. J. K. O. Ruanaidh and W. J. Fitzgerald, *Numerical Bayesian Methods Applied to Signal Processing* (Springer, New York, 1996).
[41]S. Chib and I. Jeliazkov, "Marginal likelihood from the Metropolis-Hastings output," J. Am. Stat. Assoc. **96**, 270–281 (2001).
[42]J. R. Shaw, M. Bridges, and M. P. Hobson, "Efficient Bayesian inference for multimodal problems in cosmology," Mon. Not. R. Astron. Soc. **378**, 1365–1370 (2006).
[43]I. Murray, "Advances in Markov chain Monte Carlo methods," Ph.D. thesis, Gatsby Computational Neuroscience Unit, University College London, London (2007).
[44]P. J. Green, "Reversible jump markov chain Monte Carlo computation and bayesian model determination," Biometrika **82**, 711–732 (1995).
[45]A. Malinverno and W. S. Leaney, "Monte-Carlo Bayesian look-ahead inversion of walkaway vertical seismic profiles," Geophys. Prospect. **53**, 689–703 (2005).
[46]A. Gelman and X.-L. Meng, "Simulating normalizing constants: From importance sampling to bridge sampling to path sampling," Stat. Sci. **13**, 163–185 (1998).
[47]R. M. Neal, "Annealed importance sampling," Stat. Comput. **11**, 125–139 (2001).
[48]R. E. Kass and A. E. Raftery, "A reference Bayesian tests for nested hypotheses and its relationship to the Schwarz criterion," J. Am. Stat. Assoc. **90**, 928–934 (1995).
[49]C. H. Harrison and J. A. Harrison, "A simple relationship between frequency and range averages for broadband sonar," J. Acoust. Soc. Am. **97**, 1314–1317 (1995).

Comparison of focalization and marginalization for Bayesian tracking in an uncertain ocean environment

Stan E. Dosso and Michael J. Wilmut
School of Earth and Ocean Sciences, University of Victoria, Victoria, British Columbia V8W 3P6, Canada

(Received 30 August 2008; revised 30 November 2008; accepted 3 December 2008)

This paper compares focalization and marginalization approaches to source tracking when uncertain ocean environmental parameters are included, in addition to source locations, in a Bayesian inversion formulation. Focalization consists of determining the source track that maximizes the posterior probability density (PPD) over all source and environmental parameters. An efficient focalization approach is developed by applying the Viterbi algorithm to compute the optimal track from range-depth conditional probability distributions for each realization of the environmental parameters. This allows source locations to be treated implicitly and the optimization to be applied only to environmental parameters, substantially reducing the dimensionality and complexity of the problem. Marginalization consists of first integrating the PPD over the environmental unknowns to obtain a sequence of joint marginal probability distributions over source range and depth along the track. Applying the Viterbi algorithm to these marginal distributions defines the track estimate, and the distributions themselves quantify the track uncertainty. Monte Carlo analysis of the two approaches for a test case involving both geoacoustic and water-column uncertainties indicates that marginalization provides a significantly more reliable approach to tracking in an unknown environment. [DOI: 10.1121/1.3056555]

PACS number(s): 43.30.Pc, 43.30.Wi, 43.60.Pt [AIT] Pages: 717–722

I. INTRODUCTION

It is well known that the ability to localize and track an acoustic source is strongly affected by the state of knowledge of the ocean environment and that environmental uncertainty often represents the limiting factor for localization in shallow water.[1–15] To account for environmental uncertainty in localization, unknown environmental parameters can be included, in addition to the source location(s), in an augmented inverse problem. Two general approaches, referred to here as focalization and marginalization, have been applied to solve this augmented problem for source parameters.

Focalization consists of minimizing the data misfit function over all parameters to determine the globally optimal solution, including source parameters. The method of focalization[6–8] applies this approach to source localization in an uncertain environment. However, to date, it does not appear that focalization over source and environmental parameters has been applied to track a moving source, perhaps because of the numerical effort required to optimize over a high-dimensional parameter space that includes multiple source locations. This problem is addressed here by developing an efficient tracking focalization approach that treats multiple source locations as implicit (rather than explicit) parameters in the inversion, thereby substantially reducing the dimensionality and difficulty of the optimization.

Marginalization consists of integrating the posterior probability density (PPD) over the environmental parameters, which represent nuisance parameters in localization, to obtain joint marginal probability distributions over source range and depth. Source parameters can then be extracted as the most probable values of these marginals. This general approach was originally developed and applied to localization and tracking as the optimum uncertain field processor[9–11] and the optimum uncertain field tracking algorithm,[12,13] respectively. More recent applications of marginalization to localization and tracking are found in Refs. 14 and 15. An advantage of marginalization is that the marginal distributions also quantify the uncertainty of the source localization estimates, while focalization provides no measure of uncertainty.

Focalization and marginalization represent distinct approaches in estimating parameters of interest in the presence of nuisance parameters and generally produce different solutions for nonlinear problems such as acoustic inversion. The two approaches are illustrated conceptually in Fig. 1, which considers determining the value of a model parameter of interest, m_2, in a series of toy inverse problems that also include a nuisance parameter, m_1. Figure 1(a) considers the case of a PPD that is unimodal and symmetric (e.g., a linear inverse problem). Figure 1(b) shows the corresponding marginal distribution for m_2 obtained by integrating over m_1. In this case, the m_2 value obtained by two-dimensional (2D) optimization [the m_2 value at the peak of the PPD, indicated by the dotted lines in Fig. 1(a)] and the m_2 value obtained by marginalization [the m_2 value at the peak of the marginal distribution, indicated by the dotted line in Fig. 1(b)] are identical. Figures 1(c) and 1(d) consider a nonlinear problem with a PPD that is symmetric but not unimodal. In this case the m_2 values obtained by optimization [indicated in Fig. 1(c)] and marginalization [Fig. 1(d)] do not coincide. Finally, Figs. 1(e) and 1(f) consider a nonlinear problem with a PPD that is unimodal but not symmetric; the solutions for m_2 obtained via optimization and marginalization again do not coincide.

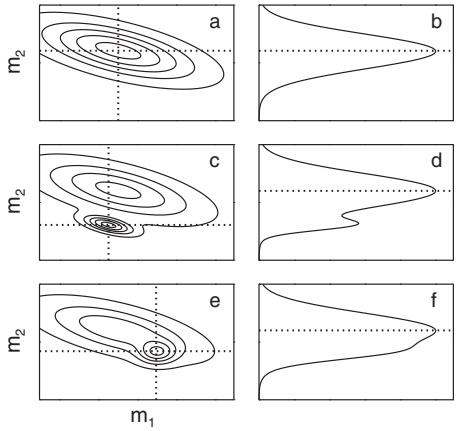


FIG. 1. Conceptual illustration of focalization (optimization) and marginalization approaches to determine a parameter of interest m_2 for a 2D toy inverse problem that includes a nuisance parameter m_1. The left column shows PPDs (contours) for three different inverse problems, with values of m_1 and m_2 that maximize the PPDs indicated by dotted lines. The right column shows corresponding marginal distributions for m_2, with m_2 values that maximize the marginals indicated by dotted lines. Note that for the two (nonlinear) examples in (c)–(f) that focalization and marginalization results for m_2 do not agree.

Although focalization and marginalization generally provide different solutions to source localization in an uncertain environment, there does not appear to have been any comparison of the two approaches to date. This paper compares the two approaches for the source-tracking problem, formulated within a unifying Bayesian framework (Sec. II). Monte Carlo methods are employed in which both tracking approaches are applied to a large number of noisy synthetic data sets computed for a test case involving uncertainty in seabed and water-column parameters. A statistical analysis of the results is carried out in terms of the probability of estimating a source track that is acceptably close to the true track (Sec. III). Marginalization is found to produce substantially better tracking results over a wide range of noise levels.

II. THEORY AND ALGORITHMS

This section summarizes a Bayesian approach to source tracking in an uncertain environment; more complete treatments of Bayesian theory can be found elsewhere.[16–19] Let $\mathbf{d}$ represent acoustic data from a sequence of source locations and $\mathbf{m}$ represent the model parameters comprised of the unknown source locations and environmental properties, with elements of both vectors considered random variables. Bayes' rule may be written as

$$P(\mathbf{m}|\mathbf{d}) \propto P(\mathbf{d}|\mathbf{m})P(\mathbf{m}), \tag{1}$$

where the PPD, $P(\mathbf{m}|\mathbf{d})$, represents the state of information for the model incorporating both data information, $P(\mathbf{d}|\mathbf{m})$, and prior information, $P(\mathbf{m})$. Interpreting the conditional probability $P(\mathbf{d}|\mathbf{m})$ as a function of $\mathbf{m}$ for the (fixed) measured data defines the likelihood function, $L(\mathbf{m}) \propto \exp[-E(\mathbf{m})]$, where E is the data misfit function. Combining data and prior as a generalized misfit

$$\phi(\mathbf{m}) \equiv E(\mathbf{m}) - \log_e P(\mathbf{m}), \tag{2}$$

the PPD can be written as

$$P(\mathbf{m}|\mathbf{d}) = \frac{\exp[-\phi(\mathbf{m})]}{\int \exp[-\phi(\mathbf{m}')]d\mathbf{m}'}, \tag{3}$$

where the domain of integration spans the multidimensional parameter space. The PPD is typically interpreted in terms of parameter estimates and uncertainties. Of interest in this paper are the maximum *a posteriori* estimate and 2D (joint) marginal probability distributions defined, respectively, as

$$\hat{\mathbf{m}} = \arg_{\max}\{P(\mathbf{m}|\mathbf{d})\}, \tag{4}$$

$$P(m_i,m_j|\mathbf{d}) = \int \delta(m_i' - m_i)\delta(m_j' - m_j)P(\mathbf{m}'|\mathbf{d})d\mathbf{m}', \tag{5}$$

where δ is the Dirac delta function. For nonlinear problems, such as acoustic inversion, numerical solutions to the optimization and integration in Eqs. (4) and (5) are required.

To define the data misfit, $E(\mathbf{m})$, consider complex acoustic-field data due to a (moving) source at S locations as measured at an array of N sensors at F frequencies, i.e., $\mathbf{d} = \{\mathbf{d}_{jk}, j=1,S, k=1,F\}$. Assuming that the data errors are complex, Gaussian-distributed random variables with variance ν, the likelihood function is given by

$$L(\mathbf{m}) = \prod_{j=1}^{S}\prod_{k=1}^{F} \frac{1}{(\pi\nu_{jk})^N} \exp[-|\mathbf{d}_{jk} - A_{jk}e^{i\theta_{jk}}\mathbf{d}_{jk}(\mathbf{m})|^2/\nu_{jk}], \tag{6}$$

where $\mathbf{d}_{jk}(\mathbf{m})$ are replica data, and A and θ represent source amplitude and phase, respectively. An unknown source can be treated by maximizing the likelihood over A, θ, and ν (i.e., setting $\partial L/\partial A = \partial L/\partial\theta = \partial L/\partial\nu = 0$) to give[20,21]

$$E(\mathbf{m}) = N\sum_{j=1}^{S}\sum_{k=1}^{F} \log_e B_{jk}(\mathbf{m}), \tag{7}$$

where $B_{jk}(\mathbf{m})$ represents the Bartlett mismatch

$$B_{jk}(\mathbf{m}) = |\mathbf{d}_{jk}|^2 - \frac{|\mathbf{d}_{jk}(\mathbf{m})^\dagger \mathbf{d}_{jk}|^2}{|\mathbf{d}_{jk}(\mathbf{m})|^2}. \tag{8}$$

The focalization and marginalization approaches to source tracking in an uncertain environment (described below) make use of the Viterbi algorithm,[22] which was first applied to marginalization in this context in Refs. 12 and 13. The Viterbi algorithm is a general and efficient dynamic-programming scheme for finding the most likely series of hidden states that correspond to a sequence of observed events. The algorithm can be applied as follows to determine the most probable track through a time-ordered series of probability surfaces over source range r and depth z, subject to constraints that the maximum radial and vertical source speeds are less than v_r^+ and v_z^+, respectively (i.e., the source location at time t is restricted to be within $v_r^+\Delta t$ in r and $v_z^+\Delta t$ in z of the location at time $t \pm \Delta t$). Consider a point on the second surface: The products of the probability of this point and the probabilities of all points on the first surface within the maximum allowable displacements in r and z are com-

puted, and the point on the first surface that maximizes the probability product is taken to be the antecedent to the point on the second surface. This procedure is carried out for every point on the second surface, with antecedents of all points as well as the surface of maximum probability products stored. Then for every point on the third surface, the products of the probability of this point and the maximum probability products of the first and second surfaces are computed for all points within constraints. The point on the second surface that produces the maximum probability product over all three surfaces is taken as the antecedent to the point on the third surface. Extending this procedure to an arbitrary number of surfaces, the maximal point on the final probability product surface defines the endpoint of the most probable track. Since antecedents on all previous surfaces are stored, the most probable source track satisfying the source-motion constraints is determined.

The focalization approach to source tracking consists of determining values for the model parameters (source locations and environment) that maximize the PPD, or, equivalently, minimize the misfit. A variety of numerical algorithms can be applied to this optimization problem. However, the key to the efficient focalization approach developed here is treating the source locations by defining the track as implicit parameters in the optimization. This is accomplished by computing a sequence of 2D conditional probability distributions over source range and depth (one distribution for each acoustic measurement time t), then applying the Viterbi algorithm to compute the optimal track from these conditionals. (For a range-independent environment, conditional distributions can be computed efficiently using normal mode methods since the modal properties are defined by the environment, not the source locations.) This procedure allows the optimization to be applied only to the environmental parameters, with the optimal track parameters computed directly for each realization of the environment, thereby greatly reducing the dimensionality and complexity of the problem. In a variety of test cases (not shown), this approach outperformed explicit optimization over all source and environment parameters by a wide margin and often gave good tracking results when explicit optimization failed to converge (even for long optimization runs). For the cases considered in this paper, the numerical optimizations were carried out using both differential evolution[23] and adaptive simplex simulated annealing[24] with similar results.

The marginalization approach to tracking consists of computing a sequence of joint marginal probability distributions over source range and depth by integrating the PPD over environmental nuisance parameters, with track constraints on the maximum source speed applied as prior information. The Viterbi algorithm is then applied to determine the most probable track from these marginals. The marginal distributions, referred to as probability ambiguity surfaces (PASs), are computed using numerical methods. A distinction between the present work and that of the original optimum uncertain field tracking algorithm[12,13] involves the integration approach applied here, described in detail in Ref. 15. In short, a powerful Markov-chain Monte Carlo method is applied here which combines Metropolis–Hastings Gibbs sampling (GS) for environmental parameters and heat-bath GS for source ranges and depths. The efficiency of Metropolis–Hastings sampling is improved by drawing parameter perturbations from a proposal distribution based on a linearized approximation to the PPD. The covariance matrix of this approximate distribution is initially computed via linearization about a starting model determined via optimization; the covariance is updated adaptively with a nonlinear estimate based on the sampling to that point, which better represents the overall nonlinear structure of the parameter space. This proposal distribution incorporates rotation to a principal-component parameter space as well as providing characteristic perturbation length scales based on principal-component parameter variance estimates. Although the theoretical linearized distribution is Gaussian, a Cauchy distribution is used for the proposal distribution to provide a higher proportion of large perturbations.[15] To further ensure complete parameter sampling, GS at nonunity sampling temperatures is employed, with appropriate weighting of the samples in evaluating the integral to ensure an unbiased estimate[15,25] (employing temperatures greater than unity samples the parameter space more intensively at the cost of increased computation time). To impose track constraints limiting radial and vertical source speeds, the r-z values for heat-bath GS for a source location at time t are restricted to be within $v_r^+\Delta t$ in r and $v_z^+\Delta t$ in z of the values for neighboring locations at times $t\pm\Delta t$ (one-sided constraints are applied at the track endpoints).

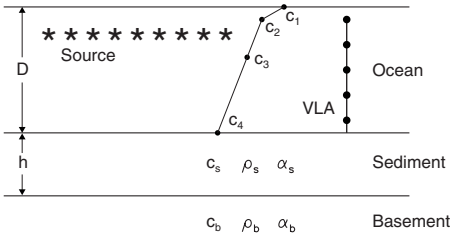


FIG. 2. Schematic of the tracking example, including unknown environmental parameters (defined in text), source locations, and vertical line array.

III. RESULTS

The Monte Carlo simulation study considered here involves tracking a quiet submerged source in shallow water with poor knowledge of the environment (seabed and water column), as illustrated in Fig. 2. Seabed geoacoustic parameters include the thickness h of an upper sediment layer with sound speed c_s, density ρ_s, and attenuation α_s, overlying a semi-infinite basement with sound speed c_b, density ρ_b, and attenuation α_b. The water depth is D, and the water-column sound-speed profile is represented by four parameters c_1-c_4 at depths of 0, 10, 50, and D m. Table I gives the true values for the environmental parameters together with the limits for the wide uniform prior distributions (search bounds) assumed for all parameters.

Simulated acoustic data are computed at a frequency of 300 Hz at a vertical array consisting of 24 sensors at 4 m spacing from 26 to 118 m depth using the normal-mode model ORCA.[26] The track consists of an acoustic source at

TABLE I. True values and bounds for uniform prior distributions for the environmental parameters used in the Monte Carlo Study.

Parameter and units	True value	Lower bound	Upper bound
h (m)	9	0	30
c_s (m/s)	1495	1450	1600
c_b (m/s)	1530	1500	1650
ρ_s (g/cm^3)	1.4	1.0	1.7
ρ_b (g/cm^3)	1.6	1.5	2.2
α_s (dB/λ)	0.2	0	1.0
α_b (dB/λ)	0.2	0	1.0
D (m)	130	128	132
c_1 (m/s)	1520	1515	1525
c_2 (m/s)	1517	1510	1520
c_3 (m/s)	1515	1510	1520
c_4 (m/s)	1510	1505	1515

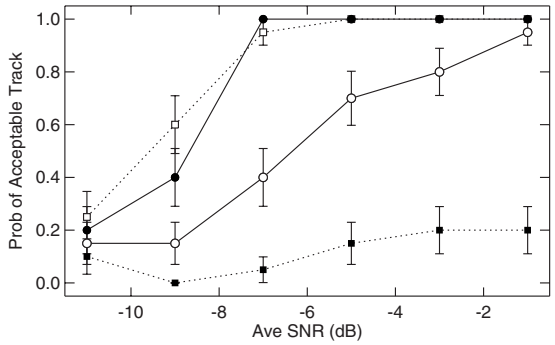


FIG. 4. PAT computed as a function of the average SNR along the track for focalization and marginalization approaches (open and filled circles, respectively), with one standard-deviation error bars. PAT values computed with exact environmental knowledge and with uncertain environmental knowledge (random parameters drawn from prior) are indicated by open and filled squares, respectively.

20 m depth moving away from the array at a constant radial speed of 5 m/s (~10 kts). Acoustic data are collected at the array once per minute for 9 min, corresponding to source-receiver ranges of $4.0, 4.3, \ldots, 6.4$ km. Prior bounds for the source location at all times are 0.1–10 km range and 2–120 m depth (numerical grids with 50 m spacing in range and 2 m spacing in depth are applied for source locations). Complex Gaussian-distributed random errors of constant variance are added to the acoustic data, with signal-to-noise ratio (SNR) decreasing as the received signal strength decreases with range by approximately 6 dB over the track, as shown in Fig. 3. (SNR is defined here as SNR $= 10 \log|\mathbf{s}|^2/|\mathbf{n}|^2$, where $\mathbf{s}$ and $\mathbf{n}$ represent the signal and noise vectors over the array, respectively.)

The focalization and marginalization approaches for source tracking were applied to 20 realizations of noisy acoustic data at six different noise levels, with track-averaged SNRs varying from approximately −1 to −11 dB. Constraints on the radial and vertical source speed were $v_r^+ = 10$ m/s and $v_z^+ = 0.06$ m/s, respectively (i.e., the source moves less than 600 m radially and 4 m vertically in the minute between data observations). Tracking algorithm performance is quantified here in terms of the probability of an acceptable track (PAT), defined here as achieving mean absolute errors in source range and depth over the track of less than 500 and 10 m, respectively. Since the goal of this work is to quantify the relative tracking performance of the conceptually different approaches of focalization and marginalization, the algorithms were run carefully to ensure that the results are not limited by lack of numerical effort. For focalization, both differential evolution and adaptive simplex simulated annealing were applied intensively, and the best result adopted (results were similar for the two algorithms). Marginalization was run at sampling temperatures of $T = 1$, 2, and 4 to ensure a complete sampling of the parameter space in integration.[15,25]

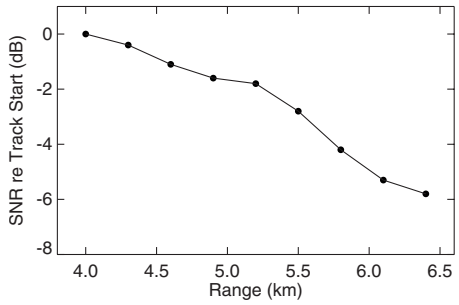


FIG. 3. Relative variation in SNR at the array as the source moves radially outward along the track (referenced to the SNR at the start of the track).

The PAT values obtained for focalization and marginalization are shown in Fig. 4 as a function of average SNR, with one standard-deviation binomial uncertainties estimated as[27]

$$\sigma_{\hat{p}} = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}, \tag{9}$$

where $\hat{p}$ represents the number of acceptable tracks obtained in n attempts (i.e., inversion of n noisy data sets). Also included in Fig. 4 for comparison are PAT values computed using exact knowledge of the environment, and PAT values that statistically represent the environmental uncertainty (i.e., environmental parameters are drawn at random from the prior distribution). Since both of these cases involve fixed environmental parameters for each noise realization, the track is computed directly via the Viterbi algorithm and there is no distinction between marginalization and focalization.

Figure 4 shows that marginalization provides significantly higher PATs than focalization for all but the lowest SNR. In fact, marginalization produced PAT values of 1.0 (i.e., 20/20 acceptable tracking results) for average SNRs of −1 to −7 dB, while the PATs from focalization decrease from approximately 0.95–0.35 over this SNR range. At −9 dB the difference between the marginalization and focalization results is smaller (0.4–0.15), and at −11 dB, where all approaches perform poorly, the difference is not statistically significant (although the average marginalization result is slightly better). Figure 4 also shows that both marginalization and focalization perform substantially better than tracking with environmental parameters drawn from the prior, which produces PAT values of 0–0.2 for over the range of SNR values. Of particular significance is the observation that the average marginalization results are comparable (within

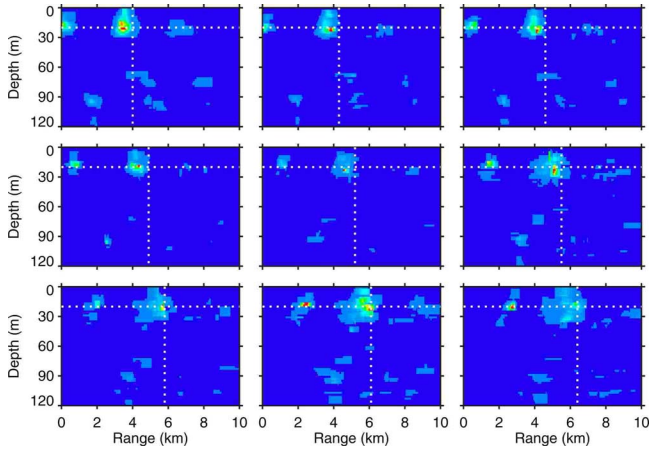


FIG. 5. (Color online) PASs computed for the nine source positions along the track for an average SNR of −9 dB. True source ranges and depths are indicated by dotted lines. Note that for display purposes, each panel is normalized independently.

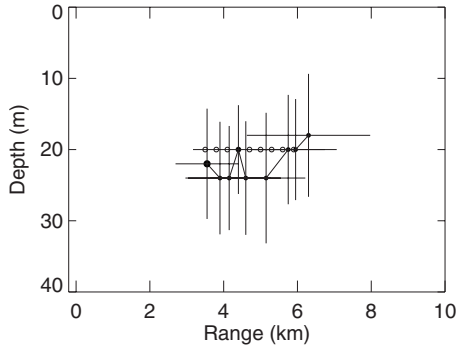


FIG. 6. Track estimate computed via marginalization for the same noise realization as Fig. 5 (average SNR of −9 dB). Closed circles indicate source location estimates (largest circle indicates the track start), with one mean-deviation error bars. The true track is indicated by open circles. Localization was carried out over the entire water column (2–120 m), but only the upper 40 m is shown.

one standard deviation) to the results of tracking with exact environmental knowledge at all SNR values. These overall results indicate that, for this example, the acoustic fields contain sufficient information to resolve the source track via marginalization despite environmental uncertainty, with noise representing the effective limiting factor.

Figure 4 indicates that marginalization significantly outperforms focalization for source tracking in an unknown environment. An additional advantage of marginalization is that the PASs from which the track estimate is derived provide a quantitative measure of tracking uncertainty. For example, Fig. 5 shows PASs at an average SNR of −9 dB computed for a noise realization for which marginalization determined an acceptable track but focalization did not. The PASs indicate two sets of strong maxima, one near the true source locations for all source ranges, and a second near the correct depth but at roughly 3 km shorter range; other weaker maxima are also evident. The maxima tend to be more tightly focused for the shorter source ranges (with higher SNRs) and spread out as range increases. For the two longest source ranges, the maxima at about 3 km shorter range represent the strongest peaks. However, the Viterbi algorithm eliminates these in determining the most probable track, shown in Fig. 6. Also included in Fig. 6 are mean (absolute) deviation uncertainties in depth and range for each source-location estimate, computed from the PASs (the mean deviation is generally considered a more meaningful measure of spread than the standard deviation for multimodal distributions). The increase in the radial mean deviations with increasing range in Fig. 6 quantifies the PAS characteristics observed in Fig. 5 (i.e., increased spread in the PAS peaks and strengthening ambiguities with increasing source range/decreasing SNR along the track).

IV. SUMMARY AND DISCUSSION

This paper considered focalization and marginalization approaches to Bayesian source tracking when uncertain environmental parameters are included as unknowns in an augmented inverse problem. In focalization, the PPD is maximized over all dimensions to provide the most probable set of parameters in the model space (but no measure of uncertainty). An efficient focalization approach was developed by applying the Viterbi algorithm to determine the optimal track for each realization of the environment, thereby allowing source parameters to be treated implicitly in the optimization. In marginalization, the posterior probability distribution is integrated over environmental nuisance parameters to produce marginal probability distributions over source range and depth, from which the source track and tracking uncertainties are extracted. The integration is carried out numerically using advanced Markov-chain Monte Carlo sampling methods. Constraints on the maximum radial and vertical source speed are applied in both focalization and marginalization approaches.

Focalization and marginalization represent distinct approaches in estimating parameters of interest in the presence of nuisance parameters and generally produce different solutions for nonlinear inverse problems. Hence, Monte Carlo analysis was applied to compare the two approaches for source tracking in an uncertain environment. Both approaches were applied to a large number of noisy synthetic data sets for a test case involving uncertain seabed and water-column parameters. Statistical analysis of the results was carried out in terms of the PAT (defined as mean range and depth errors over the track of less than 500 and 10 m, respectively). Based on this analysis, marginalization substantially outperformed focalization over a range of SNRs. In fact, marginalization results were comparable to tracking results obtained using exact knowledge of the environment. This indicates that, depending on the noise level, the acoustic data provide sufficient information to resolve the source track despite environmental uncertainty. Tracking uncertainty for marginalization was quantified in terms of PASs and as mean-deviation error estimates in range and depth.

[1]D. R. Del Balzo, C. Feuillade, and M. R. Rowe, "Effects of water-depth mismatch on matched-field localization in shallow water," J. Acoust. Soc. Am. **83**, 2180–2185 (1988).
[2]A. Tolstoy, "Sensitivity of matched field processing to sound-speed prone mismatch for vertical arrays in a deep water Pacific environment," J. Acoust. Soc. Am. **85**, 2394–2404 (1989).
[3]E. C. Shang and Y. Y. Wang, "Environmental mismatching effects on

source localization processing in mode space," J. Acoust. Soc. Am. **89**, 2285–2290 (1991).
[4]A. Tolstoy, *Matched Field Processing for Underwater Acoustics* (World Scientific, Singapore, 1993).
[5]A. B. Baggeroer, W. A. Kuperman, and P. N. Mikhalevsky, "An overview of matched field methods in ocean acoustics," IEEE J. Ocean. Eng. **18**, 401–424 (1993).
[6]M. D. Collins and W. A. Kuperman, "Focalization: Environmental focusing and source localization," J. Acoust. Soc. Am. **90**, 1410–1422 (1991).
[7]S. E. Dosso, "Matched-field inversion for source localization with uncertain bathymetry," J. Acoust. Soc. Am. **94**, 1160–1163 (1993).
[8]R. N. Baer and M. D. Collins, "Source localization in the presence of gross sediment uncertainties," J. Acoust. Soc. Am. **120**, 870–874 (2006).
[9]A. M. Richardson and L. W. Nolte, "*A posteriori* probability source localization in an uncertain sound speed, deep ocean," J. Acoust. Soc. Am. **89**, 2280–2284 (1991).
[10]J. A. Shorey, L. W. Nolte, and J. L. Krolik, "Computationally efficient Monte Carlo estimation algorithms for matched field processing in uncertain ocean environments," J. Comput. Acoust. **2**, 285–314 (1994).
[11]J. A. Shorey and L. W. Nolte, "Wideband optimal *a posteriori*, probability source localization in an uncertain shallow ocean environment," J. Acoust. Soc. Am. **103**, 355–361 (1998).
[12]S. L. Tantum and L. W. Nolte, "Tracking and localizing a moving source in an uncertain shallow water environment," J. Acoust. Soc. Am. **103**, 362–373 (1998).
[13]S. L. Tantum, L. W. Nolte, J. L. Krolik, and K. Haramci, "The performance of matched-field track-before-detect methods using shallow-water Pacific data," J. Acoust. Soc. Am. **112**, 119–127 (2002).
[14]S. E. Dosso and M. J. Wilmut, "Bayesian focalization: Quantifying source localization with environmental uncertainty," J. Acoust. Soc. Am. **121**, 2567–2574 (2007).
[15]S. E. Dosso and M. J. Wilmut, "Uncertainty estimation in simultaneous Bayesian tracking and environmental inversion," J. Acoust. Soc. Am. **124**, 82–97 (2007).
[16]A. Tarantola, *Inverse Problem Theory: Methods for Data Fitting and Model Parameter Estimation* (Elsevier, Amsterdam, 1987).
[17]M. K. Sen and P. L. Stoffa, *Global Optimization Methods in Geophysical Inversion* (Elsevier, Amsterdam, 1995).
[18]W. R. Gilks, S. Richardson, and G. J. Spiegelhalter, *Markov Chain Monte Carlo in Practice* (Chapman and Hall, London, 1996).
[19]J. J. K. O'Ruanaidh and W. J. Fitzgerald, *Numerical Bayesian Methods Applied to Signal Processing* (Springer-Verlag, New York, 1996).
[20]C. F. Mecklenbräuker and P. Gerstoft, "Objective functions for ocean acoustic inversion derived by likelihood methods," J. Comput. Acoust. **6**, 1–28 (2000).
[21]S. E. Dosso and M. J. Wilmut, "Estimating data uncertainty in matched-field geoacoustic inversion," IEEE J. Ocean. Eng. **31**, 470–479 (2006).
[22]A. J. Viterbi, "Error bounds on convolutional codes and an asymptotically optimal decoding algorithm," Proc. IEEE **61**, 268–278 (1973).
[23]K. V. Price, R. M. Storn, and J. A. Lampinen, *Differential Evolution: A Practical Approach to Global Optimization* (Springer, New York, 2005).
[24]S. E. Dosso, M. J. Wilmut, and A. L. Lapinski, "An adaptive hybrid algorithm for geoacoustic inversion," IEEE J. Ocean. Eng. **26**, 324–336 (2001).
[25]B. F. Brooks and L. N. Frazer, "Importance reweighting reduces dependence on temperature in Gibbs samplers: An application to the inverse coseismic geodetic problem," Geophys. J. Int. **161**, 12–20 (2005).
[26]E. K. Westwood, C. T. Tindle, and N. R. Chapman, "A normal mode model for acousto-elastic ocean environments," J. Acoust. Soc. Am. **100**, 3631–3645 (1996).
[27]R. E. Walpole, *Introduction to Statistics* (Macmillan, New York, 1982).

Green's function approximation from cross-correlations of 20–100 Hz noise during a tropical storm

Laura A. Brooks[a] and Peter Gerstoft
Marine Physical Laboratory, Scripps Institution of Oceanography, La Jolla, California 92093

(Received 8 May 2008; revised 2 October 2008; accepted 4 December 2008)

Approximation of Green's functions through cross-correlation of acoustic signals in the ocean, a method referred to as ocean acoustic interferometry, is potentially useful for estimating parameters in the ocean environment. Travel times of the main propagation paths between hydrophone pairs were estimated from interferometry of ocean noise data that were collected on three L-shaped arrays off the New Jersey coast while Tropical Storm Ernesto passed nearby. Examination of the individual noise spectra and their mutual coherence reveals that the coherently propagating noise is dominated by signals of less than 100 Hz. Several time and frequency noise normalization techniques were applied to the low frequency data in order to determine the effectiveness of each technique for ocean acoustic applications. Travel times corresponding to the envelope peaks of the noise cross-correlation time derivatives of data were extracted from all three arrays, and are shown to be in agreement with the expected direct, surface-reflected, and surface-bottom-reflected interarray hydrophone travel times. The extracted Green's function depends on the propagating noise. The Green's function paths that propagate horizontally are extracted from long distance shipping noise, and during the storm the more vertical paths are extracted from breaking waves. © *2009 Acoustical Society of America.* [DOI: 10.1121/1.3056563]

PACS number(s): 43.30.Pc, 43.60.Fg, 43.30.Nb [RCG] Pages: 723–734

I. INTRODUCTION

In 2001 Lobkis and Weaver[1] showed, both theoretically and experimentally, that the Green's function between two points can be determined from temporal cross-correlation within a diffuse ultrasonic field. Extraction of the Green's function by cross-correlation has since been applied to numerous areas including ultrasonics,[2,3] seismic noise,[4–12] and ocean acoustics.[13–20] The approach of inferring the Green's function between two receivers from noise cross-correlations is referred to here as ocean acoustic interferometry (OAI), due to its relationship to classical and seismic interferometries, where interferometry refers to the determination of information from the interference phenomena between pairs of signals.[11] Several source types have previously been used for OAI: active, ocean wave, biological, and ship. The relationship between the cross-correlation of sound from a column of active sources in the ocean recorded by receivers in the same plane, and the Green's function between the receivers has been demonstrated theoretically and through simulation.[13,18] Sabra *et al.*[16] cross-correlated biological noise (croaker fish) from 150 to 700 Hz data. They obtained the direct arrival between hydrophones in a bottom array and used these arrival times for array self-localization. Using a vertical array of hydrophones, ocean surface wave noise cross-correlation was used to extract seafloor structure via passive fathometry.[17,19,20]

Roux *et al.*[14] showed experimentally that for a ship track passing through the end-fire region of a pair of hydrophones, the signal from the end-fire region dominates the cross-correlation. They extracted the direct arrival between the hydrophones. Simulations for sources (ships) at various ranges along the hydrophone end-fire direction showed that the cross-correlations emphasize different Green's function arrival paths, depending on the range.[14] The Green's function that is recovered is therefore not a true Green's function because each cross-correlation peak differs from that of the corresponding Green's function arrival peak by a path dependent amplitude factor; it is therefore termed an "amplitude shaded" Green's function.

Although the theory prescribes a uniform noise distribution, good approximations of the arrival structure of the actual Green's function can still be obtained from the cross-correlation time derivative, termed the *empirical Green's function* (EGF), even for nonuniform noise distributions.[4,21,22] To obtain an EGF For ship dominated noise without directional bias, the observation time must include several ship tracks passing through the end-fire region.[14] The ocean is nonstationary, suggesting that OAI over short time periods, such as a few minutes, is sufficient if instantaneous EGFs are desired. However, the need to average over multiple ship tracks requires longer observation times, and hence an "average" EGF over a long observation time (24 h) is obtained here. When using noise from breaking waves to extract EGFs a short observation time can be used. The theory here assumes that sources all have the same amplitude and frequency contents. Nearby ships tend to be louder, and larger ships have spectra that are dominated by lower frequencies. Time and frequency preprocessing are carried out to minimize these effects.

OAI of 20–100 Hz noise is considered in detail here. Data were collected on three L-shaped arrays from 31 Au-

[a]Also at the School of Mechanical Engineering, The University of Adelaide, Australia. Present address: Institute of Geophysics, Victoria University of Wellington, New Zealand. Electronic mail: laura.brooks@vuw.ac.nz

gust to 3 September 2006 during the Shallow Water 2006 (SW06) experiment. Tropical Storm Ernesto[23] passed through the area on September 2, creating a richer noise field that is well suited for extracting EGFs.

At frequencies below about 100 Hz the noise field is usually dominated by shipping noise.[24] Nearby shipping favors higher grazing angles, while distant shipping favors more horizontally traveling wavefronts. During Tropical Storm Ernesto local ships left the region. Thus, the shipping portion of the noise field was dominated by distant vessels. For a horizontal set of hydrophones, the direct path EGF will be dominated by noise from distant ships.

The passing of Ernesto resulted in more acoustic energy at low frequencies from breaking waves and this noise is also distributed at higher propagation angles, as is evident from the beamforming on the vertical array. The higher propagating angles of the breaking wave noise enable the extraction of EGFs for more vertically traveling paths. Because of the higher noise levels and richer angular distribution from breaking wave noise and the more azimuthally uniform time-averaged shipping noise field, the EGFs extracted during the storm match the actual Green's functions more closely.

Thus, through careful processing and a longer averaging time, combined with a more evenly distributed noise field, it is possible to extract not only the direct arrivals but also higher order multiples in the water column. Furthermore, these arrivals tend to extend farther in range than previous results, with sharp arrivals and good signal-to-noise ratios.

II. THEORETICAL BACKGROUND

Both shipping noise and ocean wave generated noise originate at or near the ocean surface, and hence it is assumed that the 20–100 Hz noise considered here can be modeled as a set of sources that are uniformly and densely distributed within a horizontal plane near the surface of a waveguide. The cross-correlation of the signals recorded at two receivers, A and B, can therefore be derived following the stationary-phase methodologies of Refs. 18 and 25:

$$C_{AB}(\omega) = |\rho S(\omega)|^2 n \int\int G(\mathbf{r}_A, \mathbf{r}_S) G^*(\mathbf{r}_B, \mathbf{r}_S) dx dy, \qquad (1)$$

where $S(\omega)$ is the ship or wave source spectrum, ρ is the density of the medium, n is the number of sources per unit area, $G(\mathbf{r}_\psi, \mathbf{r}_S)$ is the Green's function between the source, S, and receiver, ψ, * denotes the complex conjugate, and x and y are the horizontal axes parallel and perpendicular, respectively, to the vertical plane containing A and B.

Application of the method of stationary phase to Eq. (1),[15,18,25,26] as well as summation over all stationary points, yields

$$C_{AB}(\omega) = in|S(\omega)|^2 \sum_{\chi_s} \left(\frac{\Gamma^{b_A+b_B} c\rho}{2\omega \cos\theta} G_f(R(\chi_s)) \right), \qquad (2)$$

where χ_s are the stationary points, c is the wave velocity, f is the acoustic frequency, $\omega = 2\pi f$ is the angular frequency, Γ is the bottom reflection coefficient, b_ψ is the number of bottom bounces for a given path, θ is the acute angle between the ray path and the vertical, $G_f(R) = e^{ikR}/4\pi R$ is the three-dimensional Green's function within a homogeneous medium, where k is the wave number and R is the distance from the source, and $R(\chi_s)$ is the path length difference between a wave traveling from χ_s to A, and a wave traveling from χ_s to B. Note that the stationary points, χ_s, satisfy the relationship $\theta_A = \pm \theta_B$. The positive relationship between θ_A and θ_B only occurs when the path to the furthest receiver passes through the closer receiver, hence the relationship between the summed cross-correlations and the Green's function between the receivers. The negative relationship corresponds to stationary-phase contributions from cross-correlations between a wave that initially undergoes a surface reflection, and one that does not.[15,18] Since ship and wave sources are near the ocean surface, these spurious arrivals will converge to almost the same time delay as the true Green's function paths, and due to the long wavelengths, will not be observed as separate peaks. The model assumes that all sources have the same spectrum. The ocean wave ambient noise field is reasonably uniform, but the ship noise field is more discrete and each ship has a different spectrum. This will cause an unknown bias in the summed cross-correlation; however, averaged over a long observation time, this bias is reduced as each ship covers a large azimuthal area. The theory presented here has also neglected curvature of ray paths due to refraction, but it has been shown by others[25] that the stationary-phase argument generalizes to a heterogenous medium with smooth velocity variations.

The cross-correlation in Eq. (2) can therefore be seen to produce an amplitude and phase shaded Green's function. The amplitude shading is dependent on the travel path through the $\Gamma^{b_A+b_B}$ and $\cos\theta$ terms, and also contains constant and frequency dependent components. The $1/\omega$ factor phase shading in Eq. (2) means that the time domain Green's function is approximately proportional to the derivative of the summed cross-correlations:[5,15,18,21]

$$\frac{\partial C_{AB}(t)}{\partial t} \sim -[G_{AB}(t) - G_{AB}(-t)]. \qquad (3)$$

The raw cross-correlation, rather than its time derivative, is often used as an approximation to the Green's function[14,17,19,27] because this is better in certain environments.[25,28]

III. EXPERIMENT

The SW06 experiments were conducted off the New Jersey shelf. The data considered here, from measurements of opportunity, were collected from August 31 to September 3 on three L-shaped arrays: SWAMI52, SWAMI32, and Shark. L-shaped arrays were used as these allow for approximation of travel paths between bottom mounted hydrophones, as well as approximation of higher grazing angle paths between bottom hydrophones and those located within the watercolumn. Array locations and orientations are shown in Fig. 1(a), and configurations are detailed in Table I. The orientations of the horizontal portion of each array varied, and thus each array was sensitive to a different propagating environment. The SE-NW and S-N oriented SWAMI52 and Shark arrays

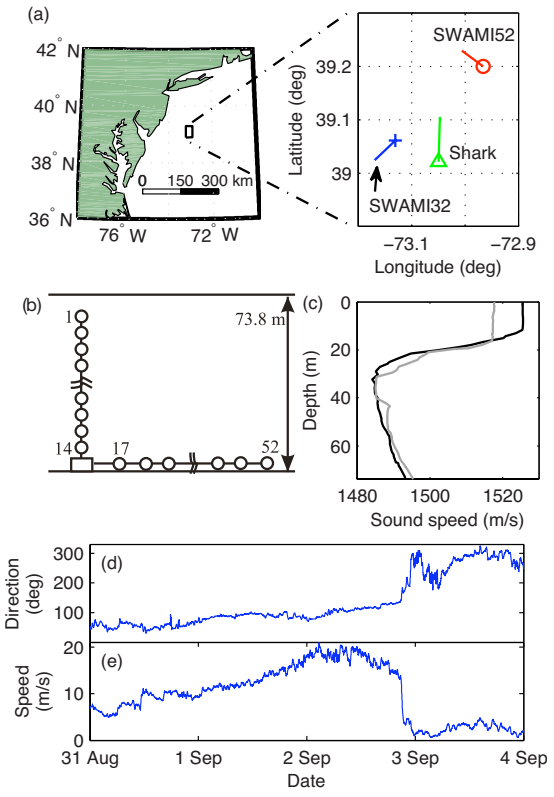


FIG. 1. (Color online) (a) Geographic location of experimental site (black rectangle) on New Jersey Shelf, with magnified view showing the relative locations of SWAMI52, SWAMI32, and Shark. The lines departing each VLA show the HLA orientation. The array length is scaled by a factor of 20. (b) SWAMI52 geometry and hydrophone numbering system. (c) Sound speed profiles near SWAMI52 for August 30 (black) and September 6 (gray). (d) Wind direction and (e) wind speed, from R/V *Knorr* ship records, from August 31 to end of September 3.

were sensitive to upslope propagation of noise from deeper waters, while the SWAMI32 array was oriented along the shelf (NE-SW).

The horizontal line arrays (HLAs) were all located on the seafloor. All vertical line array (VLA) hydrophones and the Shark HLA hydrophones are evenly spaced. SWAMI52 interhydrophone distances increase from the center, and SWAMI32 HLA interhydrophone distances increase from the hydrophone 13 (H-13) end, respectively. A sketch of the SWAMI52 geometry is shown in Fig. 1(b). Sound speed profiles that were recorded near SWAMI52 on August 30 and

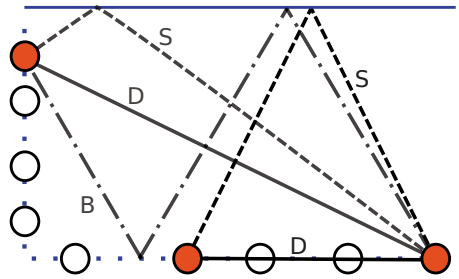


FIG. 2. (Color online) Direct (D, solid), surface-reflection (S, dashed), and surface-bottom-reflection (B, dashed-dotted) raypaths from HLA hydrophone (far right) to second HLA hydrophone (black paths, B does not exist), and to VLA hydrophone (gray paths).

September 6 are shown in Fig. 1(c). Example schematics of the raypaths that will be approximated from the cross-correlations are shown in Fig. 2.

Tropical Storm Ernesto passed through the experimental area during the data collection period, leading to large sea states and high winds. The wind direction and speed from August 31 to September 3 are shown in Figs. 1(d) and 1(e). Predominantly easterly winds gradually built up over August 31 and September 1 to a 20 m/s peak early on September 2, and then remained high until late in the day, when they dropped rapidly once the storm had passed. The decrease in speed was accompanied by a change in wind direction to south and west.

IV. ANALYSIS OF DATA PREPROCESSING METHODS

Time and frequency domain preprocessing methods were applied to the raw data to emphasize broadband ocean noise. The preprocessing techniques considered here were analyzed using the data collected on SWAMI52 throughout September 2 (Zulu time). No towed source experiments were undertaken on this day, because of Ernesto, and therefore ocean noise over a large frequency bandwidth could be considered. The data were stored and analyzed in 140 portions, each 10 min and 14 s (10:14 min) duration.

Short (10:14 min) cross-correlations were unstable above the thermocline, likely due to sound speed fluctuations resulting from the elevated levels of swell and mixing associated with Ernesto. If the noise field had been isotropic and sufficiently strong, OAI could have been performed over periods that were sufficiently short for the environment to be

TABLE I. Details of array configurations.

			VLA		HLA	
	Water depth (m)	No. hydrophones	Length (m)	Hydrophones[a]	Length (m)	Hydrophones[b]
SWAMI52	73.8	52	56.81	1:14	230	17:52
SWAMI32	68.5	32	53.55	1:10	256	13:32
Shark	79	48	64.25	0:12	465	16:47[c]

[a]Lowest numbered hydrophone is uppermost in the array. Extra hydrophones tied off just above the frame (SWAMI52: H-15 and H-16, SWAMI32: H-11 and H-12, and Shark: H-13–H-15).
[b]Lowest numbered hydrophone is closest to the array except for Shark, which is opposite.
[c]Data from H-15 and H-46 were discarded due to inconsistencies with other data.

considered stationary. The temporal change in cross-correlation could then have been related to environmental changes, in particular, changes in temperature in the upper waveguide, and tidal changes. However, discrete ships are a significant noise source here, and therefore the cross-correlations had to be performed over a long time period so that specific ship sources did not dominate (see Sec. VI).

Comparisons of time and frequency domain normalization techniques by Bensen *et al.*[12] concerned seismic noise, which is often dominated by high amplitude earthquakes and lower frequencies. Since other physical processes dominate ocean noise, the effect of these normalization techniques on the resulting EGFs will also differ. A comparative study of several techniques to the current data set was undertaken for both frequency (Sec. IV C) and time (Sec. IV D) domain preprocessing. Frequency normalization using a smoothed amplitude spectrum, and one-bit time normalization were then chosen as the preferred preprocessing methods.

A. Removal of main contamination

Depending on the particular time interval, some of the September 2 data exhibited high amplitude midfrequency signal from fixed location sound sources, amplitude clipping, and low frequency energy bursts. Electrical noise could manifest as high amplitude tonals at the hydrophone operation frequency and its harmonics, and/or as Gaussian noise across a wider frequency band. Impact noise from a fish colliding with a hydrophone or something else tapping the hydrophone array would likely be observed as sharp amplitude peaks in the time domain, and energy would be smeared across the frequency spectrum at this time. Signals from any ships near the array throughout the day would be recorded as discrete high amplitude tonals.

All discrete signals have a difference in direct path length to each hydrophone, which is less than or equal to the direct path between the hydrophones, and may be visible in the cross-correlation as spurious precursory arrivals. Preprocessing, which includes choice of bandwidth as well as time and frequency domain normalizations, ameliorates these effects (see Secs. IV B–IV D).

B. Spectra and coherence

Only signals that are received by both hydrophones will sum coherently to give a peak in the cross-correlation function. Both the signal amplitude and coherence are therefore considered here. Since the underlying statistics of the data are nonstationary, spectra and coherence over short periods are examined. Spectrograms of the signals received by H-52, and coherograms (coherence plotted as a function of frequency and time) of the signals recorded on H-52 and H-40, which are separated by 97.74 m, are shown in Fig. 3. The mean spectra and coherence for 1:33 min intervals were calculated using 2 s Hanning windowed data segments and 50% overlap. The mean of each set of spectra and coherences is shown on the far right.

High amplitude signals with high coherence are apparent at regular time intervals in the 200–410 Hz frequency range, as shown in Figs. 3(a) and 3(c). These signals are from three

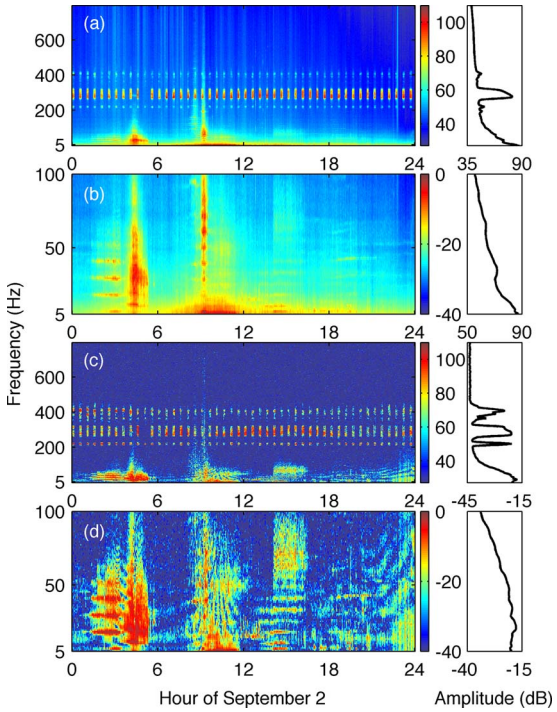


FIG. 3. (Color online) Spectrograms (dB) of signals recorded on H-52, using the entire September 2 data: (a) 5–800 and (b) 5–100 Hz. Cohero-grams (dB relative to unity linear coherence) of the data recorded on H-52 and H-40: (c) 5–800 and (d) 5–100 Hz. The average of each set of spectra or coherences is plotted to the right of each figure.

fixed location sound sources. Low signal amplitudes and coherence are observed at frequencies above 420 Hz, and also from 100 to 200 Hz. Below 100 Hz both the amplitude and coherence of the received signals are higher. This is expected since the attenuation of lower frequency ocean noise is less over long distances. A banded structure consisting of high amplitude tonals is observable in both the low frequency spectrogram and coherogram, Figs. 3(b) and 3(d), at a range of frequencies at different times throughout the day (e.g., 1:30–3:30Z and 13:30–15:30Z). This banded structure is indicative of ship noise at low frequencies.

Since the signals, apart from those emanating from the fixed sound sources, exhibit negligible amplitude and coherence above 100 Hz, the data were bandpass filtered to 20–100 Hz. The lower limit of 20 Hz was selected because frequencies below this have insufficient resolution for more closely spaced hydrophone pairs that are separated by only a few meters.

Beamformed data of 10:14 min duration, recorded on the SWAMI52 VLA at the start of each day of August 31–September 3, are shown in Fig. 4. The critical angle at the bottom can be observed from the beamformer output to be about 25°, corresponding to a sediment sound speed of 1650 m/s. The extraction of the critical angle from the beamformer seems more straightforward than extracting it from the noise cross-correlation.[29] The data from the two mornings before the storm, (a) and (b), show more horizontally traveling wave fronts, as would be expected for a noise field dominated by distant shipping and distant breaking waves. The data from the morning of the storm and the morning after, (c) and (d), show a significant increase in

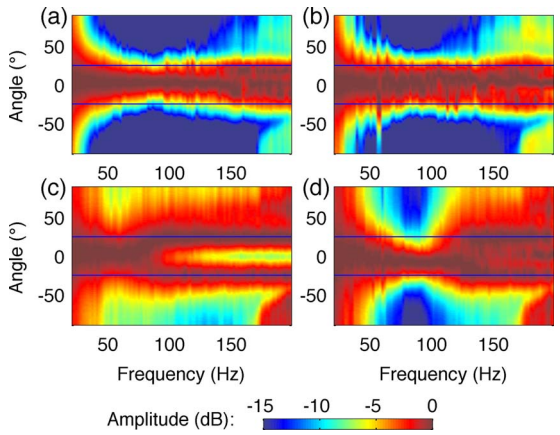


FIG. 4. (Color online) Beamformer output, normalized to maximum at each frequency, from 10:14 min of SWAMI52 VLA data at the start of (a) August 31, (b) September 1, (c) September 2, and (d) September 3. The horizontal lines are overlayed at ±25°.

higher frequency energy at and above 25°, suggesting that there is a significant increase in locally generated wave noise during the storm.

C. Frequency domain normalization

Frequency domain normalization has the dual purpose of broadening the signal bandwidth by placing higher emphasis on low amplitude signals, and of decreasing the negative impact of discrete sources. For the purpose of comparing frequency domain normalization methods, the effects of high amplitude temporal peaks, such as those observed around 4:30Z and 9Z in the low frequency spectrogram and coherogram, were minimized by setting all values of amplitude greater than 50% of the signal standard deviation to this value,[7,30] a process described as "threshold clipping," and explained further in Sec. IV D.

1. Normalization methods

OAI was performed using five different frequency domain preprocessing methods:

(a) no frequency domain preprocessing;
(b) bandpass filter, no frequency domain amplitude normalization;
(c) bandpass filter and whiten by normalizing over the entire frequency range (20–100 Hz), known as absolute whitening;
(d) bandpass filter and normalize by a smoothed version of the amplitude spectrum, known as smoothed whitening; and
(e) bandpass filter and partially normalize the data by the sum of the signal magnitude at that frequency and a mean amplitude dependent constant:

$$S(\omega) = \frac{S(\omega)}{|S(\omega)| + \beta\overline{|S|}}, \tag{4}$$

where $\overline{|S|}$ is the mean amplitude over the entire frequency range, and β determines the degree to which the data are whitened [$\beta=0$ is equivalent to absolute whitening (c), and $\beta=\infty$ to no normalization (b)].

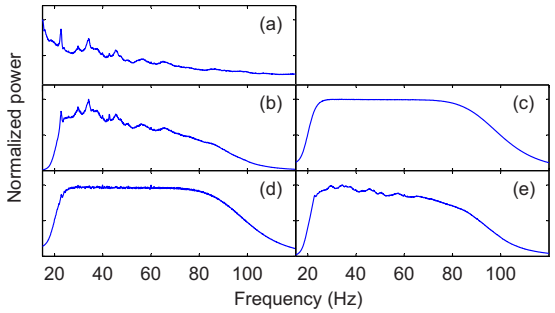


FIG. 5. (Color online) Normalized (linear) spectra of the September 2 signals recorded on H-40 (a) before prefiltering and [(b)–(e)] after prefiltering. Prefiltering methods are (b) bandpass and time domain filters only, (c) absolute whitening, (d) smoothed whitening, and (e) partial whitening ($\beta=1$).

The effect of applying each normalization technique to H-40 data can be seen in Fig. 5. Bandpass filtering without frequency normalization, method (b), maintains the general characteristics of the amplitude peaks and decay with frequency seen in the raw data, method (a). Absolute and smoothed whitening, methods (c) and (d), respectively, give approximately equal energy across the frequency band. Partial whitening, method (e), reduces extraneous tonals, but also places emphasis on signals of higher coherence (lower frequency).

2. Application to data

Data were preprocessed using each of the methods outlined in Sec. IV C 1. Individual cross-correlations were calculated and normalized by their peak value before summing so that the overall cross-correlation is not dominated by high amplitude cross-correlations from only part of the day. The cross-correlations between H-52 (tail-end HLA hydrophone) and all other hydrophones, summed over September 2, using smoothed-whitening filtering, method (d), are shown in Fig. 6(a).

The HLA cross-correlations are plotted as a function of distance from H-52. The VLA cross-correlations, which are offset by the horizontal distance between H-52 and the VLA hydrophones, are plotted as a function of height from the seafloor (note that the two vertical axes have different scales). The direct (D), surface-reflected (S), and surface-bottom-reflected (B, VLA cross-correlations only) travel times between each hydrophone, which are shown as dotted lines, were determined using OASES.[31]

Peaks in the cross-correlation are evident at both the direct and surface-reflected travel times. The EGF envelope in Fig. 6(b), on a logarithmic scale, reveals the surface-bottom reflected path to the lower VLA hydrophones.

The EGF envelope for the case of bandpass filtering only, method (b), shown in Fig. 6(c), shows only minor differences to that for smoothed whitening. Due to the higher proportion of low frequency energy, the arrivals are less sharp and the background noise level is slightly higher. In addition, the surface-reflected path is not as clear at the closer hydrophones (40–120 m). The raw signals have greater amplitude at lower frequency and this naturally assists the EGF when no frequency domain normalization is applied; the lower coherence signals have lower amplitude

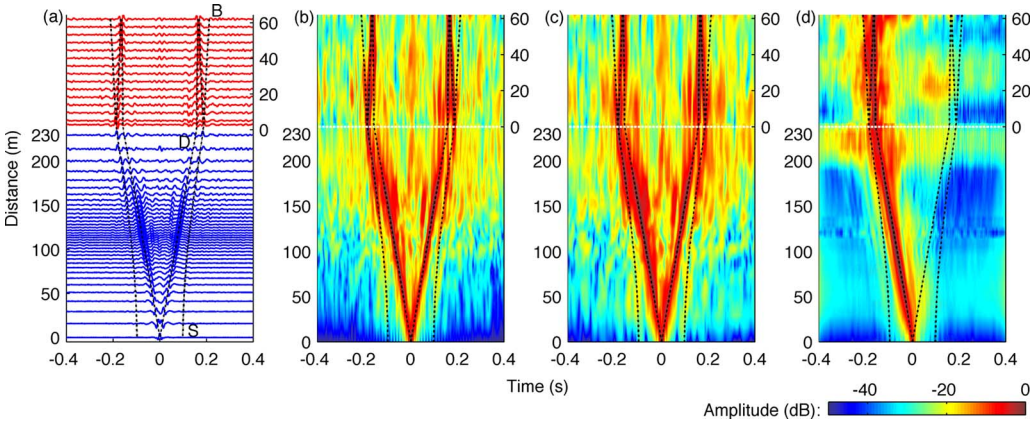


FIG. 6. (Color online) (a) Cross correlations between H-52 and all other hydrophones for September 2 data using smoothed-whitening frequency filtering (20–100 Hz bandwidth). [(b)–(d)] EGF envelope (dB relative to maximum value): (b) with smoothed whitening (20–100 Hz bandwidth), (c) with no frequency normalization (20–100 Hz bandwidth), and (d) with no frequency domain filtering or normalization. The lower traces are from EGFs with HLA hydrophones; their distances from the tail hydrophone (H-52) are shown on the left side axis. The upper traces are EGFs with VLA hydrophones; their vertical distance from the seafloor is shown on the right side axis, which is offset by the horizontal distance of the VLA from the HLA tail. The simulated travel times between the hydrophones were calculated using OASES and are overlaid as dotted lines.

and will therefore have less overall influence on the cross-correlated signal. This explains why a reasonable EGF can be determined when no spectral normalization is performed. The EGF envelope for no frequency domain filtering, method (a), shown in Fig. 6(d), gives a poor representation of the Green's function. A low frequency signal below 20 Hz from the southeast dominates the EGF to such an extent that only the direct acausal path is obtained. The EGF envelopes for absolute and partial whitening, methods (c) and (e), are not shown, but their characteristics lie between that of Figs. 6(b) and 6(c).

Smoothed whitening was selected as the optimal frequency domain filtering method for the data collected. Bensen *et al.*[12] compared no normalization and smoothed whitening for cross-correlations of seismic data and found that the improvements gained by normalization were substantially greater than here. Appropriate selection of the data bandwidth affected the result more than any frequency domain normalization because above the 100 Hz low-pass frequency the cross-correlation has almost no coherence, and therefore inclusion of higher frequencies adds to the noise floor. If no frequency domain filtering or normalization is applied this added noise is minimal, since the amplitude is negligible at higher frequencies. However, if the data are whitened but not bandpass filtered, signals of low coherence will be emphasized, and the resulting cross-correlation sum will be dominated by noise that requires very long averaging times to remove.

D. Time domain normalization

Various methods of time domain normalization have been used by others. One-bit time reversal normalization, where the sign (phase) of the waveform is retained but the amplitude is discarded, yields a higher signal-to-noise ratio than classical time reversal in some multiple scattering media.[32,33] A similar argument holds for cross-correlation analysis and therefore one-bit normalization is frequently used.[16,32,33] Another method of time domain normalization is to clip all signals above a certain threshold.[7] This minimizes the effect of energy bursts, but also retains more information than one-bit normalization. Gerstoft *et al.*[30] set their threshold as the minimum of the standard deviations measured over each day. For their data set this gave identical results to one-bit normalization. Bensen *et al.*[12] and Yang *et al.*[34] used temporally variable weighting functions. They claimed that these retain more small amplitude information and also allow for flexibility in defining the amplitude normalization in particular frequency bands.

1. Normalization methods

Six different time domain preprocessing methods and their applicability to 20–100 Hz ocean noise are compared here:

(a) no normalization;
(b) cross-correlate over short intervals with some degree of overlap, normalize the cross-correlations and then sum;
(c) clip the signal to a threshold;
(d) one-bit normalization;
(e) use of a rectangular central temporally variable weighting (RCTVW) function; and
(f) use of an exponential central temporally variable weighting (ECTVW) function.

Performing no normalization in the time domain sets a clear benchmark for the five other techniques. Cross-correlating over short intervals and then summing the normalized cross-correlations is more effective for shorter intervals. Since the greatest distance between any two hydrophones is 230 m, the direct path should be observable in under 0.2 s; hence, to ensure sufficient time for reverberant paths to be captured, 0.4 s data segments were used, with 33% overlap.

A threshold of σ, one standard deviation, was chosen as the level to which the signal would be clipped for normalization technique (c). It was noted that the results were not

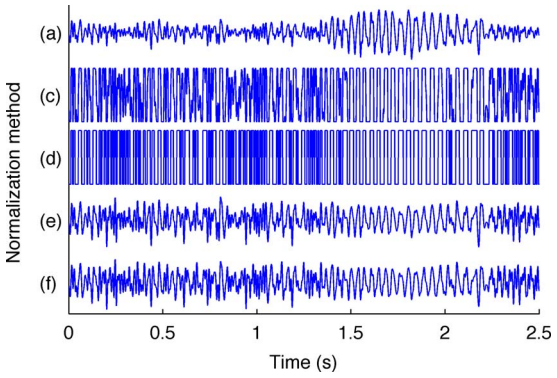


FIG. 7. (Color online) Preprocessed waveforms for 2.5 s of 20–100 Hz bandpass filtered data from H-40 (at 12:48:45Z) with normalization method: (a) none, (c) threshold clipping, (d) 1 bit, (e) RCTVW, and (f) ECTVW.

highly sensitive to the chosen threshold. Mathematical descriptions of one-bit normalization, RCTVW, and ECTVW are included in the Appendix.

2. Application to data

Example waveforms resulting from application of each time-normalization method to 2.5 s of H-40 data are shown in Fig. 7. Higher energies are observed in the time period 1.4–2.1 s, as can be seen in Fig. 7(a), and these amplitudes are all successfully reduced by the time-filtering methods, as shown in Figs. 7(c)–7(e). Normalization technique (b) is not shown here as this normalization is only applied after cross-correlating the data.

EGF envelopes for September 2 for each time-normalization method are shown in Figs. 8(a) and 8(b). The same line style has been used in the figure for all results because they are too similar to be individually discerned. The horizontal distance between (a) the HLA hydrophones H-52 and H-48 is 31.31 m, and the horizontal distance between (b)

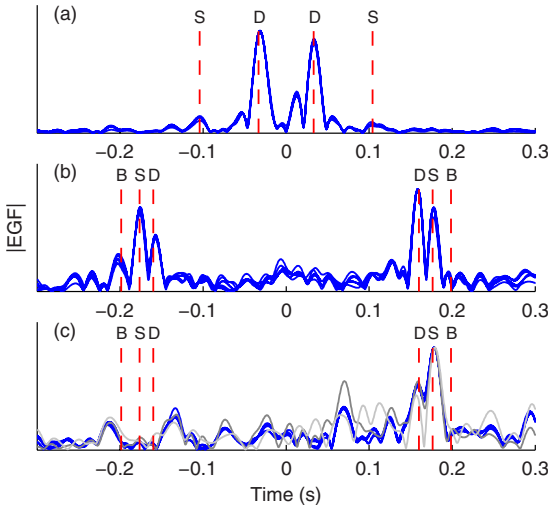


FIG. 8. (Color online) EGF envelopes, |EGF|, for all time-normalization methods for H-52 and (a) H-48 (entire day with 51.32 m horizontal separation), (b) H-8 (entire day with 230 m horizontal separation), and (c) H-8 (10:24 min from 8:30Z). Simulated travel times of direct (D), surface (S), and surface-bottom (B) paths are shown as vertical dashed lines. In (c) results for no normalization and short interval EGFs are shown in dark and light gray, respectively.

the HLA hydrophones H-52 and H-8 (located in the VLA) is 230 m. For the two closely spaced hydrophones shown in Fig. 8(a), large EGF peaks exist at the direct ray travel time, and smaller peaks at the surface-reflected travel time. The background noise is consistently low, except for one high peak at a time just less than the positive direct arrival. This could be due to a nonuniform source distribution. For the two further spaced hydrophones shown in Fig. 8(b), the EGF envelope once again peaks at the direct and surface-reflection travel times. A smaller peak can also be seen at the acausal surface-bottom travel time (i.e., the bottom-surface-reflected path from H-8 to H-52). The signal-to-noise ratio is poorer than for the more closely spaced hydrophones, but this is to be expected since decay and spreading of signals increase with distance.

The results from Figs. 8(a) and 8(b) suggest that time normalization has little influence on EGFs for this data set. Time normalization is important for seismic cross-correlations[8,12] since otherwise the results can be dominated by earthquakes. Although the ocean noise field is not perfectly diffuse, there are no equivalently energetic events for the frequency band considered, and nearby shipping is minimal on September 2. This and the intrinsic averaging introduced by summing over the entire day are two reasons why time domain normalization shows negligible benefit here.

If OAI were carried out over a time period insufficient to average out energetic events, the benefits of normalization would be greater. Consider the 10:14 min EGFs between H-52 and H-8 in Fig. 8(c). The EGFs peak at the positive direct and surface-reflected travel times only, indicating that the dominant sound field is from the tail end of the array (the northwest direction). Distinct peaks seen at −0.22, −0.14, and 0.07 s are the result of discrete sources. Since high amplitude events, which are reduced in the normalization process, are not averaged out in the shorter cross-correlation time period, the EGF envelope without normalization, method (a), and EGFs from cross-correlating over short periods and summing the normalized results, method (b), both have a higher noise level than the results for data that are normalized before cross-correlation.

Since time domain normalization techniques (c)–(f) all give similar results, and one-bit normalization is the least computationally intensive, it was selected for further processing and analysis of the data.

The EGFs between H-52 and all other hydrophones for 20–100 Hz bandpassed, one-bit normalized, smoothed whitened September 2 data are shown in Fig. 9(a). The Green's function, which was simulated using OASES, is shown convolved with a 20–100 Hz box car pulse in Fig. 9(b) for comparison purposes. The assumed model sediment density, $\rho = 1.69\ \mathrm{g/cm^3}$, was approximated from grab samples in the array vicinity,[35] and a sediment sound speed of $c = 1650\ \mathrm{m/s}$, estimated from the critical angle suggested by the VLA beamformer, was also assumed. Note that exact bottom properties are not critical as the simulations are only used here to calculate travel times, not amplitudes, of ocean-only paths. The direct arrival peaks are positive and the reflected arrival peaks are negative, which is due to the phase

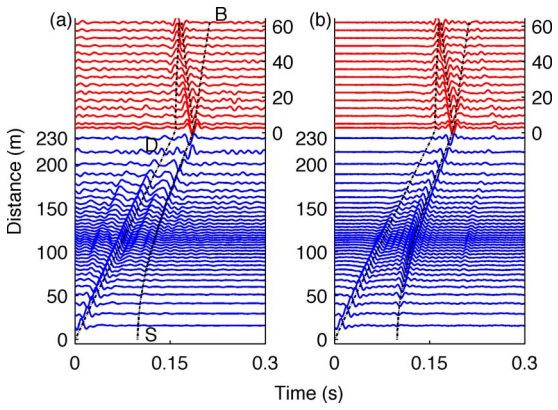


FIG. 9. (Color online) (a) EGFs between H-52 and all other hydrophones for September 2, with simulated travel times of direct (D), surface (S), and surface-bottom (B) paths shown as dotted lines. (b) Simulated Green's functions convolved with a 20–100 Hz bandwidth linear source. Vertical axis format is the same as in Fig. 6.

change at the surface. The amplitudes are not exact, though this is expected since the source field is not diffuse. The surface-reflected arrivals are apparent in the EGFs for distances greater than about 50 m. They are not observable at closer ranges, where they would be more steep, because the vertically propagating noise is weaker.

If a cross-correlation is started or finished part way through a ship's track, the EGF may be biased. Tapering of the cross-correlation amplitudes toward the start and end of the cross-correlation was therefore considered; however, for the given data set and long cross-correlation times, tapering was seen to have negligible effect.

V. GEOMETRIC COMPARISONS

Examples of EGF envelopes with respect to hydrophones other than the outermost HLA hydrophone, H-52, are shown in Fig. 10. Due to the steeper grazing angle to the furthest hydrophone, EGFs with H-34, a central HLA hydrophone, shown in Fig. 10(a), do not yield as much information about the surface-reflected path as do EGFs with H-52, shown in Fig. 9(d). Figure 10(b) reveals that EGFs with VLA H-10 show the surface-reflected path, at slightly larger times than the dominant direct path, for distances of 0–150 m

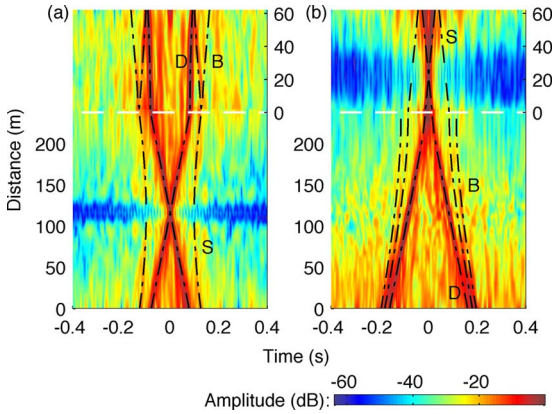


FIG. 10. (Color online) September 2 EGF envelope (dB relative to maximum amplitude) for SWAMI52 with respect to (a) H-34 and (b) H-10. Vertical axes are the same as in Fig. 6.

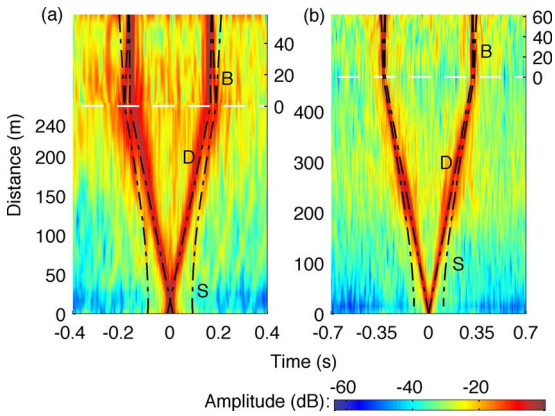


FIG. 11. (Color online) September 2 EGF envelopes (dB relative to maximum amplitude) for (a) SWAMI32, with respect to H-30, and (b) Shark, with respect to H-16. Vertical axis format is the same as in Fig. 6.

from the tail end of the HLA; however, the surface path is not as clear as that obtained when cross-correlating with H-52, which is likely due to either the decreased stability in the environment at the shallower depth of H-10, or the increased motion of the VLA hydrophones relative to the HLA hydrophones. The bottom-surface-reflected arrival from H-10 to the HLA hydrophones is also observable at a time just after the surface-reflected path, but the acausal path from the HLA to H-10 is not observable.

The September 2 EGF envelopes for SWAMI32 and Shark are shown in Fig. 11. The SWAMI32 EGFs are with respect to H-30 rather than the tail hydrophone due to high noise on the outer two hydrophones. The high noise levels are attributed to channel switching.[36] Like the SWAMI52 results shown in Fig. 9(d), the SWAMI32 and Shark array EGFs show both the direct and surface-reflected paths. The results in Figs. 9–11 show that, for all arrays, although the direct path dominates for more closely spaced hydrophones, the relative amplitude of the surface-reflected path increases at greater distances. These relative amplitudes depend on array geometry, modal distribution of acoustic energy, roughness at the surface, and, importantly, the impedance at the seafloor. As such, a relationship between the relative amplitudes of the paths and the critical angle could potentially be determined.[29]

Unlike the tapered spacing of the SWAMI52 and SWAMI32 HLA hydrophones, the Shark HLA hydrophones are evenly spaced at 15 m intervals. The September 2 EGFs between all HLA pairs separated by 345 m are plotted in Fig. 12(a). The traces are similar, and all display EGF peaks at approximately ± 0.24 s. The median value of the signals is plotted against a shaded area encompassing the range of all signal values in Fig. 12(b). A magnified view of part of the signal is provided in Fig. 12(c) and shows that the signal variation is minimum near the direct path travel time.

VI. TEMPORAL VARIATIONS

The September 2 data EGFs with H-52 were compared with those from adjacent days, and are shown in Fig. 13. The September 2 data, shown in Fig. 13(c), peak at the expected direct and surface-reflected paths and exhibit the least back-

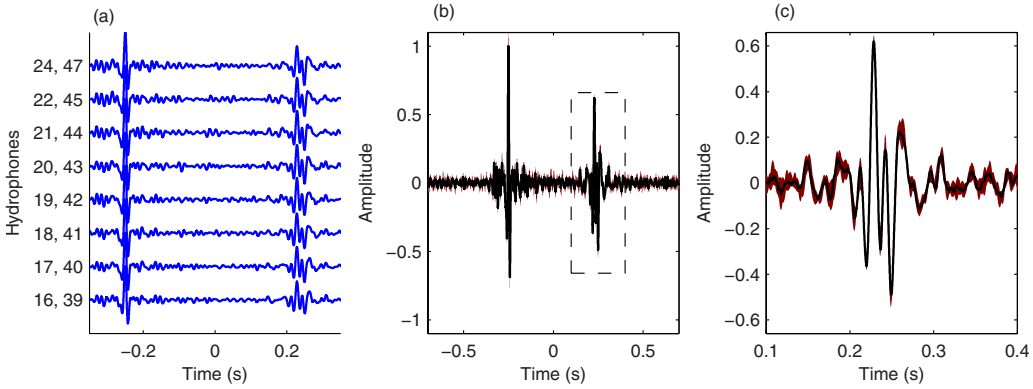


FIG. 12. (Color online) (a) September 2 EGFs for Shark hydrophone pairs separated by 345 m. (b) The median EGF from (a) overlies a shaded region between which all values lie. (c) Data from the dashed box in (b) are magnified.

ground noise. Results from September 1, shown in Fig. 13(b), are not as clear as those from September 2, but are still better than from September 3 and August 31, shown in Figs. 13(d) and 13(a), respectively, suggesting that the sound field is more diffuse during the storm.

Short time EGFs were calculated for data from pairs of hydrophones for all three arrays from 0Z August 31 to 12Z September 3. SWAMI52 hydrophones H-52 and H-17, SWAMI32 hydrophones H-30 and H-15, and Shark hydrophones H-16 and H-35 were chosen as their separation distances are all similar (between 200 and 285 m). Time segments corresponding to one data file were used for SWAMI52 and SWAMI32 (10:14 and 6:24 min, respectively), and quarter files (8:34 min) were used for Shark. The corresponding EGF envelopes are plotted as a function of time in Figs. 14(a)–14(c), along with the EGF envelope of the summed normalized cross-correlations over the 84 h period.

The EGF envelope is dominated by discrete sources, as indicated in Figs. 14(a)–14(c) by the high amplitude peaks that occur throughout the days at times less than the direct interhydrophone travel times. Hydrophone spectrograms from times corresponding to the largest peaks are dominated by a banded structure indicative of ship noise. As an example, spectrograms of 60 s duration from 3:36:40Z August 31 for SWAMI52 H-52 and SWAMI32 H-30 are shown in Figs. 14(d) and 14(e). Noise from a large ship, with a primary tonal at just under 40 Hz, dominates both spectrograms. The ship is visible as a peak in the EGF envelope for all three arrays from 0 to 4Z August 31. It was ascertained from the time of the EGF envelope peak that during this period the ship moved from southwest of the arrays to north of the arrays. The peak in EGF time due to a ship changes with the ship's azimuth to the array, with peak times approaching the interhydrophone travel time as the ship approaches end-fire. Hence, the signals from a ship are apparent as curves when plotted as a function of experimental and correlation times. The "pattern" of curves that is visible in Figs. 14(a)–14(c) is therefore due to a multitude of ship tracks. Most shipping occurred along the coast and this is apparent from the greater proportion of ship tracks visible at positive travel times in Figs. 14(a) and 14(b), corresponding to NW and SW directions for SWAMI32 and SWAMI52, respectively. The Shark array was parallel to the coast, and as such, the ship tracks in Fig. 14(c) do not appear to have a preferred direction.

Toward the end of September 1 and on September 2 the EGF envelope is more stable, as observed by the main peaks in the EGF being more consistently closer to the dashed interhydrophone travel times and also by the amplitude and number of smaller peaks in the EGF being reduced. Fewer shipping tracks are seen, and faint arrivals are observable at the interhydrophone travel times. This is during the period of high wind [see Fig. 1(e)] and sea conditions from Tropical Storm Ernesto. The reduction in number of nearby ships and the increase in wave energy result in a greater proportion of

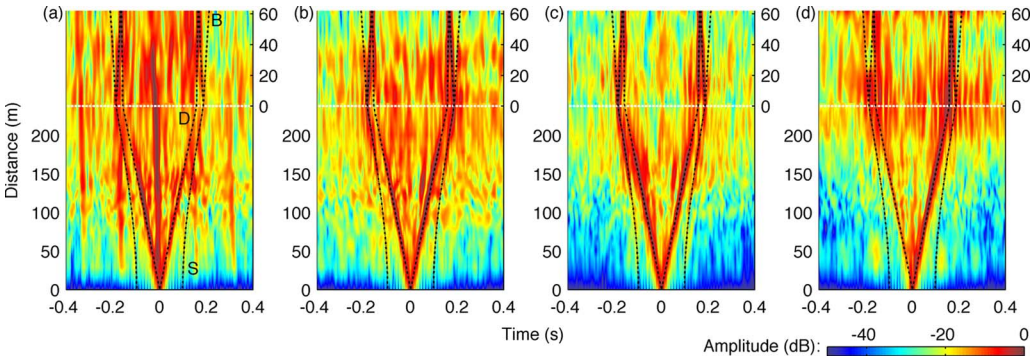


FIG. 13. (Color online) EGF envelope (dB relative to maximum amplitude) with respect to H-52 for (a) August 31, (b) September 1, (c) September 2, and (d) the first 12 h of September 3. Vertical axes are the same as in Fig. 6.

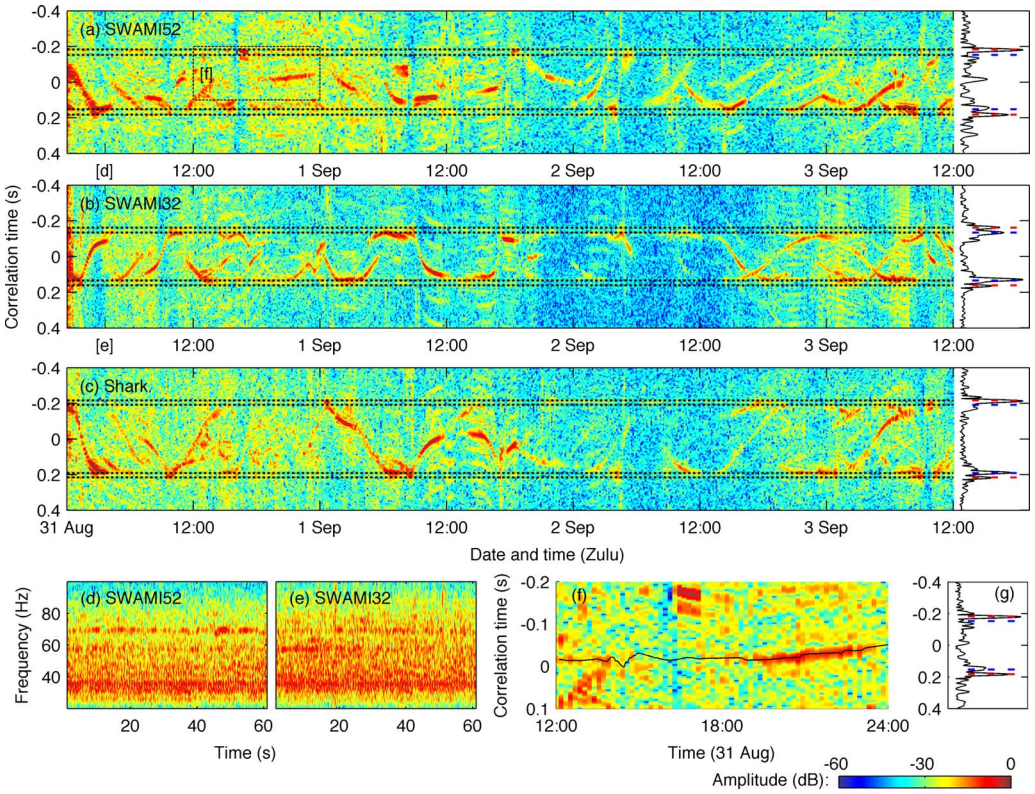


FIG. 14. (Color online) EGF envelope (dB relative to maximum amplitude) plotted for (a) SWAMI52 H-52 and H-17 (230 m separation), (b) SWAMI32 H-30 and H-15 (200 m separation), and (c) Shark H-16 and H-35 (285 m separation). Simulated direct and surface-reflected travel times (dashed lines) faintly overlayed. The envelope of the time derivative of the sum of all cross-correlations (normalized by their peak amplitudes to minimize the effects of dominant signals) is shown at the right of each plot. [(d)–(e)] 20–100 Hz spectrograms from 3:36:40Z August 31 for SWAMI52 H-52 and SWAMI32 H-30 [times denoted on (a) and (b) time axes as "d" and "e"], respectively. (f) Enlarged view of SWAMI52 EGF envelope [boxed area from (a)] showing a dominant near-side signal, with calculated travel time difference (black line) from R/V *Oceanus* to the hydrophone pair. (g) The envelope of the time derivative of the sum of all cross-correlations, excluding the period 12–24Z August 31 for SWAMI52 data.

acoustic energy in the ocean at these lower frequencies being from breaking waves and cumulative noise from distant shipping, and therefore the noise field is more diffuse. During this period faint peaks are frequently observed at times corresponding to the simulated surface-reflection travel times, such as between 22Z September 1 and 3Z September 2 in the acausal signal of Fig. 14(c).

Although the short term EGF envelope rarely yields the modeled interhydrophone travel time, based on the measured sound speed profile, throughout the 84 h period, the EGF envelopes of the summed cross-correlations for this period do peak at times near the simulated travel times, as shown at the far right of Figs. 14(a)–14(c). The surface-reflected path is particularly strong. This is because the EGF is dominated by nearby ships, and these shorter ranges favor higher acoustic grazing angles.

A strong signal is observed at a correlation time of slightly less than zero for all EGFs for August 31 in Fig. 13(a), suggesting that there is a high amplitude signal from near broadside (either SW or NE) of the array during that day. A corresponding peak in the summed SWAMI52 EGF envelope is seen at −0.0175 s at the far right of Fig. 14(a). The EGFs reveal that this peak is a result of signals from 12–14Z and 18–24Z on August 31 [see box "f" and Fig. 14(f)]. The Shark and SWAMI32 EGF envelopes do not show a strong signal at these times. Hence the dominant signal seen in the SWAMI52 data is likely from a source significantly closer to that array than the others. R/V *Oceanus* was located NE of SWAMI52 (in the region 39.25–39.28°N, 72.8–72.9°W) from 12 to 24Z August 31, about 10 km away. This is the closest that R/V *Oceanus* came to any of the arrays during the experiment. R/V *Oceanus* moved slowly in the experimental area and as such is an unusual ship noise source. The expected difference in travel time from this near-broadside location to SWAMI52 matches the short time EGF envelope peaks, as can be seen in Fig. 14(f). Thus, the high amplitude spurious signals in Fig. 13(a) are attributed to R/V *Oceanus*. The amplitude of the anomalous −0.0175 s peak in the EGF envelope of the summed normalized cross-correlations shown in the far right of Fig. 13(a) decreases to the background noise level when the period 12–24Z August 31 is excluded, as can be seen in Fig. 13(g).

Figures 13 and 14 suggest that the observation time period to obtain a stable EGF envelope depends on the distribution of the noise. Summing over September 2 yields a good approximation, as shown in Fig. 14(c), but summing over any of the other days or even summing over the entire 84 h period gives poorer results due to the increased proportion of directional bias of dominant events in the total received signal. Hence, when specific events dominate the

EGFs, either data from the times during which they occur should be discarded, or the cross-correlations need to be summed over an even longer period so that the effects of individual events are negligible. At any given time, except during the storm, the cross-correlation is generally dominated by one or two high amplitude events, and eliminating these from the data is difficult. For the case considered here with ship noise near the coast being a significant component of the noise field, the cross-correlations summed over many days or longer could show some directionality, corresponding to preferred shipping routes.

VII. CONCLUSION

Shallow water OAI in the ship dominated 20–100 Hz frequency band was considered using data from the Shallow Water 2006 experiment. Theory indicates that the time derivative of the cross-correlation yields an EGF, an approximation of an amplitude shaded Green's function. For an appropriate bandwidth, different time and frequency domain normalization methods yielded similar cross-correlation results. A major reason for this is the spatial averaging of the noise field, which occurs when noise from many ship tracks are recorded. Since ship noise is discrete, long cross-correlation periods were required to give sufficient averaging for the emergence of the Green's function. EGFs were here computed over 1 day, but shorter observation times could potentially be used. The EGFs are therefore average rather than instantaneous Green's functions.

Most ambient noise processing has focused on extracting the direct arrival, but the careful processing combined with the strong noise here allowed for extraction of higher order arrivals. Direct and surface-reflected paths between HLA hydrophones, as well as bottom-surface-reflected paths between HLA and VLA hydrophones, were determined from the EGF for three L-shaped arrays, in agreement with simulated travel times. The EGFs between equispaced HLA hydrophone pairs in a linear array are shown to have minimal variation. Analysis of temporal variations in the EGFs for horizontally propagating noise is generally dominated by one or two sources. The richer angular distribution of the breaking wave noise enabled construction of more vertically traveling paths. The EGFs obtained from data recorded during Tropical Storm Ernesto were clearer than those obtained before and after the storm.

The work here has focused on extracting high-quality arrivals from noise. These can potentially be used for array element localization,[6,36] ocean acoustic monitoring, and estimating sediment structure.

ACKNOWLEDGMENTS

Work supported by the Office of Naval Research under Grant No. N00014-05-1-0264, and by the Department of Energy National Energy Technology Laboratory via the Gulf of Mexico Hydrates Research Consortium, University of Mississippi. SWAMI data supplied by the Applied Research Laboratories, the University of Texas at Austin. Shark data supplied by Woods Hole Oceanographic Institute. L.A.B. is appreciative of support from a Fulbright Postgraduate Award in Science and Engineering, and from the Defence Science and Technology Organisation, Australia.

APPENDIX: TIME DOMAIN NORMALIZATION METHODS

This appendix outlines the mathematical details of three of the time domain normalization methods that were compared in Sec. IV D.

One-bit normalization, which uses only the sign of the signal, increases the signal-to-noise ratio of the data:

$$s_n(t) = \begin{cases} -1 & \text{if } s(t) < 0 \\ 1 & \text{if } s(t) > 0, \end{cases} \tag{A1}$$

where $s(t)$ is the raw signal at time t, and subscript n denotes the normalized signal.

RCTVW and ECTVW are the most computationally intensive time domain normalization techniques considered here. RCTVW normalizes each point by the sum of the unweighted mean of the absolute value of N preceding and succeeding values ($2N+1$ points overall):

$$s_n(t) = \frac{s(t)}{\omega(t)}, \tag{A2}$$

where

$$\omega(t) = \sum_{\tau=t-N}^{t+N} |s(\tau)| = \omega(t-1) - |s(t-N-1)| + |s(t+N)|. \tag{A3}$$

A normalization window of 0.05 s, the time interval of the maximum period, corresponding to the minimum frequency of 20 Hz, was found to be suitable. A normalization vector length of $2N+1=257$ was therefore used.

ECTVW places more emphasis on points closer to the point of interest. It normalizes in the same manner as RCTVW, the only difference being that it applies a weighting filter with an amplitude that decreases exponentially in both directions from the data point of interest:

$$\omega(t) = (1-\alpha)^N |s(t-N)| + \cdots + (1-\alpha)|s(t-1)| + |s(t)| + (1-\alpha)|s(t+1)| + \cdots + (1-\alpha)^N |s(t+N)|, \tag{A4}$$

where $\alpha = 2/(N+1)$ is the exponential smoothing factor. In order to use previously calculated sums to determine subsequent weights, the exponential is split up into two parts, the increasing exponential prior to and including the current point, and the decreasing exponential after the current point. These are then summed to give the overall weighting.

[1] O. I. Lobkis and R. L. Weaver, "On the emergence of the Green's function in the correlations of a diffuse field," J. Acoust. Soc. Am. **110**, 3011–3017 (2001).
[2] R. L. Weaver and O. I. Lobkis, "Ultrasonics without a source: Thermal fluctuation correlations at MHz frequencies," Phys. Rev. Lett. **87**, 134301 (2001).
[3] K. van Wijk, "On estimating the impulse response between receivers in a controlled ultrasonic experiment," Geophysics **71**, SI79–SI84 (2006).
[4] M. Campillo and A. Paul, "Long-range correlations in the diffuse seismic coda," Science **299**, 547–549 (2003).
[5] R. Snieder, "Extracting the Green's function from the correlation of coda

waves: A derivation based on stationary phase," Phys. Rev. E **69**, 046610 (2004).
[6]N. M. Shapiro, M. Campillo, L. Stehly, and M. H. Ritzwoller, "High-resolution surface-wave tomography from ambient seismic noise," Science **307**, 1615–1618 (2005).
[7]K. G. Sabra, P. Gerstoft, P. Roux, W. A. Kuperman, and M. C. Fehler, "Surface wave tomography from microseisms in Southern California," Geophys. Res. Lett. **32**, L14311 (2005).
[8]K. G. Sabra, P. Gerstoft, P. Roux, and W. A. Kuperman, "Extracting time-domain Green's function estimates from ambient seismic noise," Geophys. Res. Lett. **32**, L03310 (2005).
[9]K. Wapenaar and J. Fokkema, "Green's function representations for seismic interferometry," Geophysics **71**, SI33–SI46 (2006).
[10]E. Larose, A. Khan, Y. Nakamura, and M. Campillo, "Lunar subsurface investigated from correlation of seismic noise," Geophys. Res. Lett. **32**, L16201 (2005).
[11]A. Curtis, P. Gerstoft, H. Sato, R. Snieder, and K. Wapenaar, "Seismic interferometry-turning noise into signal," The Leading Edge **25**, 1082–1092 (2006).
[12]G. D. Bensen, M. H. Ritzwoller, M. P. Barmin, A. L. Levshin, F. Lin, M. P. Moschetti, N. M. Shapiro, and Y. Yang, "Processing seismic ambient noise data to obtain reliable broad-band surface wave dispersion measurements," Geophys. J. Int. **169**, 1239–1260 (2007).
[13]P. Roux and M. Fink, "Green's function estimation using secondary sources in a shallow water environment," J. Acoust. Soc. Am. **113**, 1406–1416 (2003).
[14]P. Roux, W. A. Kuperman, and the NPAL Group, "Extracting coherent wave fronts from acoustic ambient noise in the ocean," J. Acoust. Soc. Am. **116**, 1995–2003 (2004).
[15]K. G. Sabra, P. Roux, and W. A. Kuperman, "Arrival-time structure of the time-averaged ambient noise cross-correlation function in an oceanic waveguide," J. Acoust. Soc. Am. **117**, 164–174 (2005).
[16]K. G. Sabra, P. Roux, A. M. Thode, G. D'Spain, W. S. Hodgkiss, and W. A. Kuperman, "Using ocean ambient noise for array self-localization and self-synchronization," IEEE J. Ocean. Eng. **30**, 338–347 (2005).
[17]M. Siderius, C. H. Harrison, and M. B. Porter, "A passive fathometer technique for imaging seabed layering using ambient noise," J. Acoust. Soc. Am. **120**, 1315–1323 (2006).
[18]L. A. Brooks and P. Gerstoft, "Ocean acoustic interferometry," J. Acoust. Soc. Am. **121**, 3377–3385 (2007).
[19]P. Gerstoft, W. S. Hodgkiss, M. Siderius, C.-F. Huang, and C. H. Harrison, "Passive fathometer processing," J. Acoust. Soc. Am. **123**, 1297–1305 (2008).
[20]C. H. Harrison and M. Siderius, "Bottom profiling by correlating beam-steered noise sequences," J. Acoust. Soc. Am. **123**, 1282–1296 (2008).
[21]P. Roux, K. G. Sabra, W. A. Kuperman, and A. Roux, "Ambient noise cross correlation in free space: Theoretical approach," J. Acoust. Soc. Am. **117**, 79–84 (2005).
[22]K. Wapenaar, "Green's function retrieval by cross-correlation in case of one-sided illumination," Geophys. Res. Lett. **33**, L19304 (2006).
[23]J. Traer, P. Gerstoft, P. D. Bromirski, W. S. Hodgkiss, and L. A. Brooks, "Shallow-water seismoacoustic noise generated by tropical storms Ernesto and Florence," J. Acoust. Soc. Am. **124**, EL170–EL176 (2008).
[24]R. J. Urick, *Principles of Underwater Sound* (McGraw-Hill, New York, 1975).
[25]R. Snieder, K. Wapenaar, and K. Larner, "Spurious multiples in seismic interferometry of primaries," Geophysics **71**, SI111–SI124 (2006).
[26]C. M. Bender and S. A. Orszag, *Advanced Mathematical Methods for Scientists and Engineers: Asymptotic Methods and Perturbation Theory* (McGraw-Hill, New York, 1978).
[27]A. Derode, E. Larose, M. Tanter, J. de Rosny, A. Tourin, M. Campillo, and M. Fink, "Recovering the Green's function from field-field correlations in an open scattering medium," J. Acoust. Soc. Am. **113**, 2973–2976 (2003).
[28]N. Harmon, P. Gerstoft, C. A. Rychert, G. A. Abers, and K. M. Fischer, "Phase velocities from seismic noise using beamforming and cross correlation in Costa Rica and Nicaragua," Geophys. Res. Lett. **35**, L19303 (2008).
[29]S. E. Fried, W. A. Kuperman, K. G. Sabra, and P. Roux, "Extracting the local Green's function on a horizontal array from ambient ocean noise," J. Acoust. Soc. Am. **124**, EL183–EL188 (2008).
[30]P. Gerstoft, K. G. Sabra, P. Roux, W. A. Kuperman, and M. C. Fehler, "Green's functions extraction and surface-wave tomography from microseisms in southern California," Geophysics **71**, SI23–SI31 (2006).
[31]H. Schmidt, *OASES Version 3.1 User Guide and Reference Manual*, Department of Ocean Engineering, Massachusetts Institute of Technology, 2004.
[32]A. Derode, A. Tourin, and M. Fink, "Ultrasonic pulse compression with one-bit time reversal through multiple scattering," J. Appl. Phys. **85**, 6343–6352 (1999).
[33]E. Larose, A. Derode, M. Campillo, and M. Fink, "Imaging from one-bit correlations of wideband diffuse wavefields," J. Appl. Phys. **95**, 8393–8399 (2004).
[34]Y. Yang, M. H. Ritzwoller, A. L. Levshin, and N. M. Shapiro, "Ambient noise Rayleigh wave tomography across Europe," Geophys. J. Int. **168**, 259–274 (2007).
[35]J. A. Goff, B. J. Kraft, L. A. Mayer, S. G. Schock, C. K. Sommerfield, H. C. Olson, S. P. S. Gulick, and S. Nordfjord, "Seabed characterization on the New Jersey middle and outer shelf: Correlatability and spatial variability of seafloor sediment properties," Mar. Geol. **209**, 147–172 (2004).
[36]L. A. Brooks, P. Gerstoft, and D. P. Knobles, "Multichannel array diagnosis using noise cross-correlation," J. Acoust. Soc. Am. **124**, EL203–EL209 (2008).

Range-dependent geoacoustic inversion of vertical line array data using matched beam processing

Kyungseop Kim, Woojae Seong,[a] and Keunhwa Lee
Department of Naval Architecture and Ocean Engineering, Seoul National University, Seoul 151-744, Korea

Seongil Kim
Agency for Defense Development, Jinhae 645-016, Korea

Taebo Shim
Soongsil University, Seoul 156-743, Korea

(Received 23 May 2008; revised 23 October 2008; accepted 2 December 2008)

This paper describes the results of range-dependent geoacoustic inversion using vertical line array data obtained from the 4th Matched Acoustic Properties and Localization Experiment conducted in the East Sea of Korea. The narrowband multitone continuous-wave signal from the towed source was analyzed to estimate the range-dependent geoacoustic properties along the radial track. The primary approach is based on the sectorwise inversion scheme. The inversion region up to 7.5 km from the vertical line array was divided into several segments, and the subinversions for each segment were performed sequentially. To reduce the dominance of low-angle arrivals, which bears little information for the bottom segment in question, matched beam processing with beam filtering was used for the cost function. The performance of proposed algorithm was tested using simulated data for an environment representative of the experimental site. The inversion results for the experimental data were consistent with the geophysical database and were validated from matched-field source localization using frequencies different from those used in the inversion.
 [DOI: 10.1121/1.3056551]

PACS number(s): 43.30.Pc, 43.60.Pt [RAS] Pages: 735–745

I. INTRODUCTION

Estimation of seabed properties is important in applying acoustic systems to shallow water because of bottom-interacting sound propagation. Until recently, various inversion techniques have been developed to estimate the geoacoustic bottom properties from remote sensing of the acoustic field. One of the most effective model-based techniques is the matched-field (MF) inversion method, which is generally based on sensor array signal processing. The MF inversion has been successfully applied to estimate bottom characteristics in various shallow water environments.[1–12]

Actual shallow water environments usually have range variability in seabed properties as well as in bathymetry or ocean sound speed. However, in conventional geoacoustic inversions using a vertical line array (VLA) and a single source, the geoacoustic properties have been typically assumed to be range independent. In this case, averaged outputs between source and receiver are obtained.[2,3] However, as the source-receiver range increases, spatial variability causes modeling errors, resulting in degraded inversion performance. To prevent such degradation in inversion modeling accuracy, research has been conducted on range-dependent geoacoustic inversion.[4–12] Towed horizontal line array (HLA) schemes are one such range-dependent inversion method. The towed HLA system has several advantages over a VLA, such as mobility and manageability, and provides simplicity of the inversion modeling based on assuming range independency in local regions between the source and receiver.[4,5] On the contrary, a VLA configuration is relatively stable and capable of receiving sound fields from all directions and can probe large areas around the VLA with a towed source or multiple sources.[6–11] Although there is some computational complexity in propagation modeling, VLA configurations are still effective in range-dependent inversion.

In this paper, a range-dependent geoacoustic inversion using a VLA and a moving source is proposed. In previous research for range-dependent inversion with VLA systems, usually only bathymetry variability was considered, or multiple sources and VLAs spatially distributed (tomographic inversion) were used. In our work, the variability of bottom properties in the radial direction defined by the towed source was estimated by a sectorwise inversion scheme with relatively long-range data. The basic strategy of this sectorwise scheme is similar to the tomographic inversion method suggested by Pignot and Chapman.[12] However, they used a time-domain inversion method based on the separation and filtering of signal arrivals using multiple broadband shot data within a short range of the VLA (<1 km), generally difficult to apply to long-range data.

Data used in this inversion are from the 4th Matched Acoustic Properties and Localization Experiment (MAPLE 4), one of MAPLE series[13] projects conducted by the Agency for Defense Development of Korea. The purpose of the experiment was to investigate the performance of matched-field geoacoustic inversion and source localization. A narrowband

[a]Electronic mail: wseong@snu.ac.kr

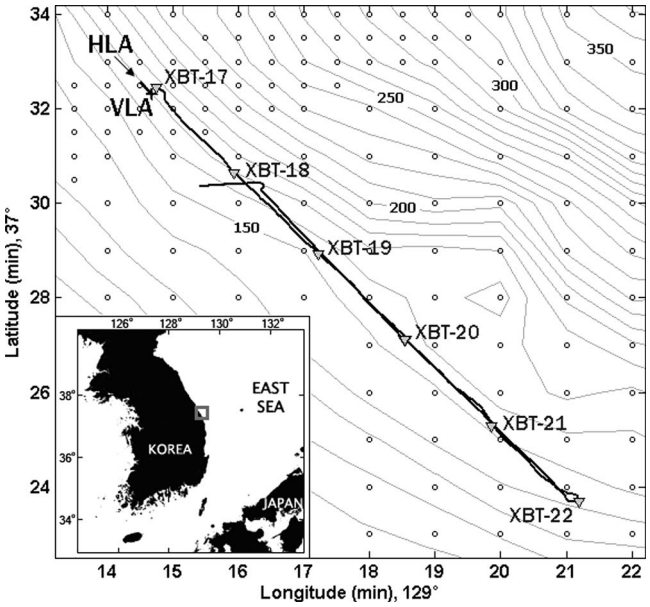


FIG. 1. Bathymetry (in meters) of the MAPLE 4 experimental site and onboard global positioning system (GPS) track of the R/V Sunjin. The triangles indicate the positions of XBT-measurements during the source towing, and the open circles denote coring positions for the geophysical database.

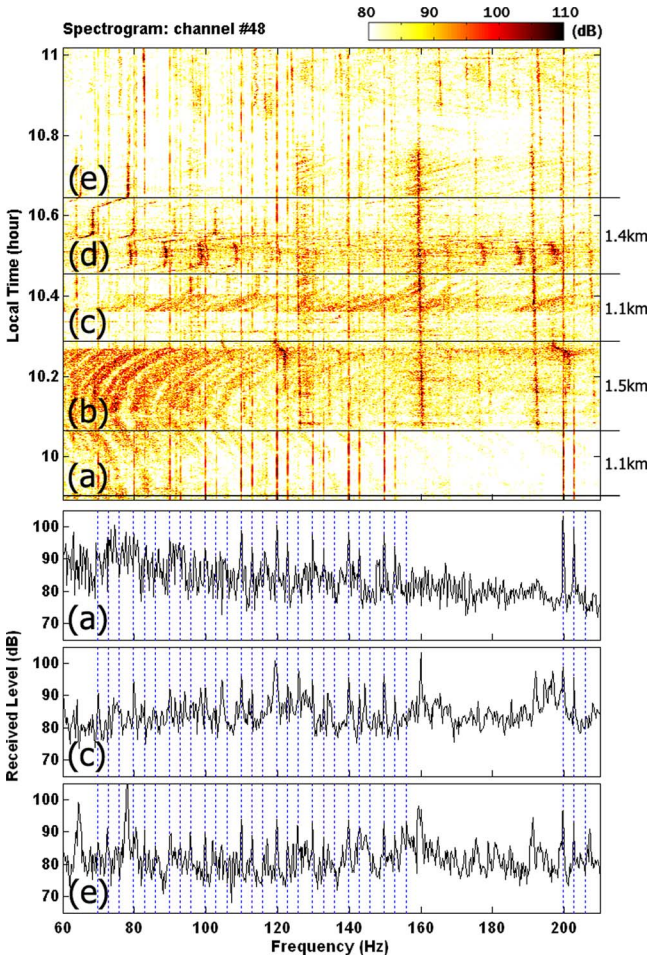


FIG. 2. (Color online) Frequency spectrogram for the initial 67 min duration of the received signal at the uppermost channel (ch. 48) (top panel), and snapshots of frequency spectrum (averaged over entire 48 channels of the VLA) at different selected ranges (lower three panels). The horizontal solid lines [(a)–(e)] in the top panel represent signal frames used in the inversion and vertical dotted lines in the lower panels represent the tones of source signal.

low-frequency acoustic source was towed along the radial track over a slightly inclined bottom with varying geophysical properties. The total length of the track was 18.5 km, and bottom properties in the initial 7.5 km of the track were inverted through the sequential subinversions for several segments. As the dominance of the uninformative portion of the received signal lowered parameter sensitivity in our sectorwise method, matched beam processing (MBP) was applied for the cost function to filter out such acoustic arrivals.[14–17]

Descriptions of the experiment, measured data, and geophysical database are presented in Sec. II. Inversion procedures including the basic scheme, beam filtering, and environmental modeling are described in Sec. III. Preliminary synthetic data simulation and the geoacoustic inversion results for experimental site are presented in Sec. IV. Lastly, conclusions are presented in Sec. V.

II. EXPERIMENT AND DATA

A. Geometric description

The experiment was carried out in May 2005, 10 km off the coast in the East Sea of Korea. An L-shaped receiving line array was deployed on the seabed at a depth of approximately 170 m. The array consists of two legs, one vertical and the other horizontal, lying on the seabed, with 48 hydrophones in each subarray. In the inversion, 48 channel data from the VLA were used. The VLA part has a 2.5 m hydrophone spacing and a 117.5 m aperture which spans the lower 70% of the water column. The low-frequency acoustic source was towed at a nominal depth of 50 m, by R/V Sunjin along the slightly inclined bathymetry track. During the towing, the ship maintained a constant speed of about 2 m/s and constant bearing aligned with the end-fire direction of the HLA, as shown in Fig. 1. The ship moved radially outward from the receiving array and returned along the same course. The bathymetry information was obtained from an echosounder of the R/V during the towing.

B. Acoustic and oceanographic data

The towed source transmitted narrowband multitone continuous-wave (cw) signals in outbound towing and narrowband pulse signals in inbound towing. In this paper, only the measurements of cw signals are considered. The cw tones have a total of 30 frequency components and can be divided into three groups. The tones of the first group at ten frequencies within the 70–200 Hz band are pilot signals with high source levels (150–160 dB). The second and third group tones, at 3 and 6 Hz higher frequencies, respectively, than the first group, have 5–20 dB lower source levels. A frequency spectrogram for the initial 67 min part of received signal at the uppermost channel (ch. 48), which covers a source-receiver range from 2.4 to 10.4 km, is given in the top panel of Fig. 2. Besides the transmitted signal, there are strong broadband noise signals and narrowband tones that are generated from ferry and cargo vessels passing near the receiv-

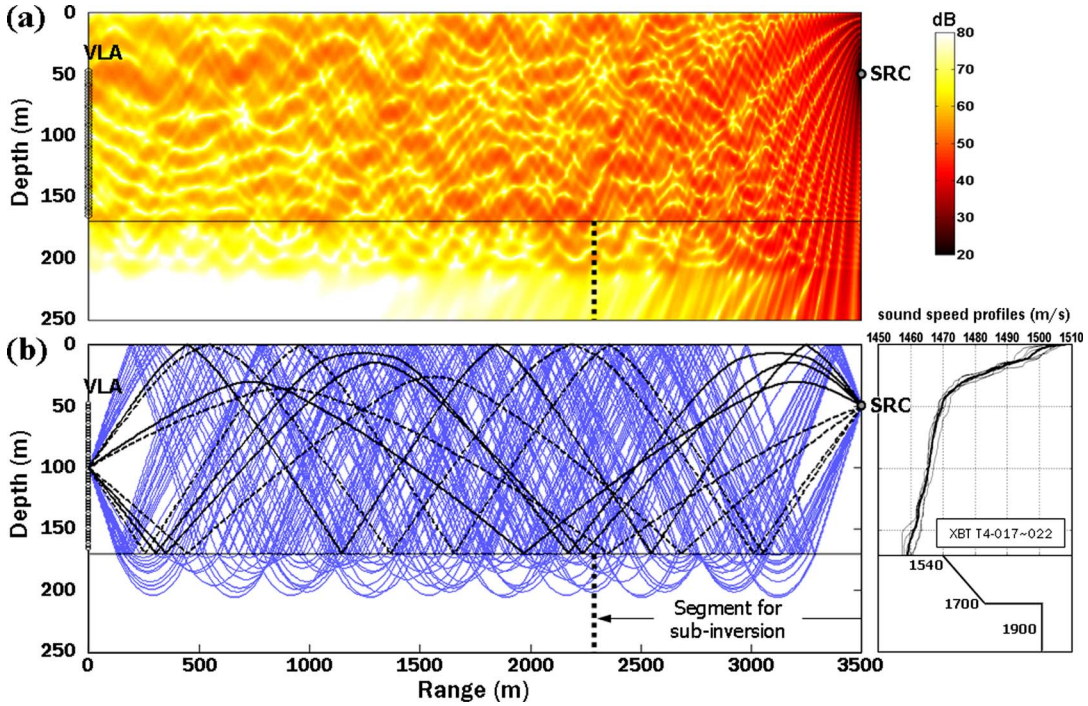


FIG. 3. (Color online) (a) 2D acoustic field (150 Hz) and (b) an example of significant eigenray trajectories in an environment representative of the experimental site. In the ray plot, the thick-solid and thick-dashed lines are upward and downward transmitted rays having initial grazing angle below 15°.

ing array. Three snapshots of frequency spectrum (averaged over entire 48 channels of the VLA) at different selected ranges are presented in the lower panels of Fig. 2. It shows that the emitted tones have high signal-to-noise ratio (SNR) except for relatively low-frequency tones below 100 Hz. The sampling frequency was 4 kHz, and 5s fast Fourier transforms were applied with no overlap. Due to the outgoing ship motion, a Doppler shift of about −0.2 Hz (frequency bin width is 0.2 Hz) was observed, and this frequency shift was reflected in the data processing. In this analysis, selected frequency tones at 110, 150, and 200 Hz of the first group were used for the inversion (the lower frequency data could not be used due to the contamination from surface interferences).

The water column sound speed profiles (SSPs) for the track were obtained using XBT and CTD measurements. The temperatures were measured from the 6 XBT casts during the source towing as in Fig. 1. A set of six profiles is given in the right panel of Fig. 3. A typical profile has a strong thermocline in the upper 40 m layer that generates downward refracting propagation. There is also significant variability between the measured profiles with a maximum difference of approximately 10 m/s in the thermocline. Because water column SSP generally has a large effect on sound propagation (especially for the downward refracting structure), this fluctuation of the profile makes the environmental modeling difficult for the inversion of experimental data from relatively long source-receiver ranges.

C. Geophysical database

Prior to the experiment, geophysical surveys were conducted in the East Sea by the Korea Institute of Geoscience and Mineral Resources (KIGAM). The survey contained bottom surface coring for a large area of 37°13′–38°3′N and 128°59′–129°55′E, including the site of this experiment. The coring positions around the source track are represented as open circles in Fig. 1 (1′-interval in longitude/latitude and 0.5′-interval near the array). Then, the geophysical properties of the sea bottom at each coring position were estimated by combining the results of high-resolution seismic profiles from the sub-bottom profiling (2–7 kHz sonar and air gun), shallow and deep cores (grab, piston, and portable remote operated drilling), and outcrop geology with field surveys.[18] The area near the receiving array consists of mud/sand-type thick sediment layer with a sound speed of about 1500–1700 m/s above the sub-bottom. Bottom properties also have range-dependent variability along the source track up to 12 km from the VLA.

III. INVERSION PROCEDURE

A. Inversion scheme

The objective of this work is to estimate the geoacoustic bottom properties in the range-dependent environment using the data received at the moored VLA. In this work, the parabolic equation method RAM[19] (SAGA-implemented version[20]) was used as a propagation model. The basic principle of the inversion scheme is sector-by-sector inversion for the propagation path along the radial source track. On the assumption that the inversion region can be divided into several segments, the entire inversion procedure consists of sequential subinversions for each segment, as illustrated in Fig. 4. First, the geoacoustic model parameters of the nearest segment from the receiving array are estimated using the initially received signal from the towed source. Then, fixing the

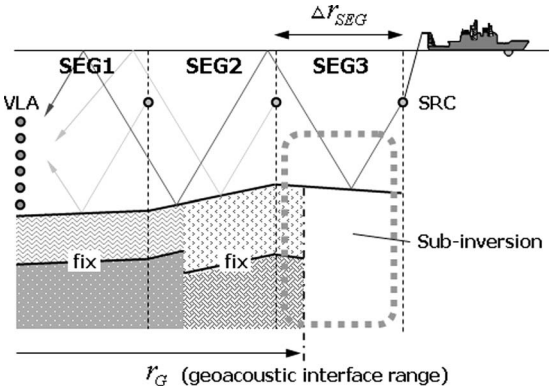


FIG. 4. Illustration of the basic scheme of the sectorwise inversion.

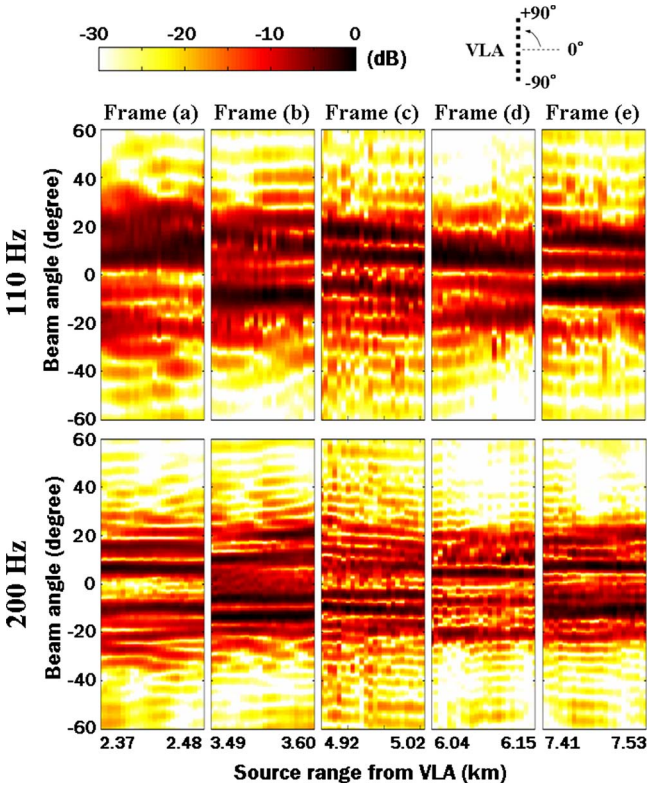


FIG. 5. (Color online) Vertical directionality of the received signal at the VLA for 110 and 200 Hz. In each panel, conventional beamforming was performed for 1 min data around the five signal frames shown in the top panel of Fig. 2.

estimated parameters in the first segment, the subinversion for the second segment is performed using data from the next signal frame. The same procedure is repeated for the following segments. In order to properly model an abrupt sediment property change, the geoacoustic interface range r_G is introduced as an additional inversion parameter, as indicated in Fig. 4. The estimated bottom parameters in the former segment are extended up until this interface. In the slightly and continuously varying seabed environment, this additional parameter may not be necessary. However, separation of the segment geometric boundary (i.e., position where the signal frame is used) and the geophysical interface turned out to improve the convergence of searching parameters.

B. Angle filtering of received signal

In this sectorwise inversion approach, the geoacoustic parameters of the latest segment (i.e., nearest to the source position) are estimated in each subinversion step. An example of two-dimensional (2D) acoustic field calculated by the PE and of ray traces in an environment representative of the experimental site is illustrated in Fig. 3. The 2D field plot shows that there are acoustic energies that reach the VLA without any interaction with the bottom in question. This can be seen more schematically in the ray-trace plot. In the figure, the thick-solid and thick-dashed lines are the upward and downward transmitted rays having initial grazing angles below 15°. The downward transmitted rays bounce directly off the segment for subinversion, while almost all the upward rays have no bottom bounce within this segment. Though these upward rays take up a large portion in the energy of received signals, they do not provide much information about the bottom properties of the inversion segment. Using the "full-field" data, which include these arrivals, causes a heavy loss in the sensitivity of the searching parameters

The conventionally beamformed outputs of the received cw signal at the VLA show vertical directionality (Fig. 5). In Fig. 5, beams are separated into several groups and low-angle arrivals can be distinguished from relatively steep angle arrivals (more clearly at 200 Hz). In this inversion, MBP was applied to filter out these uninformative low-angle arrivals. In relation to the signal filtering, mode filtering which is conceptually the same as beam filtering can also be considered. Beam filtering, however, is usually more effective in shallow water because mode decomposition is often difficult due to the limited aperture of the VLA.[15] Matched beam processor with beam filter for the inversion is given by[16]

$$B(\mathbf{\Phi}) = \frac{\left| \int_{|\theta|>|\Theta_0|} A^{\text{model}^*}(\theta,\mathbf{\Phi}) A^{\text{data}}(\theta-\theta_0) d\sin\theta \right|^2}{\left[\int_{|\theta|>|\Theta_0|} |A^{\text{model}}(\theta,\mathbf{\Phi})|^2 d\sin\theta \right] \left[\int_{|\theta|>|\Theta_0|} |A^{\text{data}}(\theta-\theta_0)|^2 d\sin\theta \right]}, \quad (1)$$

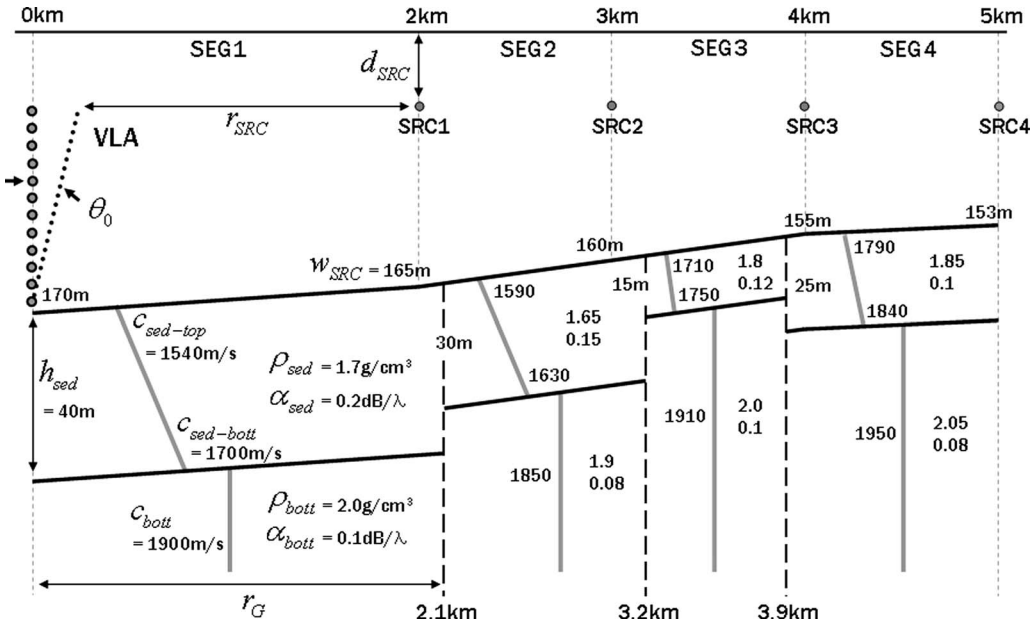


FIG. 6. Environment for the synthetic data and inversion parameters of one-layer geoacoustic model.

where $A^{\text{model}}(\theta,\mathbf{\Phi})$ and $A^{\text{data}}(\theta)$ are the beam outputs of replica and measured pressure field, $\mathbf{\Phi}$ is the inversion parameter vector, θ_0 is the beam angle shift corresponding to the array tilt, and Θ_0 is the cutoff angle for beam filtering. The cutoff angle can be determined from the analysis of the eigenray and the beam structure of the received signal. Too high cutoff angle excessively increases the sensitivity of parameters and lowers the array gain. In Eq. (1), if the length of the vertical array is sufficient, MBP with full beams can be reduced to conventional MFP.[15] The cost function for our matched beam inversion is an incoherently frequency-averaged and signal-frame-averaged (i.e., averaged over the signal frames at multiple ranges near the source position for each segment) beam processor given as follows:

$$E(\mathbf{\Phi}) = 1 - \frac{1}{N_s}\frac{1}{N_f}\sum_{j=1}^{N_s}\sum_{i=1}^{N_f} B_{ij}(\mathbf{\Phi}), \tag{2}$$

where, N_f is the number of frequencies and N_s is the number of signal frames. Using incoherent averaging reduces the instability of the cost function caused by filtered signal in addition to obtaining the broadband effect.

C. Environmental modeling

Based on previous results[21] from the experiments conducted near the MAPLE 4 experimental site, the geoacoustic model was determined as a one-sediment layer overlying a half-space sub-bottom. In the geoacoustic model, shown in Fig. 6, *p*-wave sound speed in the sediment layer has a constant gradient whereas density and attenuation are constant values. The half-space also has constant sound speed, density, and attenuation. The searching parameters are sediment thickness h_{sed}, sediment top and bottom sound speeds $c_{\text{sed-top}}$ and $c_{\text{sed-bott}}$, sediment density ρ_{sed}, sediment attenuation α_{sed}, sub-bottom sound speed c_{bott}, sub-bottom density ρ_{bott}, and sub-bottom attenuation α_{bott}. In addition, geometric parameters were included in searching parameters for inversion: source range r_{SRC}, source depth d_{SRC}, water depth at source position w_{SRC}, array tilt θ_0, and range of geophysical interface r_G.

The optimization of the cost function was performed through adaptive simplex simulated annealing,[22] which combines the strength of both simulated annealing and downhill-simplex method. This hybrid algorithm is simple and provides effective accommodation to correlated parameters and widely ranging parameter sensitivities. Based on the results of several tests conducted prior to the inversion, the annealing schedule and tuning parameters for the algorithm were adjusted. The annealing schedules for each inversion had 700–1000 temperature steps prior to quenching from the initial temperature of 0.3, and temperature reduction factor of 0.995, with five perturbations per temperature. The computational grid spacings for the parabolic equation model were dr=10 m and dz=0.25 m.

IV. INVERSION RESULTS

A. Synthetic data

The sequential inversion algorithm was tested using simulated data for an environment representative of the MAPLE 4 experimental site, which is shown in Fig. 6. The noise-free and noisy synthetic data at 100, 150, and 200 Hz, representative of the frequency band of the experimental data, were generated using RAM. Spatially white-complex Gaussian noise with zero mean was added in the noisy data with SNR=20 dB. The number of signal frames was N_s=1 in this simulation. The modeled environment has slightly inclined bathymetry from 170 m at VLA (range of 0 km) to 153 m at 5 km from the VLA. There are four equally spaced source positions from 2 to 5 km and their depths are also equally 50 m. Thus, the segment interval is 1 km except for the first segment (2 km). All segments are assumed to have the one-sediment layer and the water column SSP was fixed as the measured profile in the experiment. As the acoustic

TABLE I. Model parameters and inversion results for the synthetic data.

		SEG1 $\lvert\theta\rvert \le 45$ deg			SEG2 17 deg $\le \lvert\theta\rvert \le 40$ deg			SEG3 17 deg $\le \lvert\theta\rvert \le 35$ deg			SEG4 17 deg $\le \lvert\theta\rvert \le 35$ deg		
Parameters	Search bound	True	Noise-free	SNR 20 dB	True	Noise-free	SNR 20 dB	True	Noise-free	SNR 20 dB	True	Noise-free	SNR 20 dB
r_{SRC} (km)	[True ± 0.1]	2.000	2.000	2.009	3.000	3.002	3.014	4.000	4.001	4.011	5.000	4.999	5.019
d_{SRC} (m)	[True ± 2.0]	50.0	49.9	50.3	50.0	50.0	49.9	50.0	50.0	50.0	50.0	50.0	50.4
w_{SRC} (m)	[True ± 2.0]	165.0	165.0	165.8	160.0	160.0	159.6	155.0	155.0	155.3	153.0	153.0	154.1
θ_0 (deg)	[−3.0 3.0]	0.000	−0.006	0.003	0.000	0.001	−0.022	0.000	0.015	−0.020	0.000	0.008	−0.025
r_G (km)	[Seg-length][a]	⋯	⋯	⋯	2.100	2.102	2.124	3.200	3.207	3.213	3.900	3.889	3.853
h_{sed} (m)	[5 50]	40.0	40.0	39.8	30.0	30.1	29.9	15.0	15.7	22.4	25.0	24.8	25.5
$c_{\mathrm{sed\text{-}top}}$ (m/s)	[1500 1900]	1540.0	1540.3	1540.4	1590.0	1591.5	1591.4	1710.0	1695.5	1698.6	1790.0	1798.3	1776.8
$c_{\mathrm{sed\text{-}bott}}$ (m/s)	[1500 1900]	1700.0	1699.3	1695.5	1630.0	1627.7	1632.1	1750.0	1776.7	1801.5	1840.0	1825.7	1867.6
c_{bott} (m/s)	[1600 2100]	1900.0	1900.9	1893.9	1850.0	1854.2	1830.4	1910.0	1914.7	1929.0	1950.0	1946.3	1988.0
ρ_{sed} (g/cm^3)	[1.3 2.1]	1.700	1.695	1.705	1.650	1.648	1.777	1.800	1.717	1.792	1.850	1.860	1.739
ρ_{bott} (g/cm^3)	[1.5 2.2]	2.000	1.952	1.964	1.900	1.797	1.903	2.000	1.907	1.920	2.050	1.678	1.901
α_{sed} dB/λ	[0.0 0.5]	0.200	0.201	0.194	0.150	0.153	0.205	0.120	0.136	0.188	0.100	0.111	0.036
α_{bott} dB/λ	[0.0 0.5]	0.100	0.236	0.119	0.080	0.260	0.280	0.100	0.180	0.367	0.080	0.198	0.363
E (Cost function value)			2.7×10^{-4}	3.9×10^{-3}		2.5×10^{-4}	8.2×10^{-3}		1.1×10^{-3}	8.4×10^{-3}		7.8×10^{-4}	5.7×10^{-3}

The search interval of r_G is from $[r_{\mathrm{SRC}}]_{\mathrm{previous\_seg}}-0.2$ (km) to $[r_{\mathrm{SRC}}]_{\mathrm{current\_seg}}-0.2$ (km).

arrivals at the VLA do not need any filtering in the first segment inversion, low-angle beams were included in the calculation of the cost function (beams above approximately 45° were not considered because their low energy content did not affect the inversion result). After the first segment, low-angle beams were filtered out.

The filtering beam angles, search intervals, and inversion results for four segments are summarized in Table I. The geometric parameters were accurately estimated over all segments for the noise-free data, but there were slight differences for the noisy data except for the array tilt. For the geoacoustic parameters, the sediment thickness and sound speed showed good convergence over all. As might have been expected, the result for the third segment is worst (both for the noise-free and noisy data) owing to its short interval between the geophysical interfaces in the segment. For the noisy data case in this segment, the sediment thickness and correlated sound speed have some errors, although the location of geophysical interface was well estimated. However, the cost function value of this local minimum has not deteriorated severely from that of the former segments. The density and attenuation were not well determined sometimes due to their low sensitivities. Especially estimations for the sub-bottom attenuation had relatively large errors except for the first segment, because the segment interval is too short to extract accurate information of the sub-bottom attenuation from the received signal. The scatter plots (cost function value as a function of parameter values) for noisy data case are shown in Fig. 7.

One of the main issues in this sequential processing is the accumulation of errors. Due to the nature of sequential inversion, the modeling errors in the prior subinversions are accumulated and affect the result of subsequent subinversions. And this effect of error accumulation will be more distinct for the sensitive parameters. This is identified by the obtained results which describe that the estimation accuracy of the geoacoustic parameters decreases with increasing range of the segment. Detailed error accumulation analysis is beyond the scope of current work, but the derived results for subsequent segments show that they are within tolerant inversion results when compared with typical geoacoustic synthetic inversions.

B. Experimental data

Matched beam inversions were performed for the first five segments along the source track using the received signals from different ranges. The three frequency tones at 110, 150, and 200 Hz were selected for the inversion. Assuming that no *a priori* information about the distribution of bottom characteristic is available, it is reasonable to set the segment intervals (Δr_{seg} in Fig. 4) equidistant as in the synthetic simulation. However, the equal segment intervals could not be achieved because some parts of the received signals were contaminated by the broadband interferences. Here, the signal frames were chosen as Figs. 2(a)–2(e) depending on the clarity of received signal, and the segment intervals correspond to 1.1, 1.5, 1.1, and 1.4 km for each segment. Using the same technique as in the simulation, beams corresponding to those interacting with the segment in question were used. The number of signal frames was $N_s=6$ using different signal frames at 5 s intervals over 30 s of data for each segment.

As shown in Fig. 3, the water column SSPs have considerable variability, and these profiles cause quite different pressure fields at the VLA. Since it is difficult to model this spatiotemporal fluctuation directly from measurements, an empirical orthogonal function (EOF) analysis[6,23] was performed. For the first segment, the starting cw signal from the source at 2.4 km and measured water column SSPs (XBT-017 and XBT-018) were used. After the first segment, the first four EOF coefficients were included in inversion parameters to approximate a representative SSP between the source and receiver.

The filtering beam angles, search bounds for the parameters, and inversion results are summarized in Table II.

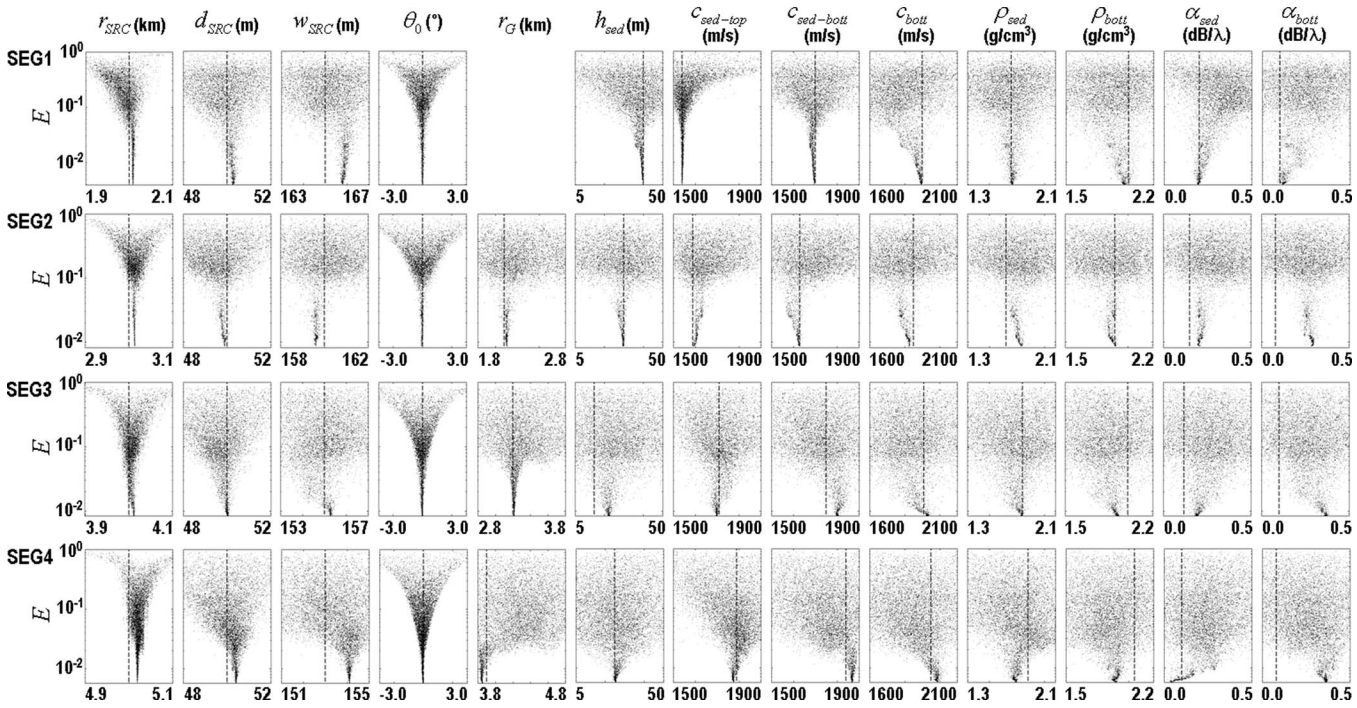


FIG. 7. Sensitivity analysis for the sequential inversion of synthetic noisy data. The dotted lines indicate true values.

Search bounds for the geometric parameters were tightly constrained based on the measured values, while such wide bounds were determined for the geophysical parameters in order to consider range dependently the varying properties based on the *a priori* information that the experimental site has sand-type sediment layers over hard sub-bottom. Search bounds for the EOF coefficients were determined based on the measured SSPs.

In Fig. 8, scatter plots and one-dimensional (1D) sensitivity curves of selected inversion parameters are shown for the five segments; the scatter plot just contains sample points in the simplex at the end of each temperature step. For the sensitivity curves, one parameter is perturbed at a time while all other parameters are fixed at the optimal inversion results. These 1D curves give the relative sensitivity and convergence of the estimates over their chosen bounds. For all the subinversions, the array tilt (fourth column in Fig. 8) was estimated to be small ($<1°$) with strong convergence and other geometric parameters also showed relatively high sensitivity. The sediment top sound speed $c_{\text{sed-top}}$ was the most sensitive geoacoustic parameter, while all the sub-bottom parameters (c_{bott}, ρ_{bott}, and α_{bott}) showed very low sensitivities

TABLE II. Sequential inversion results for the MAPLE 4 experimental data.

Parameters	Search bound	SEG1 $\lvert\theta\rvert \le 45$ deg	SEG2 17 deg $\le \lvert\theta\rvert \le 45$ deg	SEG3 17 deg $\le \lvert\theta\rvert \le 45$ deg	SEG4 17 deg $\le \lvert\theta\rvert \le 40$ deg	SEG5 17 deg $\le \lvert\theta\rvert \le 40$ deg
r_{SRC} (km)	[Measure ±0.1]	2.443	3.535	5.002	6.079	7.415
d_{SRC} (m)	[Measure ±1.5]	48.5	48.6	48.4	47.5	44.7
w_{SRC} (m)	[Measure ±3.0]	169.9	166.6	155.6	148.9	158.3
θ_0 (deg)	[−3.0 3.0]	0.49	0.20	0.77	0.25	0.58
r_G (km)	[Seg-length][a]	⋯	2.257	3.728	5.452	6.241
h_{sed} (m)	[5 60]	55.9	44.0	18.6	46.9	26.3
$c_{\text{sed-top}}$ (m/s)	[1500 1850]	1514.9	1544.2	1706.4	1609.7	1811.2
$c_{\text{sed-bott}}$ (m/s)	[1550 2000]	1984.5	1747.6	1893.9	1683.1	1766.3
c_{bott} (m/s)	[1700 3000]	2480.9	2129.5	2329.6	2164.8	2250.1
ρ_{sed} (g/cm³)	[1.3 2.1]	1.768	1.769	1.974	1.804	1.641
ρ_{bott} (g/cm³)	[1.6 2.8]	2.417	2.291	2.263	1.783	2.477
α_{sed} (dB/λ)	[0.0 0.5]	0.442	0.242	0.189	0.339	0.271
α_{bott} (dB/λ)	[0.0 0.5]	0.240	0.184	0.401	0.242	0.212
EOF1	[−30 30]	⋯	2.984	−20.311	−26.339	−2.968
EOF2	[−20 20]	⋯	0.059	2.652	11.595	−7.513
EOF3	[−15 15]	⋯	2.321	6.571	13.405	−10.849
EOF4	[−15 15]	⋯	2.884	0.816	0.389	10.955
E (Cost function value)		0.120	0.206	0.256	0.146	0.320

[a]The search interval of r_G is from $[r_{\text{SRC}}]_{\text{previous\_seg}} - 0.2$ (km) to $[r_{\text{SRC}}]_{\text{current\_seg}} - 0.2$ (km).

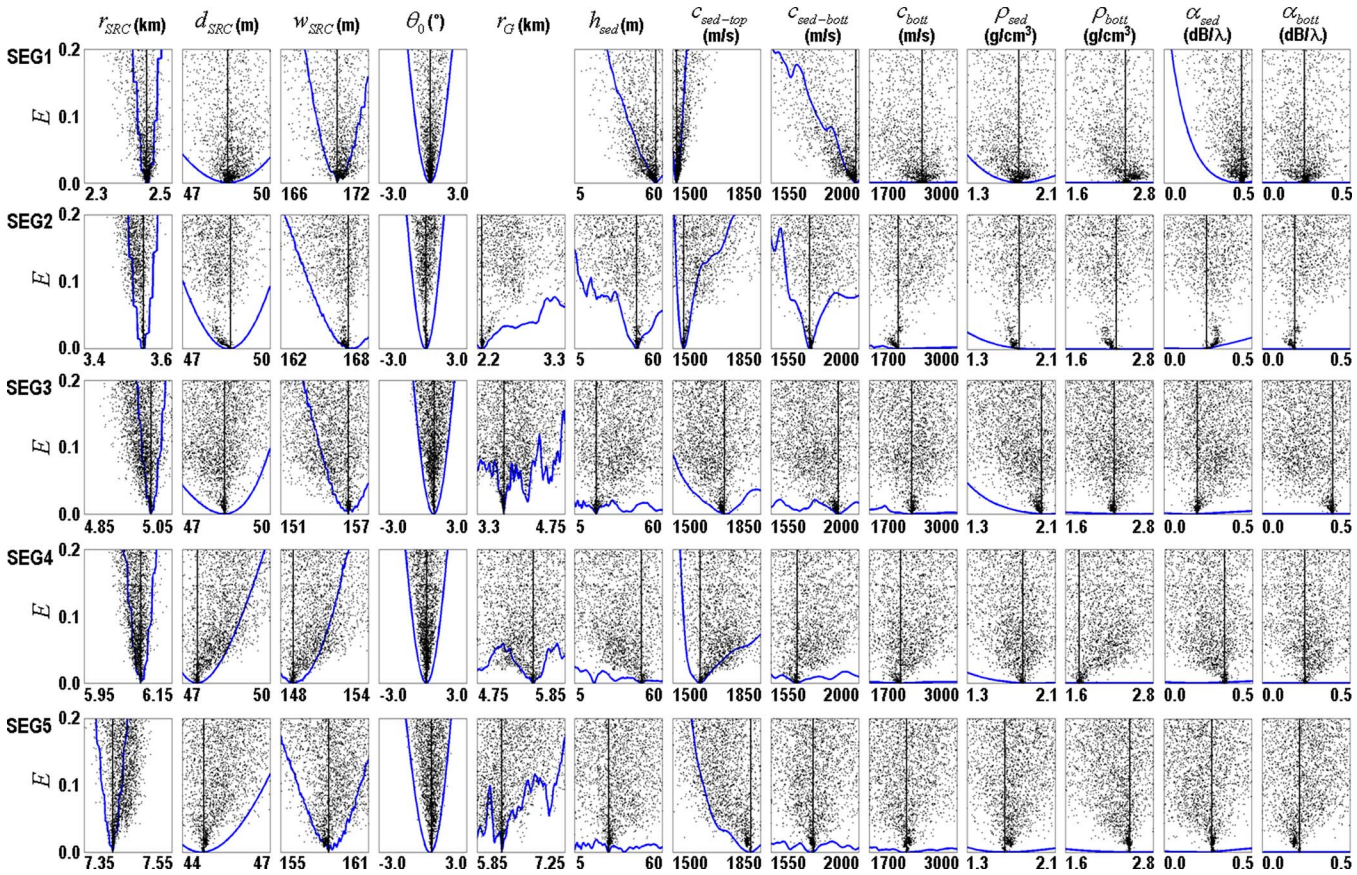


FIG. 8. (Color online) Scatter plots and 1D sensitivity curves of selected inversion parameters for the first five segments. The scatter plot just contains sample points in the simplex at the end of each temperature step. For the sensitivity curves, one parameter is perturbed at a time while all other parameters are fixed at the optimal inversion results (vertical line). For comparison, cost function values were normalized to have a minimum value of zero.

as is usual in typical inversion results. The geophysical interface r_G shows a particular characteristic in Fig. 8; it is a relatively sensitive parameter but its sensitivity curves have strong convergences at several discrete ranges. These uncertainties can be seen more specifically in 2D distributions of the cost function values. The 2D distributions for selected pairs of the geoacoustic parameters in the second and third segments are shown in Fig. 9; that of the fourth and fifth segments shows similar characteristic with that of the third segment. In the figure, the r_G-h_{sed} distribution for the second segment shows two local minima, and these are represented as two positive correlation (strong upper and weak lower) "valleys" having different slopes in the h_{sed}-$c_{\text{sed-top}}$ and h_{sed}-$c_{\text{sed-bott}}$ distributions. The r_G-h_{sed} distribution for the third segment reveals discrete three local minima along the r_G-axis. However, these are also represented as same number of valleys in the h_{sed}-$c_{\text{sed-top}}$ and h_{sed}-$c_{\text{sed-bott}}$ distributions. There is, on the other hand, only one valley in the distributions of h_{sed}-$c_{\text{sed-top}}$ and h_{sed}-$c_{\text{sed-bott}}$ for the first segment (not shown here). These are plausible results because the averaged values of the geoacoustic properties within the segment would be influenced according to the specific value of the geophysical interface which is allowed to change.

In Fig. 10, the SSPs obtained using estimated EOF coefficients for each segment were compared with the measured SSPs. Note that, as the source moves out (i.e., goes to subsequent segments), the approximated profile shifts toward the measured profiles having lower sound speed with increased fluctuations, especially in the thermocline. This

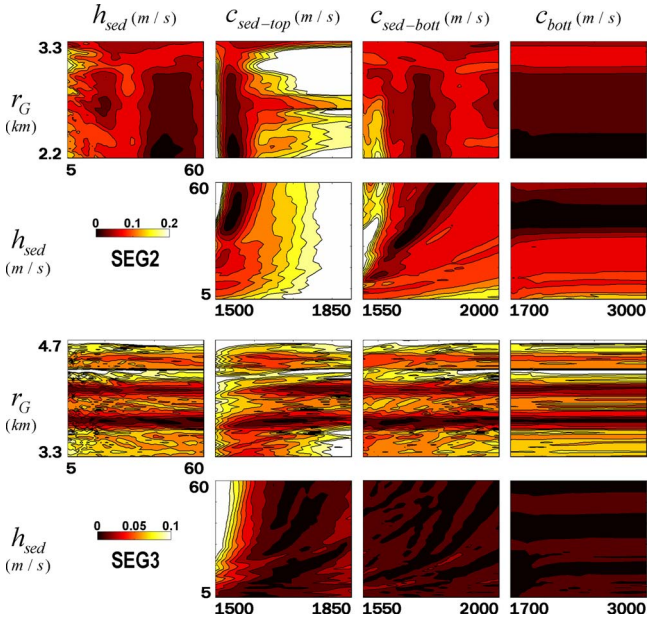


FIG. 9. (Color online) 2D distributions of cost function values for several pairs of the geoacoustic parameters in the second and third segments. The cost function values were re-evaluated using the optimal inversion results and normalized to have a minimum value of zero.

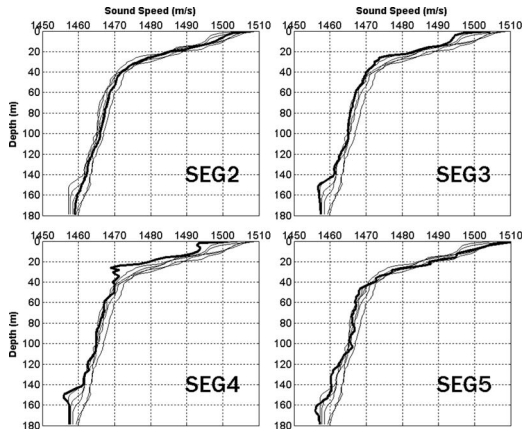


FIG. 10. The SSP obtained using estimated EOF coefficients for each segment (thick-solid lines) and measured profiles (thin-solid lines).

means that there were considerable spatial (or temporal) variability in the water column SSP and the result profiles were adapted to such variability.

C. Validations

Inversion results for the bottom sound speed, density, and attenuation are compared to those of the geophysical database in Fig. 11. The geophysical database at the coring positions were interpolated along the source track, and the interpolated *p*-wave SSPs are represented in the top panel of the figure with the estimated bathymetry and geophysical interfaces from the inversion. In Figs. 11(a)–11(c), the borders of the shaded area indicate min-max values of the property within each segment (between the geophysical interfaces). The attenuation of the database is at 400 Hz, and thus the inversion result was converted into dB/m kHz unit for comparison. Note that there is an intrusion of uplifted bed

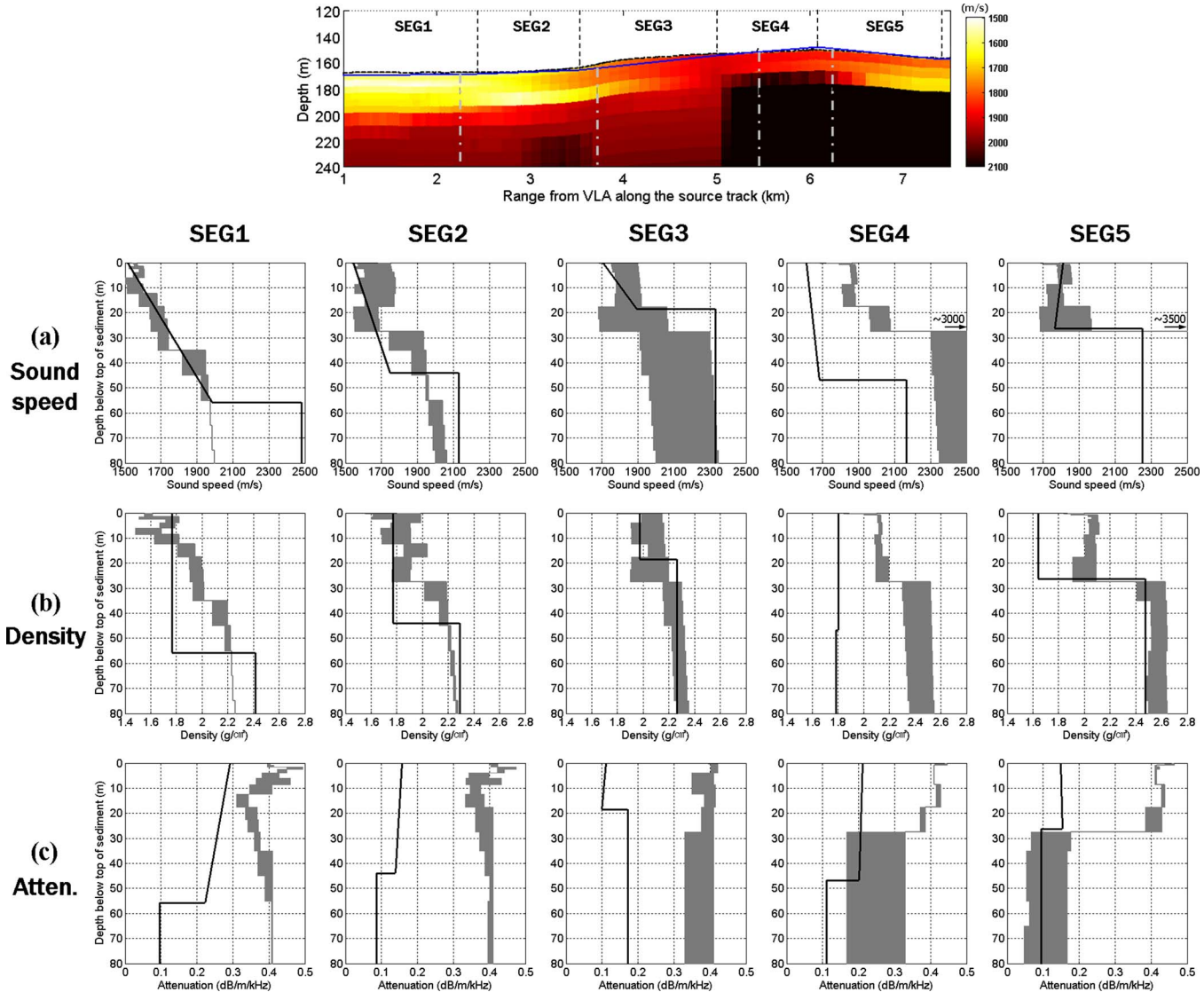


FIG. 11. (Color online) Comparison of bottom properties obtained from the geophysical database (shaded areas) and the inversion results (solid lines) for each segment; (a) sound speed, (b) density, and (c) attenuation. The geophysical database at the coring positions were interpolated along the source track, and the min-max values of the properties within each segment (between the geophysical interfaces) are represented by the shaded area. The top panel shows the interpolated *p*-wave SSPs of geophysical database as well as estimated bathymetry and geophysical interfaces from the inversion. The attenuation of the database is for 400 Hz; the inversion result was converted into dB/m kHz unit for comparison.

rock that begins at 5 km range from the VLA. The bed rock rises up to 20–30 m below the bottom surface and is covered with sand-type consolidated layers. (The p-wave and s-wave speeds of the bed rock are approximately 3500 and 2400 m/s.)

In the first and second segments, the estimated sediment thickness is thicker and the sub-bottom sound speed is higher than that of the database. The sub-bottom sound speed in the first segment could not be determined with precision due to the relatively thick sediment which has continuously increasing sound speed, with large (estimated) attenuation factor, as demonstrated in the geophysical database. Note that h_{sed} and $c_{\text{sed-bott}}$ of the first segment converged simultaneously toward the end of the search bound in Fig. 8; the 2D distribution of h_{sed}-$c_{\text{sed-bott}}$ showed strong positive correlation between the two parameters. Although the sediment attenuation α_{sed} in the first segment is relatively sensitive compared with that of other segments, the estimated value is smaller than that of the database which was evaluated at 400 Hz; the estimated values of the sediment attenuation in other segments are less reliable due to their low sensitivities. However, the inversion results for the first and second segments accurately characterize the soft bottom environment near the receiving array.

Unlike the inversion results for other segments, the result for the fourth segment seems to be in disagreement with the model from the database. In this segment, a thick sediment layer with the sediment sound speed of 1610–1683 m/s was estimated, which forms a pronounced contrast to that of adjacent (third and fifth) segments. According to the database, there is very thin (below 0.5 m) surface layer with a sound speed of 1600 m/s over lying a hard bottom (1900 m/s) in this region. There are two possibilities for the cause of this difference. One is the estimation error from the lack of information caused by the short effective length of the segment in question (r_G in the fourth segment was estimated toward the end of the search bound.). The other is the modeling error; one-layer model with constant gradient sound speed in the sediment is inadequate, particularly for the sediment structure in this segment. Although other estimated parameter values differ from that of the database model, it is interesting that such a low sediment top sound speed of 1610 m/s was estimated despite the low-frequency source signals which have a minimum wavelength of 7.3 m. It seems that this owes to using the high-grazing angle signals in the inversion, because a difference of the reflection coefficients between soft and hard bottoms is larger at high-grazing angles than at low-grazing angles.

Although the shear effect was not considered in the modeling, the effect of the bed-rock intrusion on the inversion was also one of our concerns. The scatter plots for the geophysical interface in the fourth and fifth segments shows relatively good convergences. However, it is not positive about the effect of the intrusion from the inversion results because the sensitivities of the sub-bottom properties are low and the estimated values of the sub-bottom sound speed and density are much smaller than that of the database. It is thought that the signal could not sufficiently penetrate into the sub-bottom due to the upper consolidated sandy sediment layers.

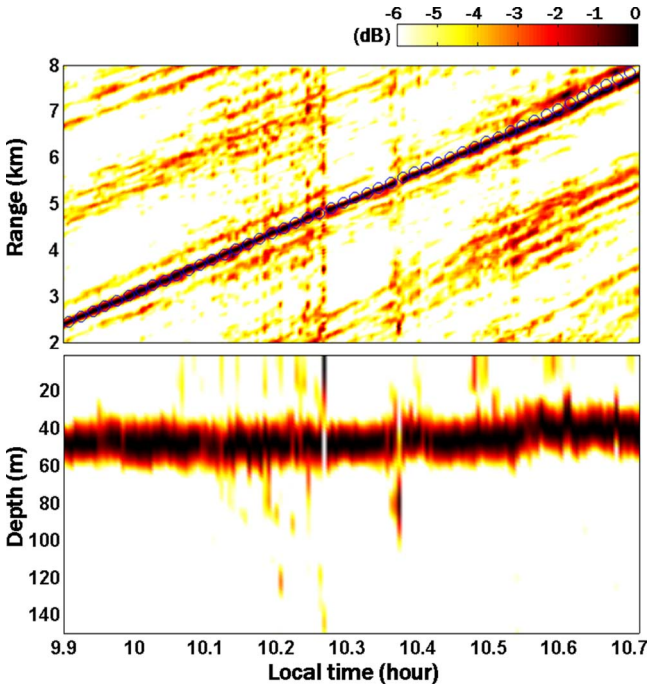


FIG. 12. (Color online) MF-derived source range and depth tracking results for incoherent averages over 80, 90, and 100 Hz (data for 49 min length). The circles in range tracking plot (upper panel) indicate the measured (GPS) source-receiver range. The mean measured source towing depth was 48.8 m.

Another typical method to validate the inversion results is via MF source localization (backpropagation) using different frequency signals than those used in the inversion. Here, frequency tones that are outside the band of frequencies used in the inversion are used for the source localization. The replicas of the full pressure (not beam filtered) field were generated using the set of best models. To check the consistency of the inversion results through all segments, the localizations were performed for the data over the source range from 2.4 to 8 km (49 min length). All the data at 20 s intervals were used, and the range and depth slices at the peak position in each Bartlett AMS were stacked. The MF-derived range and depth tracking results for incoherent averages over 80, 90, and 100 Hz are shown in Fig. 12. Although there are some side lobes due to the high noise signal level, the localized source positions match well with the measured source track.

V. CONCLUSIONS

In this paper, we presented the geoacoustic inversion results using VLA data in a range-dependent environment. The range variability of bottom properties in the experimental site was well resolved by the sectorwise inversion scheme and MBP with beam filtering. Although the validation of inversion results showed that this sectorwise inversion method can be applied successfully to range-dependent environments, several issues remain to be addressed. One is an accumulation of errors. The effect of error accumulation has been investigated via the simulation using synthetic data. However, an accurate quantification of error accumulation is a difficult problem and should be considered with a stochastic analysis (e.g., Bayesian inversion), which is beyond the

scope of this paper. The use of water column SSP data is another issue in the range-dependent inversion with a VLA that needs to be considered. In the fluctuating ocean environment, it is difficult to apply SSP data obtained from stationary (temporal or spatial) points into the propagation modeling. In this case, an *ad hoc* method such as EOF analysis, which was used in this paper, can considerably improve the inversion results. However, it should be noted that it can fail to approximate the variability in very long-range problems.

ACKNOWLEDGMENTS

This work was supported by a grant from the Agency for Defense Development of Korea. The authors would like to thank the crew of R/V Sunjin and the other participants in MAPLE 4.

[1]M. D. Collins, W. A. Kuperman, and H. Schmidt, "Nonlinear inversion for ocean bottom properties," J. Acoust. Soc. Am. **92**, 2770–2883 (1992).
[2]A. Tolstoy, N. R. Chapman, and G. E. Brooke, "Workshop '97: Benchmarking for geoacoustic inversion in shallow water," J. Comput. Acoust. **6**, 1–28 (1998).
[3]N. R. Chapman and M. Taroudakis, "Special issue: Geoacoustic inversion in shallow water," J. Comput. Acoust. **8**, 259–388 (2000).
[4]M. Siderius and P. L. Nielsen, "Range-dependent seabed characterization by inversion of acoustic data from a towed receiver array," J. Acoust. Soc. Am. **112**, 1523–1535 (2002).
[5]C. Park, W. Seong, P. Gerstoft, and M. Siderius, "Time domain geoacoustic inversion of high-frequency chirp signal from a simple towed system," IEEE J. Ocean. Eng. **28**, 468–478 (2003).
[6]P. Gerstoft and D. F. Gingras, "Parameter estimation using multifrequency range-dependent acoustic data in shallow water," J. Acoust. Soc. Am. **99**, 2839–2850 (1996).
[7]J. M. Ozard, M. L. Yeremy, N. R. Chapman, and M. J. Wilmut, "Matched-field processing in a range-dependent shallow water environment in the Northeast Pacific Ocean," IEEE J. Ocean. Eng. **21**, 377–383 (1996).
[8]M. Siderius and J. P. Hermand, "Yellow Shark Spring 1995: Inversion results from sparse broadband acoustic measurements over a highly range-dependent soft clay layer," J. Acoust. Soc. Am. **106**, 637–651 (1999).
[9]M. Musil, N. R. Chapman, and M. J. Wilmut, "Range-dependent matched-field inversion of SWellEX-96 data using the downhill simplex algorithm," J. Acoust. Soc. Am. **106**, 3270–3281 (1999).
[10]M. Siderius, M. Snellen, D. G. Simons, and R. Onken, "An environmental assessment in the Strait of Sicily: Measurement and analysis techniques for determining bottom and oceanographic properties," IEEE J. Ocean. Eng. **25**, 364–386 (2000).
[11]N. R. Chapman, S. Chin-Bing, D. King, and R. B. Evans, "Benchmarking geoacoustic inversion methods for range-dependent waveguides," IEEE J. Ocean. Eng. **28**, 320–330 (2003).
[12]P. Pignot and N. R. Chapman, "Tomographic inversion of geoacoustic properties in a range-dependent shallow-water environment," J. Acoust. Soc. Am. **110**, 1338–1348 (2001).
[13]S. Kim, J. Park, Y. Kim, E. Kim, Y. Kim, M. Park, and Y. Na, "Source localization using an L-shaped array in shallow water," in Proceedings of the Ninth Western Pacific Acoustics Conference, Seoul, Korea, 2006, p. 213.
[14]T. C. Yang and T. Yates, "Matched-beam processing: Application to a horizontal line array in shallow water," J. Acoust. Soc. Am. **104**, 1316–1330 (1998).
[15]T. C. Yang and T. Yates, "Matched-beam processing: Range tracking with vertical arrays in mismatched environments," J. Acoust. Soc. Am. **104**, 2174–2188 (1998).
[16]T. C. Yang and T. Yates, *Full Field Inversion Methods in Ocean and Seismo-Acoustics* (Kluwer Academic, Dordrecht, 1995), pp. 323–328.
[17]Y. Jiang, N. R. Chapman, and H. A. DeFerrari, "Geoacoustic inversion of broadband data by matched beam processing," J. Acoust. Soc. Am. **119**, 3707–3716 (2006).
[18]W. Ryang, Y. Kwon, J. Jin, H. Kim, C. Lee, J. Jung, D. Kim, J. Choi, Y. Kim, and S. Kim, "Geoacoustic characteristics of p-wave velocity in Donghae City—Ulleung Island line, East Sea: Preliminary results," J. Acoust. Soc. Korea **26**, 44–49 (2007).
[19]M. D. Collins, "A split-step Padé solution for parabolic equation method," J. Acoust. Soc. Am. **93**, 1736–1742 (1993).
[20]P. Gerstoft, *SAGA User Manual 5.1: An Inversion Software Package* Marine Physical Laboratory (Scripps Institution of Oceanography, University of California at San Diego, 2004).
[21]K., Kim, C., Park, W., Seong, and S., Kim, "Geoacoustic inversion via transmission loss matching and matched field source tracking," in Proceedings of the International Conference Underwater Acoustic Measurements, Crete, Greece, 2005.
[22]S. E. Dosso, M. J. Wilmut, and A. S. Lapinski, "An adaptive-hybrid algorithm for geoacoustic inversion," IEEE J. Ocean. Eng. **26**, 324–336 (2001).
[23]L. R. LeBlanc and F. H. Middleton, "An underwater acoustic sound velocity data model," J. Acoust. Soc. Am. **67**, 2055–2062 (1980).

Tracking of geoacoustic parameters using Kalman and particle filters

Caglar Yardim,[a] Peter Gerstoft,[b] and William S. Hodgkiss[c]
Marine Physical Laboratory, Scripps Institution of Oceanography, La Jolla, California 92093-0238

(Received 9 June 2008; revised 14 November 2008; accepted 20 November 2008)

This paper incorporates tracking techniques such as the extended Kalman, unscented Kalman, and particle (PF) filters into geoacoustic inversion problems. This enables spatial and temporal tracking of environmental parameters and their underlying probability densities, making geoacoustic tracking a natural extension to geoacoustic inversion techniques. Water column and seabed properties are tracked in simulation for both vertical (VLA) and horizontal (HLA) line arrays using the three tracking filters. Filter performances are compared in terms of filter efficiencies using the posterior Cramér–Rao lower bound. Tracking capabilities of the geoacoustic filters under slowly and quickly changing environments are studied in terms of divergence statistics. Geoacoustic tracking can provide continuously environmental estimates and their uncertainties using only a fraction of the computational power of classical geoacoustic inversion schemes. Interfilter comparison show that while a high-particle-number PF outperforms the Kalman filters, there are many cases where all three filters perform equally well depending on the inversion configuration (such as the HLA versus VLA and frequency) and the tracked parameters.
 [DOI: 10.1121/1.3050280]

PACS number(s): 43.30.Pc, 43.60.Pt, 43.60.Wy [AIT] Pages: 746–760

I. INTRODUCTION

Geoacoustic inversion is a technique used to extract information about the ocean environment by analyzing the acoustical field propagation in that medium. Typically, water column and seabed parameters such as sound speed profiles (SSPs), sediment densities, layer thicknesses, and attenuations are estimated by finding an environmental model that generates an acoustic field that matches closely the measured field. There are different configurations that are typically used in geoacoustic inversion, each with its own advantages and drawbacks. Some of the most commonly used ones include vertical (VLA) or horizontal (HLA) line arrays, bottom moored or towed arrays, and active or passive source configurations that use either a separate towed source or ship self-noise for inversion.[1–7] While some of the inversion techniques focus on obtaining the optimum solution with minimum computation time using efficient global optimizers such as genetic algorithms[8] or simulated annealing,[9] the others estimate the probability densities of the environmental parameters to compute the uncertainty in the estimated parameters.[10,11] This enables them to project this environmental uncertainty into parameters-of-interest such as the uncertainties in transmission loss and statistical sonar performance prediction.[12]

This paper reformulates the geoacoustic inversion algorithms that estimate the geoacoustic environment between the source and the receiver array at a given time into tracking the evolution of these parameters and their associated uncertainties in space and time. This is achieved by merging geoacoustic inversion techniques with tracking algorithms such as the Kalman and particle filters (PFs). These filters have been used previously in estimation[13] and temporal tracking[14] of the ocean SSP and similar acoustic applications.[15,16]

Here, the geoacoustic tracking problem is formulated in a Kalman framework, and depending on the source/receiver configuration, the acoustic field is calculated using either the normal mode code SNAP (Ref. 17) or the complex normal mode code ORCA (Ref. 18) for near-field calculations. This interaction between the environmental parameters and the acoustic field can involve a high level of nonlinearity. In addition, it is known from previous studies[19–21] that the posterior probability densities (PPDs) of geoacoustic parameters can be non-Gaussian. Therefore, geoacoustic tracking is a challenging task and requires tracking filters that can handle nonlinear, non-Gaussian systems. This paper studies the suitability of three such filters, namely, the extended Kalman filter[22] (EKF), the unscented Kalman filter[23] (UKF), and the PF (Ref. 24) in geoacoustic tracking. All three filters use different schemes to deal with such complex systems. The EKF extends the best possible filter in a linear/Gaussian system, i.e., the Kalman filter (KF), into the nonlinear/non-Gaussian domain by analytical linearization of the problem. Instead, the UKF uses statistical linearization with unscented transform. Finally, the PF propagates a large number of particles to represent the evolving probability density function (PDF) of the environmental parameters. In this paper, the tracked parameters are restricted to environmental parameters since detection and tracking of a target/source using tracking filters are already well-studied fields in applications involving sonar and radars.

Most cases that require consecutive geoacoustic inversions to obtain the spatial/temporal variation of geoacoustic parameters effectively can be reformulated as tracking prob-

[a]Electronic mail: cyardim@ucsd.edu
[b]Electronic mail: gerstoft@ucsd.edu
[c]Electronic mail: whodgkiss@ucsd.edu

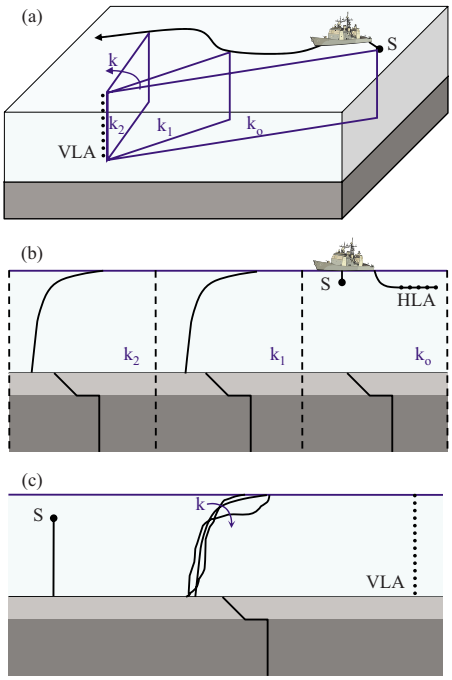


FIG. 1. (Color online) Geoacoustic tracking for three configurations: (a) Temporal tracking of the average, range-independent environment using a fixed-VLA-receiver and a towed-source, (b) spatial tracking of range-dependence using a towed-HLA-receiver and a towed-source, and (c) temporal tracking of the ocean sound speed profile for a fixed-VLA-receiver and a fixed-source.

lems. Three examples are shown in Fig. 1:

(a) Figure 1(a) shows a typical fixed hydrophone array (VLA or HLA) configuration and a separate towed source with the aim of capturing the environment between the moving source and the receiver array. The two dimensional environment between the source and the array changes as the source is towed, resulting in an evolution of the range-independent model in time, as shown in the figure as step indices k_i. For example, such a scenario could transform the following geoacoustic inversion approaches into geoacoustic tracking:
- a towed source and fixed VLA [e.g., SWARM'95 (Ref. 21)],
- a towed source and fixed HLA [e.g., SWAMI'98 (Ref. 2) and Barents Sea'03 (Ref. 6)],
- the test cases used in the Geoacoustic Inversion Techniques Workshop with a moving source and fixed HLA (Ref. 25), and
- a single hydrophone mounted on seafloor receiving transmission from a towed source [e.g., SCARAB'98 (Ref. 26)].

(b) Figure 1(b) represents the type of configuration that is designed to capture range-dependent environmental parameters at small range increments. The illustration given in the figure uses a HLA and a source close to the array, both towed by the same ship. Hence, the HLA captures the near-field acoustic field that is affected only by a small section of the water column and seabed. It is possible to take each of these sections as a step index in range and assume that the environment is constant within each k_i, and therefore turn it into a range-dependent environment tracking problem. The update rate can be increased using overlapping blocks but for the sake of simplicity, a nonoverlapping scheme is shown here. This type of geoacoustic tracking could be used in
- a towed source and HLA [e.g., MAPEX2000 (Refs. 1, 5, and 7)],
- tow-ship self-noise data acquired via a towed HLA [e.g., MAPEX2000 (Ref. 3)], and
- passive fathometer using the ocean ambient noise field measured by a drifting VLA [e.g., ASCOT'01 and Boundary'03 (Ref. 27)].

(c) The last example, given in Fig. 1(c), estimates the evolution in time of the water column SSP along a fixed path. Such a scenario could be implemented in the following examples:
- a fixed source and fixed VLA [e.g., Yellow Shark'94 (Ref. 28)] and
- KFs used to track the SSP in a similar configuration during the MREA/BP'07 experiment (Ref. 14)

The main objective of this paper is to incorporate tracking filters into the geoacoustic inversion problem and test the effectiveness of each filter in geoacoustic parameter tracking.

II. GEOACOUSTIC ESTIMATION AS A TRACKING PROBLEM

Geoacoustic inversion requires a measurement equation relating the simulated acoustic field to the observed data through a forward model. This is represented using[10,11]

$$\mathbf{d}^{\text{obs}} = s\mathbf{d}(\mathbf{m}) + \mathbf{e}, \tag{1}$$

where $\mathbf{d}^{\text{obs}}$ represents the complex-valued acoustic data vector along the array, s is the complex source magnitude, $\mathbf{d}(\mathbf{m})$ is the simulated field obtained using the acoustic propagation model for an environment represented by the environmental model vector $\mathbf{m}$, and $\mathbf{e}$ is complex Gaussian noise. Using Eq. (1), a geoacoustic inversion algorithm defines an objective function to be used in the inversion to obtain the best possible model $\hat{\mathbf{m}}$.

Geoacoustic tracking, on the other hand, uses two dynamic equations to characterize the system:

- An equation modeling the evolution of the environmental model parameters governed by the physical processes in the medium such as ocean currents and mixing, bathymetry, and the expected rate of change in seabed parameters in range.
- An acoustic measurement equation similar to Eq. (1). However, this is a dynamic equation that includes a continuous stream of data $\mathbf{d}_k^{\text{obs}}$, where k represents the temporal or spatial step index.

Following standard KF notation,[29] the error $\mathbf{e}$ in the measurement equation, the environmental model $\mathbf{m}$, and the acoustic field across the hydrophone array $\mathbf{d}^{\text{obs}}$ at step k henceforth will be denoted by $\mathbf{w}_k$, $\mathbf{x}_k$, and $\mathbf{y}_k$, respectively. Therefore, the set of equations at step k are

$$\mathbf{x}_k = \mathbf{f}(\mathbf{x}_{k-1}) + \mathbf{v}_{k-1}, \tag{2}$$

$$\mathbf{y}_k = \mathbf{h}(\mathbf{x}_k) + \mathbf{w}_k = s_k \mathbf{d}(\mathbf{x}_k) + \mathbf{w}_k, \tag{3}$$

where $\mathbf{f}(\cdot)$ is a known function of the state vector $\mathbf{x}_{k-1}$, and $\mathbf{h}(\cdot)$ is the nonlinear function that relates the environmental parameters $\mathbf{x}_k$ to the acoustic measurement vector $\mathbf{y}_k$. Hence, $\mathbf{h}(\cdot)$ includes both the unknown source term s_k and the known forward model $\mathbf{d}(x_k)$ as given in Eq. (1). $\mathbf{v}_k$ and $\mathbf{w}_k$ are the process/state and the measurement noise vectors, respectively, with

$$E\{\mathbf{v}_k \mathbf{v}_i^T\} = \mathbf{Q}_k \delta_{ki},$$

$$E\{\mathbf{w}_k \mathbf{w}_i^T\} = \mathbf{R}_k \delta_{ki},$$

$$E\{\mathbf{v}_k \mathbf{w}_i^T\} = \mathbf{0}, \quad \forall\, i,k, \tag{4}$$

where $\mathbf{Q}_k$ and $\mathbf{R}_k$ are the covariance matrices at time k of the corresponding noise terms. Similar to geoacoustic inversion, the state vector $\mathbf{x}_k$ is composed of the n_x parameters that describe the environment at step k. In geoacoustic tracking, however, these model parameters also evolve so the algorithm continuously updates the best estimate and the uncertainties in these estimates.

Equation (2) is the state equation modeling the evolution of the environmental parameters. Any prior knowledge about the environment and its evolution are modeled here. The values of the environmental parameters at step k are related to their values in the previous step $k-1$ by the function $\mathbf{f}(\cdot)$. If the environmental model evolution is linear, $\mathbf{f}(\cdot)$ can be modeled by the matrix $\mathbf{F}$, which is assumed in this paper. This assumption is reasonable if the parameter update rate is higher than the rate of change in the environment. The process noise $\mathbf{v}_k$ is a function of how correctly the evolution is modeled, which is usually taken as a zero-mean Gaussian PDF, allowing the filter to capture the changes per step that are not included in the evolution model. For example, the geoacoustic parameters are assumed to vary slowly with $\mathbf{F} = \mathbf{I}_{n_x \times n_x}$. Even though many geoacoustic parameters such as the sediment layer thickness may satisfy this condition most of the time, there can be sudden jumps at the boundaries of geological formations, violating the evolution model selected here. To continue tracking the parameters successfully through the sudden jump, the geoacoustic tracking filters will have to incorporate a state noise term with a high covariance $\mathbf{Q}_k$. The initial density $p(\mathbf{x}_0)$ can be obtained by running a Markov chain Monte Carlo geoacoustic inversion at $k=0$.

Equation (3) is the measurement equation relating the environmental model parameters to acoustic measurements. This process involves the selection of a suitable forward model that propagates acoustic fields and simulates the field observed across the receiver array for a given environmental model. The forward model is selected, taking into account the complexity of the model chosen to represent the environment, the selection of the source and receiver array configurations, and the available computational power. The most commonly used propagation models are normal mode (complex versions when the near-field is needed, adiabatic versions for mildly range-dependent configurations), ray tracing, and parabolic equation (using either split-step fast Fourier transform or Padé coefficients). SNAP is used here with the long-range VLA configuration, and complex ORCA is used with the near-field HLA simulations. The noise term is assumed complex Gaussian with covariance $\mathbf{R}_k$ obtained from the array signal-to-noise ratio (SNR) [see Eq. (9)]. Since the synthetic data used here are generated using the same forward model and environmental parameters, there is no modeling error in the examples. Working with real data will include unavoidably the modeling uncertainty, resulting in an increase in the noise term.

The KFs necessitate a linear/Gaussian framework, whereas any distribution can be used for the PF. This means that the prior PDF $p(\mathbf{x}_0)$, the state variables, state, and measurement noise all have to be Gaussian to run any KF algorithm as a geoacoustic tracking filter. PFs can work with any PDF. However, in order to compare these two types of filters under identical initial conditions, the prior densities at $k=0$ are taken as Gaussian PDF in this paper. The results of the KFs will all be Gaussian, while the PF densities can be of any distribution.

III. THEORY

For a system with linear state and measurement equations and Gaussian PDFs the KF (Ref. 22) is the optimal filter in a minimum mean square error (MSE) sense. However, for nonlinear, non-Gaussian problems such as geoacoustic tracking, it may not be possible to find an optimal estimator. Therefore, three suboptimal filters are investigated:

- EKF that uses analytical linearization where the measurement equation is linearized using the first order Taylor series expansion,
- UKF that uses statistical linearization where the nonlinearity in the parabolic equation is kept but PDFs are restricted to be Gaussian, and
- PF or sequential Monte Carlo (SMC), which uses a sequential importance resampling (SIR) or bootstrap filter to track the nonlinear, non-Gaussian system.

Each of these algorithms has their advantages and drawbacks for different tracking applications. See Appendix A for filter descriptions and implementation details.

The filters can be compared to each other using the root mean square (RMS) error between the true environment $\mathbf{x}_k$ and the filter estimate $\hat{\mathbf{x}}_{k|k}$. However, this only shows if one filter is doing better than the others, giving no indication about whether and to what extent the information available through previous states and current measurements are exploited by the filter, especially given the fact that all three of these filters are suboptimal. Therefore, it is desirable to have a tool that can not only assess the performances of these techniques but also provide a limit to achievable performance for a given environment.

This is done by using the posterior or Bayesian Cramér–Rao lower bound (PCRLB) introduced by van Trees[30] (see Appendix B). PCRLB is the Bayesian counterpart of the

classical Cramér-Rao lower bound (CRLB) defined in a non-Bayesian framework as the inverse of the Fisher information matrix. There are studies on the calculation of both the CRLB (Ref. 31) and the PCRLB (Ref. 32) for geoacoustic inversion problems. Any filter that achieves a MSE equal to the PCRLB is called an efficient estimator. For a linear and Gaussian system, the KF is an efficient estimator. It may not be possible to attain the PCRLB for a nonlinear, non-Gaussian system.

The performance metrics used in this paper are

$$\mathrm{RMS}_k(i) = \left[\sum_{j=1}^{n_{\mathrm{MC}}} \frac{(\hat{\mathbf{x}}_k^j(i) - \mathbf{x}_k^j(i))^2}{n_{\mathrm{MC}}} \right]^{1/2}, \tag{5}$$

$$\eta_k(i) = \mathbf{J}_k^{-1/2}(i,i)/\mathrm{RMS}_k(i), \tag{6}$$

$$\mathrm{RTAMS}(i) = \left[\sum_{k=k_1}^{k_2} \sum_{j=1}^{n_{\mathrm{MC}}} \frac{(\hat{\mathbf{x}}_k^j(i) - \mathbf{x}_k^j(i))^2}{(k_2 - k_1 + 1) n_{\mathrm{MC}}} \right]^{1/2}, \tag{7}$$

$$\mathrm{Improv} = \frac{\mathrm{RTAMS}_{\mathrm{EKF}} - \mathrm{RTAMS}_{\mathrm{filter}}}{\mathrm{RTAMS}_{\mathrm{EKF}}}, \tag{8}$$

where $\mathbf{x}_k^j(i)$ is the ith parameter of the true state vector $\mathbf{x}$ at time index k for the jth MC run, RMS_k, $\mathbf{J}_k$, and η_k are the root mean square error, the Fisher information matrix (inverse of the PCRLB), and the filter efficiency, respectively, at step k. RTAMS is the root time averaged mean square error[29] calculated for the interval $[k_1, k_2]$, and Eq. (8) calculates the performance improvement of a filter with respect to the EKF.

IV. EXAMPLES

This section is composed of three geoacoustic tracking examples that either spatially or temporally track the evolving environment and the PPD. The first two use the configuration in Fig. 1(a), and the last one uses the one in Fig. 1(b). The simulation parameters such as the array structure, water depth, and source frequencies are selected similar to the ones that are used in Refs. 1 and 20. Each example evaluates and compares different aspects of the tracking algorithms. These three examples are

(1) Temporal tracking. Filter efficiencies, PCRLB calculations, performance limitation analysis, computational costs, effects of increasing the particle size in PF, and interfilter comparison of uncertainty propagation are studied using temporal tracking of an effective range-independent environment (with n_x=4 unknown parameters at each step k).
(2) Divergence analysis. For both slowly varying and abruptly changing environments, a divergence analysis is carried out using n_x=7 unknown parameters at each step k.
(3) Spatial tracking. The effects of selection of geoacoustic setup on filters and tracking performance of individual geoacoustic parameters are investigated using spatial tracking of a range-dependent environment represented by n_x=7 unknown parameters at each spatial step k.

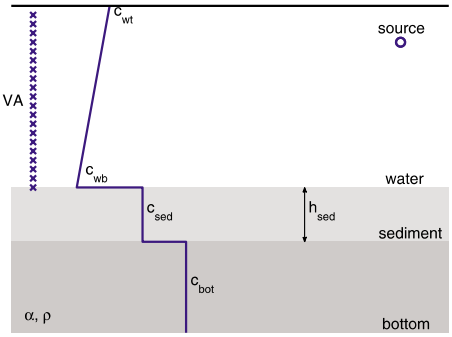


FIG. 2. (Color online) Seven-parameter geoacoustic model used in the simulations.

A. Example 1: Temporal tracking using a VLA

This example compares the performance of the EKF, UKF, and PF with the best possible limit given by the PCRLB in terms of the RMS error and the filter efficiency. The range-independent environment model used is given in Fig. 2. Note that the selection of the environmental model is arbitrary, and multiple more complex models can be incorporated into filters such as the multiple model particle filter (MMPF).[29]

Only the four parameters representing the sediment layer, namely, sound speed, thickness, attenuation, and density, are tracked in this example. A sandy silt with medium-fine to fine sand sediment is used in the tracking.[7] A VLA spanning the entire 100 m water column with 20 hydrophones is used. A frequency of 250 Hz is selected. All the environmental constants, state variables, their initial means and covariances, and the filter parameters are given in Table I. The covariance of the measurement error term $\mathbf{R}=\nu\mathbf{I}$ is computed from the array SNR (Ref. 20) defined as

$$\mathrm{SNR} = 10 \log \frac{[s\mathbf{d}(\mathbf{x})]^H [s\mathbf{d}(\mathbf{x})]}{\nu}. \tag{9}$$

The PCRLB and the filter performances are calculated using the Monte Carlo (MC) analysis as discussed in Appendix B. First, n_{MC}=100 evolving environments (each one a MC trajectory) are created using the state equation, with starting values selected from a Gaussian with a mean of $\mathbf{x}_0$ and covariance $\mathbf{P}_0$. These trajectories are given in Fig. 3(a). Then the PCRLB is computed using Eq. (B2) where the first term ($\mathbf{D}_{k-1}^{22}$) is estimated using Eq. (B8). Each of these 100 trajectories is also tracked by the EKF, UKF, and PFs using 200, 2000, and 10 000 particles designated by PF-200, PF-2000, and PF-10 000, respectively. The normal mode code SNAP is selected as the forward model.

The evolution of the RMS error in Eq. (5) of each parameter is computed for each filter and is given in Fig. 3(b) as a function of step index k. Note that the region below the square root of the PCRLB is shaded as unobtainable RMS values. Also note that the filter RMS error estimates can initially get lower than this limit before they increase and stabilize to their real values. Hence, this region is discarded in the calculations by setting the $[k_1, k_2]$ interval as [100, 150] min for the RTAMS in Eq. (7) and their following improvement-over-EKF computations in Eq. (8).

TABLE I. Environmental and simulation parameters used in example 1.

Environment					
		State variables			
Constants		$\mathbf{x}$	$E[\mathbf{x}_0]$	$\mathbf{P}_0^{1/2}$	State noise ($\mathbf{Q}_k^{1/2}$)
c_{wt}	1480 m/s	c_{sed} (m/s)	1600	1	0.35
c_{wb}	1460 m/s	h_{sed} (m)	15	0.5	0.35
h_w	100 m	α_{sed} (dB/λ)	0.25	0.01	0.0015
c_{bot}	1700 m/s	ρ_{sed} (g/cm^3)	1.8	0.1	0.03

Simulation parameters			
Source depth	20 m	Source frequency	250 Hz
Source range	5 km	Array SNR ($\mathbf{R}_k$)	40 dB
Receiver type	VLA	Track length	2.5 h (k=30)
No. of hydrophones	20	Track frequency	1 measurement/5 min
Array start, Δz	5 m, 5 m	MC runs	100

The results given in Fig. 3(b) show that the PFs perform better than the EKF and the UKF. While sediment thickness and sound speed tracking using PF is clearly superior to the KF variants, only PF with a large number of particles outperforms the EKF and UKF tracking of the sediment density, and all three types of filters perform well for attenuation tracking, closely following the theoretical limit set by the PCRLB. The RMS errors in sediment parameters, the average efficiency after 2.5 h of tracking, RTAMS values, and improvement-over-EKF percentages are given in Table II. Due to its inherent limitations, the EKF achieves an average filter efficiency of 52%. The UKF performs only slightly better with a 2% improvement over the EKF. With an efficiency of 63%, the PF-200 is 19% better than the EKF. Increasing the particle number improves performance to 80% efficiency in the PF-2000. The PF-10 000 results show that further increase in the particle number does not result in an increase in the performance, with a 37% improvement over the EKF out of a theoretical upper limit of 48% dictated by the PCRLB. PF-10 000 results are not shown in Fig. 3(b) but are given in Table II.

Even though PF performs better than the KFs in terms of RMS errors, it is also important to compare the computational cost of each algorithm both with each other and with

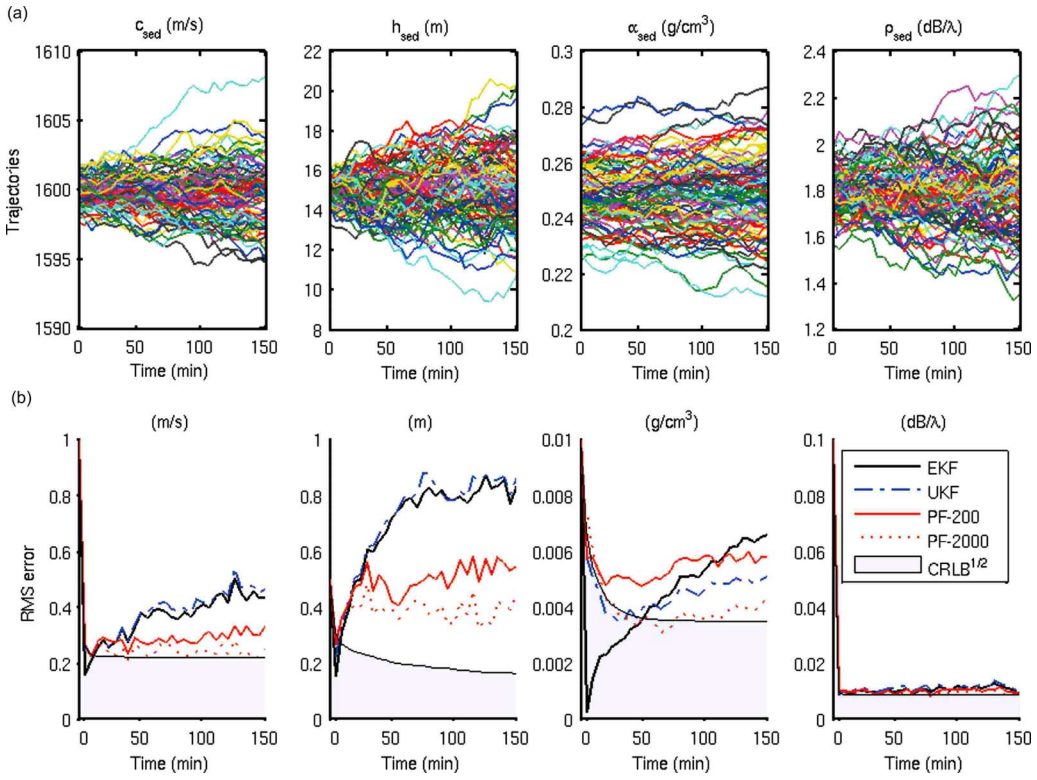


FIG. 3. (Color online) Example 1: Comparison of the tracking algorithms: (a) Evolution of 100 different environments (Monte Carlo trajectories), and (b) RMS errors for the EKF, UKF, 200-point PF, and 2000-point PF obtained from tracking each of these 100 trajectories along with the theoretical lower limit for the RMS error, the square root of the posterior CRLB.

TABLE II. Performance comparison for example 1.

Method	RMS at $t=150$ min c_{sed} (m/s)	h_{sed} (m)	α_{sed} (dB/λ)	ρ_{sed} (g/cm^3)	Avg. η (%)	RTAMS (100–150 min) c_{sed} (m/s)	h_{sed} (m)	α_{sed} (dB/λ)	ρ_{sed} (g/cm^3)	Avg. % Improv. over EKF
EKF	0.43	0.77	6.5×10^{-3}	10.7×10^{-3}	52	0.44	0.82	6.1×10^{-3}	11.3×10^{-3}	0
UKF	0.45	0.80	5.0×10^{-3}	11.1×10^{-3}	55	0.46	0.84	4.9×10^{-3}	11.8×10^{-3}	2
PF-200	0.30	0.53	5.8×10^{-3}	9.9×10^{-3}	63	0.31	0.54	5.8×10^{-3}	10.4×10^{-3}	19
PF-2000	0.22	0.39	4.2×10^{-3}	9.3×10^{-3}	80	0.24	0.39	3.9×10^{-3}	9.6×10^{-3}	36
PF-10 000	0.22	0.39	4.2×10^{-3}	9.3×10^{-3}	81	0.24	0.39	3.9×10^{-3}	9.6×10^{-3}	37
$\sqrt{\text{PCRLB}}$	0.22	0.16	3.5×10^{-3}	8.8×10^{-3}	100	0.22	0.17	3.5×10^{-3}	8.8×10^{-3}	48

previous classical geoacoustic inversion techniques. The forward model runs are by far the most computationally intensive section in a geoacoustic inversion or tracking problem. Therefore, the computational costs are given in terms of the number of forward models needed to run at each step k. First, if these parameters were to be inverted as a geoacoustic inversion problem using a global optimizer such as genetic algorithms, one would need around 10 000–40 000 forward model runs[1,8] for each step k. If the uncertainties or the parameter PDF are also required, techniques such as importance,[10] Gibbs,[4] and Metropolis–Hastings sampling[11,12,33] would be needed at each step requiring typically 100 000–1 000 000 samples per k. Using a hybrid genetic algorithms–Markov chain Monte Carlo (GA-MCMC) sampler may reduce this number,[34] but still the required number of forward model runs is large compared to the techniques discussed here. Instead, geoacoustic tracking requires an initial mean and a PDF that is obtained by running a classical geoacoustic inversion at $k=0$ and then tracking this density, and the optimum solution is done using the filter. The EKF requires $2\times n_x$ forward model runs at each step to compute the Jacobians needed for linearization and one forward model at the prediction step (nine forward model runs per k for this particular example). Similarly, the UKF uses $(2\times n_x+1)$ sigma points to propagate the mean and covariance using the unscented transformation (UT). The PF-200, −2000, and −10 000 require factors of 20, 200, and 1000 more CPU time than those of the EKF, respectively, for this scenario. Therefore, the selection of the filter type is a trade-off between the gain in performance and the extra computational burden of the PF.

Since the PF is computationally expensive compared to the KFs, it is desirable to know how much improvement can be obtained by using PF with a large number of particles. Unfortunately, the optimum n_p is scenario dependent. The effect of increasing particle size for this problem is shown in Fig. 4. A value of around 2000 particles per k provides maximum filter efficiency (81%) with minimum computational cost.

B. Example 2: Divergence analysis for slowly and fast changing environments

Divergence is an important issue in tracking problems. There are many reasons a track will diverge, such as the limitations in the filter (e.g., a KF structure in a highly non-Gaussian problem), errors in the forward model, and incorrect assumptions about the state and/or measurement noise. A frequently encountered problem is the error in the state equation model. The state equation models how we expect the state parameters to behave with k. If the real values of the state parameters evolve differently from this state evolution model, the filter may be unable to track these changes. Even though the measurement equation may tell the filter that the parameters are changing in an unexpected way, the filter may ignore the measurement information coming from Eq. (3) if this contradicts the state evolution model in Eq. (2). The filter type selected will affect the level of state modeling error that can be handled without resulting in divergence.

Geoacoustic tracking is no different. The state model used here assumes that the environment is evolving slowly. Therefore, comparing filter behavior under both slowly and fast changing environments is desirable. The environment in Fig. 2 is selected with $n_x=7$ environmental parameters to track, with the same VLA configuration and forward model as in the previous example. The simulation and measurement configuration parameters are selected from the sensitivity

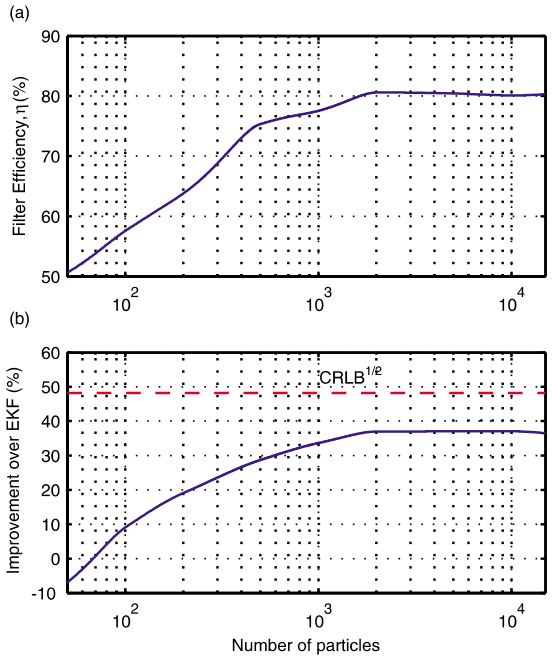


FIG. 4. (Color online) Example 1: Performance improvement of PF as a function of number of particles expressed in terms of (a) filter efficiency, and (b) improvement over the EKF. The dashed line shows the attainable improvement limit.

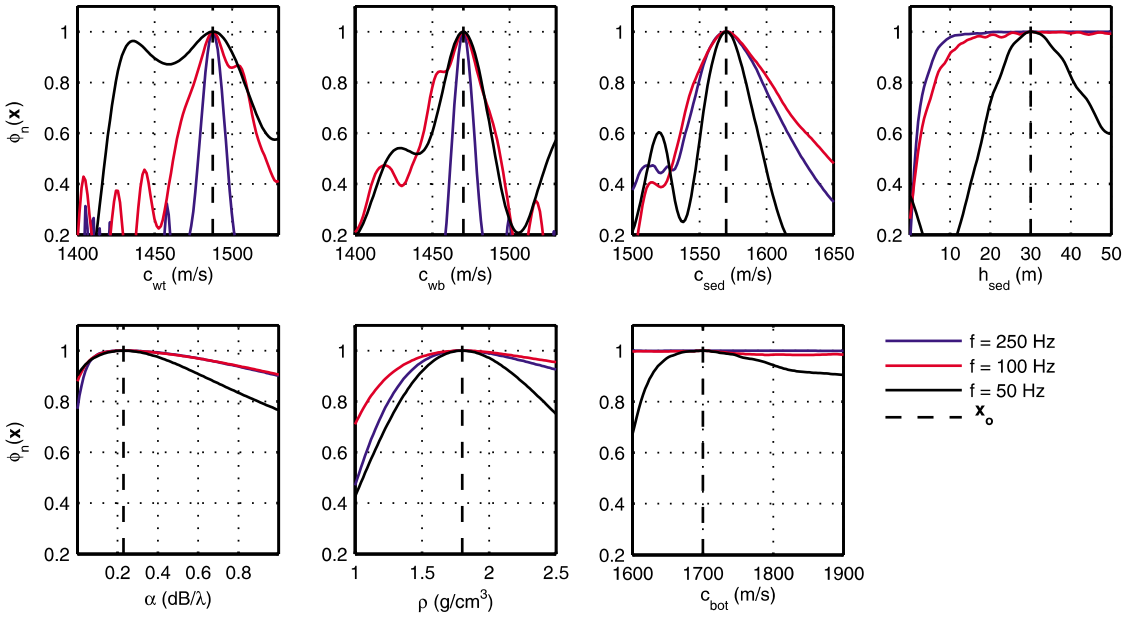


FIG. 5. (Color online) Example 2: Normalized objective functions for three different frequencies. The cost function for each parameter is obtained by fixing all other parameters to their true values (dashed line).

plots at $t=0$, given in Fig. 5. These plots are obtained by varying one parameter at a time from their values at $t=0$, while the other six are the same as those of $\mathbf{x}_0$. The normalized objective function used here is similar to ones used in Refs. 20 and 10 as a MSE metric. It is given as

$$\Phi_n(\mathbf{x}) = 1 - \frac{1}{\mathbf{y}_k^H \mathbf{y}_k} \left\| \mathbf{y}_k - \frac{\mathbf{d}(\mathbf{x}_k)^H \mathbf{y}_k}{\|\mathbf{d}(\mathbf{x}_k)\|^2} \mathbf{d}(\mathbf{x}_k) \right\|^2. \tag{10}$$

The objective functions in Fig. 5 are given for three different frequencies at 50, 100, and 250 Hz, respectively. Note that the penetration depth of the field decreases as the frequency is increased. Since the source and VLA are separated by 5 km, most of the high-incidence angle deep penetrating modes have attenuated at longer ranges and may not be detected by the receiver array. Note that $\Phi_n(\mathbf{x})$ for sediment thickness becomes insensitive after a certain value, which decreases with increasing frequency. The same also applies to the bottom sound speed. For the given environment, most of the signal is restricted to the sediment, not penetrating deep enough; hence $\Phi_n(\mathbf{x})$ is not sensitive to the bottom parameters.

Simulation parameters different from the previous example are provided in Table III. The tracking is carried out for 200 min with one update every 2 min. A frequency of 250 Hz is selected for the tracking problem. At this frequency, the bottom parameters give an entirely flat sensitivity plot, and sediment thickness above around 20 m is poorly determined. The evolutions of the seven parameters are given as solid lines in Fig. 6. These variations include a fluctuation in the top water sound speed, simultaneous gradual variations in all seven parameters, and a simultaneous sudden jump in two sediment parameters, sediment thickness from 30 to 20 m followed by a similar increase in the sediment sound speed. Note that one of the two environmental parameter jumps is in the sediment, a poorly determined parameter. Therefore, the filters are expected to give high divergence percentages due to the selection of such an environment and frequency, enabling a comparison between them under conditions difficult for tracking purposes. The evolving environment is tracked using the EKF, UKF, and PF that use 200, 2000, and 5000 particles, respectively. PF-2000 results are not shown in Fig. 6 but are given in Table IV.

The corresponding temporal evolution of the amplitude of the vertical acoustic field at 5 km as a function of time is given in Fig. 7. Note how the vertical mode structure evolves with time. Also note that only a sampled version of this field is used in tracking, as shown in the figure as circles representing the vertical hydrophone locations of the VLA. A lower spatial sampling frequency of the vertical field may result in the loss of some of the evolving trends in the field and higher divergence rates.

A typical track result for each filter is given in Fig. 6 along with the true trajectories of the parameters, and the results are summarized in Table IV. Some of the important

TABLE III. Simulation parameters for example 2.

Parameter	c_{wt} (m/s)	c_{wb} (m/s)	c_{sed} (m/s)	h_{sed} (m)	α (dB/λ)	ρ (g/cm^3)	c_{bot} (m/s)
State noise $\mathbf{Q}_k^{1/2}$	0.5	0.5	0.5	1.0	0.002	0.02	1.5
Initial cov. $\mathbf{P}_0^{1/2}$	1.0	1.0	1.0	1.0	0.002	0.02	3.0
Divergence threshold	1.0	1.0	1.0	1.0	0.01	0.05	10.0

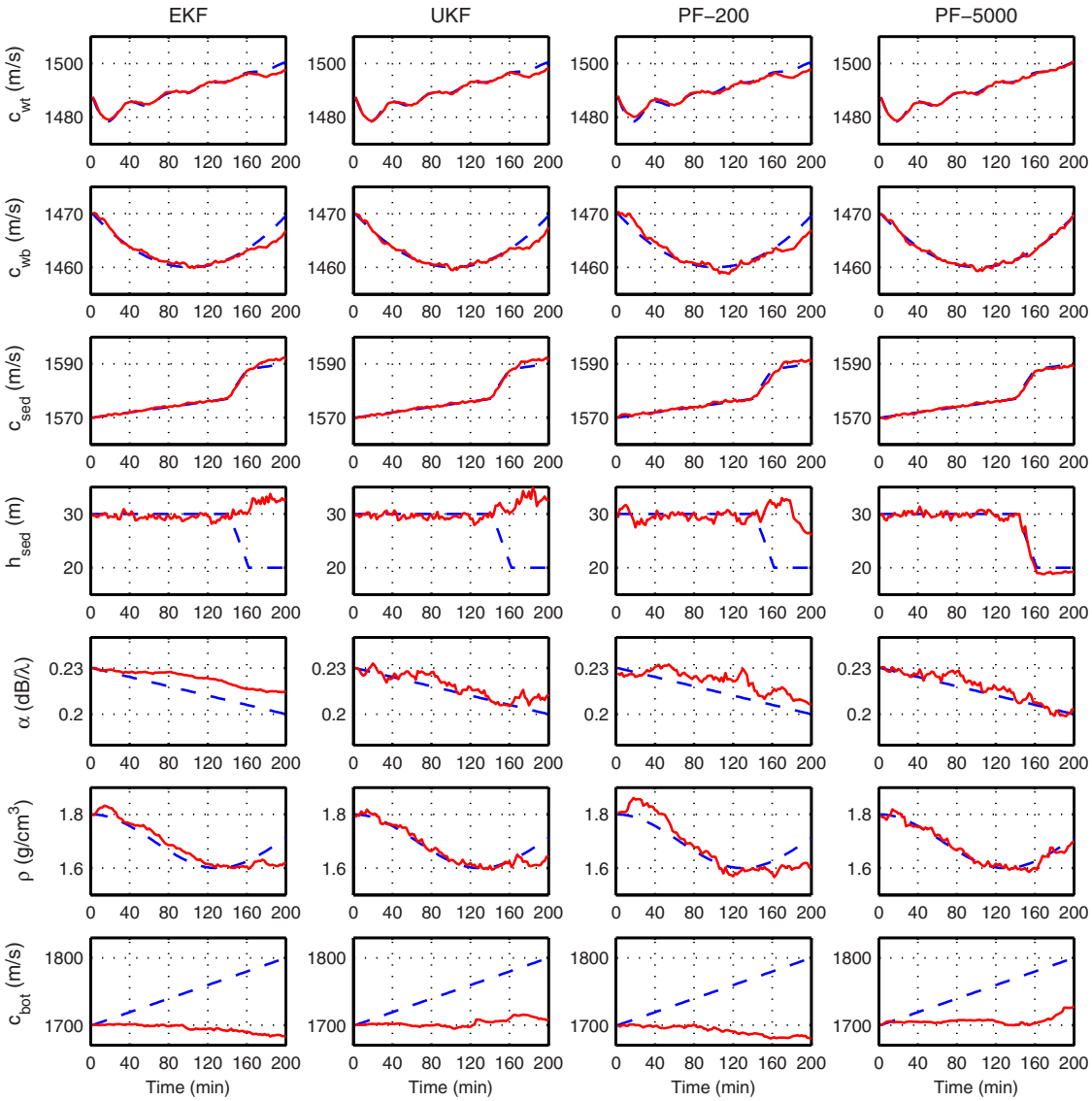


FIG. 6. (Color online) Example 2: Tracking results of EKF, UKF, PF-200, and PF-5000 for the seven-parameter environment given in Fig. 1 using the long range VLA. True trajectories (dashed) are provided along with the tracking filter estimates (solid).

features in this figure are as follows:

- All four filters are sensitive to the water column sound speed parameters and are able to track them. Water column parameters only start to diverge after the jump at $t = 140$ min for the EKF, UKF, and PF-200 because these filters are unable to track some of the sediment parameters that are coupled to the water column sound speed values. The PF-5000 is able to track these parameters perfectly both during slow ($t < 140$ min) and rapid ($t >= 140$ min) changes. Although the PF-200 could track the slowly changing sound speed values, the track is much noisier than the KF filters and the high-particle PF. A similar pattern emerges for the sediment density.
- All four filters are mostly able to track the sediment sound

TABLE IV. Results for example 2.

	After 140 min			After 200 min					
						% divergence			
Method	RTAMS h_{sed} (m)	% Imp. EKF	% Avg. Diverg.	RTAMS h_{sed} (m)	% Imp. EKF	$c_{w,sed}$	h_{sed}	α	ρ
EKF	0.75	0	16	10.6	0	82	100	68	6
UKF	0.71	24	0	11.2	15	62	100	1	12
PF-200	2.92	–70	39	9.2	3	58	90	61	48
PF-2000	0.83	16	1	4.8	42	29	40	7	12
PF-5000	0.82	20	0	3.1	60	11	19	1	5

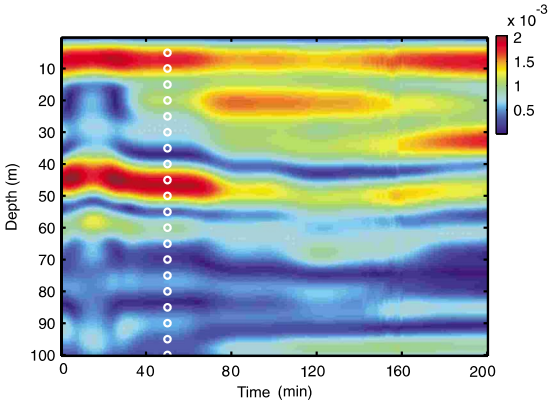


FIG. 7. (Color online) Example 2: Evolution of the magnitude of the vertical acoustic field in the water column at the receiver array as the environment evolves in time. Hydrophone locations (circle) show the vertical sampling interval of the time-varying field.

speed, including the sudden jump in the parameter. Again, the track given by PF-5000 is superior to the other three.

- As expected, the first three filters fail to track the sudden jump in the sediment thickness. Only PF-5000 is able to track the true trajectory. Also note how noisy the track is even for the PF-5000 due to the low sensitivity predicted in Fig. 5.
- Attenuation is the only parameter where there is a marked difference between the EKF and the UKF. The improvements introduced by the UKF over the linearized EKF enable it to track the attenuation, whereas the EKF divergence rates are much higher. PF-200 performance lies somewhere between the two KFs, and PF-5000 performance is very similar to the UKF performance, except for the superior performance after the jump due to divergence of other parameters in the UKF.
- All four filters are unable to track the bottom sound speed. This is an expected result, taking into account the entirely flat sensitivity curve given in Fig. 5.

The divergence percentages are given in Table IV for slowly changing (before $t=140$ min) and fast changing (after $t=140$ min) environments. A parameter track is declared diverged if the RMS error is greater than the corresponding threshold given in Table III for any 30 consecutive min (i.e., 15 samples). All the average values in Table IV are computed using the first six parameters, excluding the bottom sound speed, which always diverge. Note how the KFs have low RTAMS for the sediment thickness compared to the RTAMS of PFs before the jump. The average improvement over EKF is 20% for PF-5000, and overall, the UKF performs best in this region. The UKF, PF-2000, and PF-5000 almost always successfully track the trajectory, while the average divergence rates are 16% and 39% for the EKF and PF-200, respectively.

However, both KFs have difficulties at the jump in the sediment thickness. The UKF still outperforms the EKF by 15%, but the improvement goes up to 60% for the PF-5000 (Table IV). The average divergence in the water column and sediment layer sound speed values (designated as $c_{w,\mathrm{sed}}$) are given after the jump. The UKF diverges less than the EKF in sound speed tracking and more in attenuation tracking, both filters have a 100% divergence for the sediment thickness, and the UKF still tracks the attenuation whereas the EKF diverges 68% of the time after the jump. The PF-5000, on the other hand, diverges only 19% of the time for the hard-to-track sediment thickness, and overall, the PF performs much better than the KF structures after the jump.

It is also of interest to observe the underlying PDFs of the evolving parameters and examine how the uncertainty in parameters change with filter. The evolving PPD of the sediment thickness as a function of time is given for a PF-10 000 and the EKF in Fig. 8. They start with the same initial Gaussian PDF as seen at $t=0$. Both filters are able to follow the parameter until the sudden decrease in the sediment thickness. Note that the PDF of the EKF is always a Gaussian (due to the initial Gaussian assumption and linearization), whereas the PF density can take different forms, which enables the filter to simultaneously follow multiple regions in the state space with high likelihoods (such as at $t=116$ and 152 min. As the parameter starts to evolve quickly, the EKF is unable to follow, and it diverges, as can be seen from the large error in the PDFs given after $t=140$ min between the PF and the EKF. Note how stable the PDF evolution in Fig. 8(b) is at $h_{\mathrm{sed}}=20$ m compared to the 30 m sediment thickness region due to the flat sensitivity curve for larger sediment thickness values.

C. Example 3: Spatial tracking using a HLA

The final example uses a HLA towed together with the source to map the spatially evolving environment. The configuration in Fig. 1(b) is used with a HLA of 254 m and a distance of 300 m from the source.[1] A nonoverlapping spatial partitioning with each step k representing 500 m is selected. Since the source and HLA are close to each other, the complex normal mode code ORCA capable of computing the near field is used as the forward model. The simulation parameters different from the previous examples are summarized in Table V. Again the seven-parameter environment in Fig. 2 is used. To compare the effects of different configurations on an identical geoacoustic tracking problem, the evolution of the environmental parameters is the same as in the previous example.

A typical track for each type of filter is shown in Fig. 9. Note how the tracking capabilities of the filters for individual parameters change from the previous long-range VLA configuration to the short-range HLA configuration used here. Geoacoustic tracking behaves very similarly to previous studies comparing geoacoustic inversions using HLA versus VLA in that a parameter that is not readily estimated by geoacoustic inversion will also be poorly tracked.[1,5,35] The major difference of a source close to the receiver is the ability of the receiver to detect higher order modes with large incidence angles that can penetrate deeper into the sediment since the signal does not propagate enough to attenuate these fields. This means that the field across the HLA is much more sensitive to some of the sediment and bottom parameters such as the bottom sound speed and sediment thickness. Notice how all four filters are able to track, in general, the

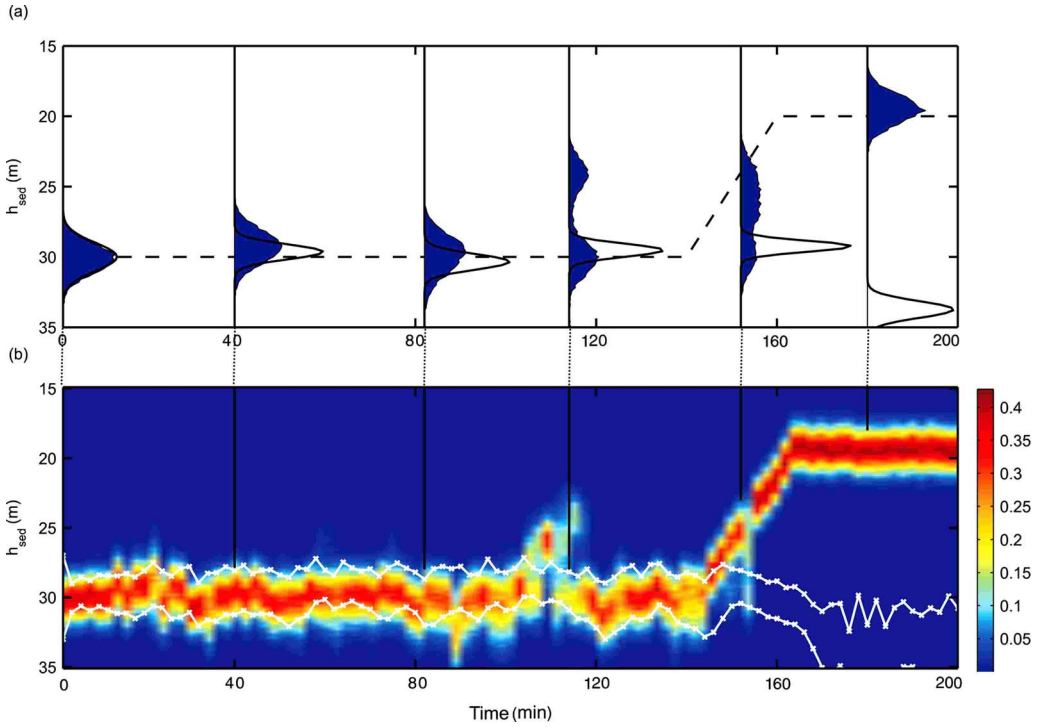


FIG. 8. (Color online) Example 2: Posterior probability density evolution $p(\mathbf{x}_{k|k})$ for the sediment thickness h_{sed} for a 10000-point particle filter and the EKF: (a) Six snapshots at $t=0$, 40, 84, 116, 152, and 180 min with the local particle histograms representing the PF distribution (solid) and the EKF Gaussian PDF (line) along with the true trajectory (dashed). (b) The continuous evolution of the PF PPD together with the local mean ± 3 standard deviation ($\hat{\mathbf{x}}_{k|k} \pm 3\sqrt{\mathbf{P}_{k|k}}$) of the EKF Gaussian.

bottom sound speed, sediment thickness, sediment sound speed, and density both in slowly and fast changing environments. Since the field is not that sensitive to the attenuation, it is now a relatively poorly determined parameter and the EKF fails to track it, while the UKF and PF-5000 are able to maintain the track, albeit a noisy one. Similarly, the filters are unable to track the top sound speed value most of the time. Only PF is able to track this parameter on occasion.

The improvement percentages of the filters are obtained by repeating the track using 100 MC realizations. The results are given in Table V. The improvement of the UKF over EKF is similar to the previous example with 25% and 33% for slowly and fast changing regions, respectively. PF-200 performs poorly due to an insufficient number of particles used in tracking. On the other hand, the PF-5000 outperforms the EKF by 60%.

V. DISCUSSION

It is possible to extend the state space from just the environmental parameters by appending other parameters-of-interest such as the source range, depth, and speed. Also a single frequency is used throughout the paper. However, multiple frequencies are frequently employed for geoacoustic inversion due to the varying levels of sensitivities to different frequencies and robustness. It is possible to include multiple frequencies by appending the array data at different frequencies forming a long measurement vector $\mathbf{y}_k$ and a forward model $\mathbf{h}(\mathbf{x}_k)$ composed of multiple normal mode runs at different frequencies.

The filter performance strongly depends on where $\mathbf{x}$ is in the state space. The most common scenario is where the performance improves from the EKF to the UKF to a PF with enough particles. However, there are regions in the state space where the KFs give better tracking results depending on the local linearity of the forward model and the Gaussian nature of the densities involved.

Although not given here, there are some special cases in geoacoustic tracking that can result in track divergence. One example observed during spatial tracking using the HLA configuration (example 3) is when a layer gets thin and then

TABLE V. Simulation parameters and percent improvement of filters for example 3.

Simulation parameters				Method	% Imp. over EKF	
					35 km	50 km
Source depth	20 m	Source frequency	100 Hz			
Receiver type	HLA	Array SNR ($\mathbf{R}_k$)	40 dB	EKF	0	0
Receiver depth	26 m	Array start, Δr	300 m, 2 m	UKF	25	33
No. of hydrophones	128	Track length	50 km ($k=100$)	PF-200	−12	−73
MC runs	100	Track frequency	1 meas./500 m	PF-5000	60	64

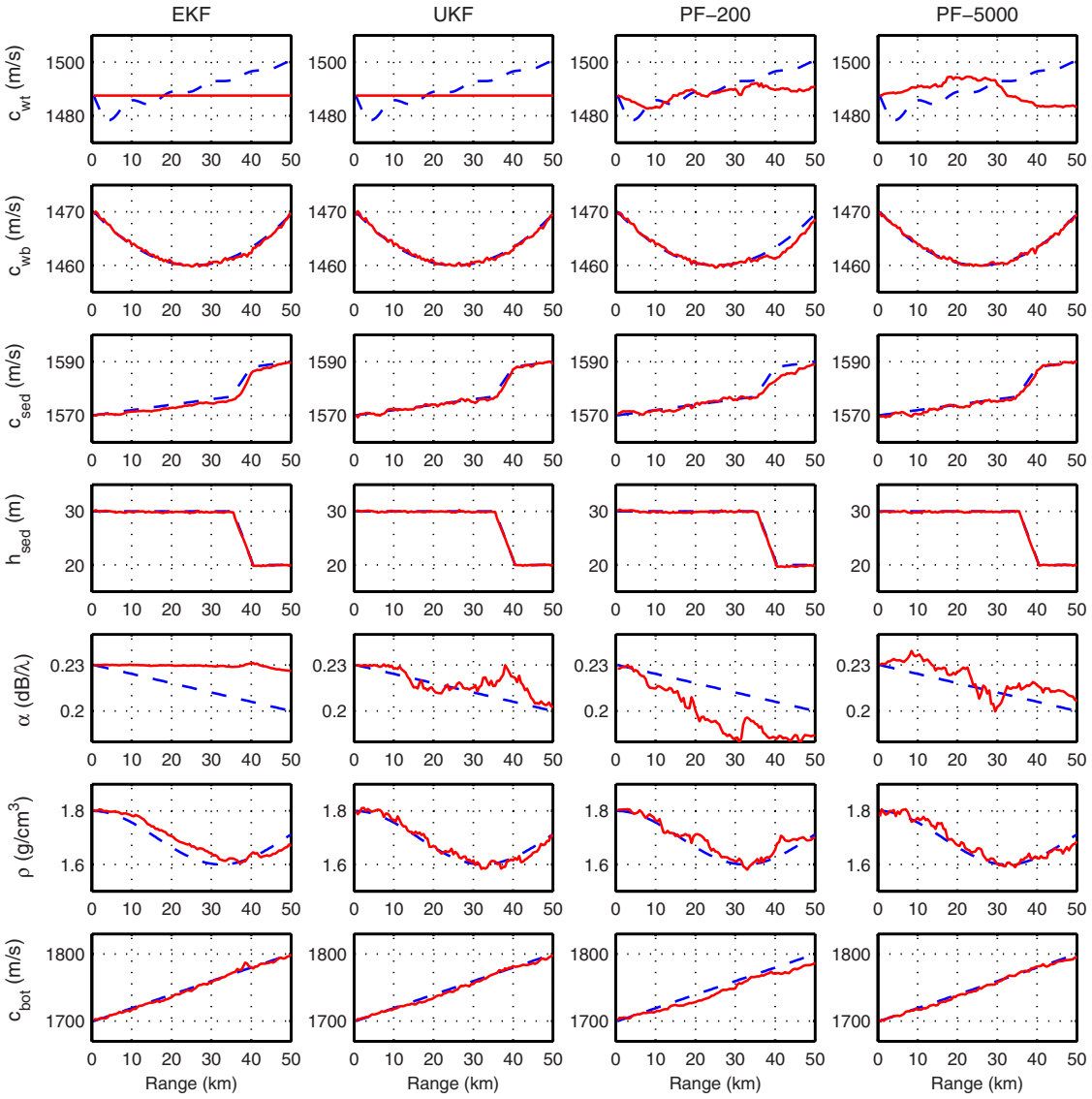


FIG. 9. (Color online) Example 3: Tracking results of EKF, UKF, PF-200, and PF-5000 for the seven-parameter environment given in Fig. 1 using the short range HLA configuration. True trajectories (dashed) are provided along with the tracking filter estimates (solid).

thickens again. When the layer gets thin, other parameters such as the sound speed, attenuation, and density characterizing the layer have little or no effect on the acoustic field across the array, temporarily making the field insensitive to that layer's parameters. This results in deviations from their true values for these parameters, and when the layer starts to thicken again the filters diverge since the starting points for the layer parameters other than the layer thickness are too far from their current true values.

Another case is the coupling between the parameters. When the sediment thickness increases, less signal reaches the bottom layer, resulting in degrading performance of these bottom parameters similar to the previous case and may cause divergence as the sediment gets less thick again. In general, PFs show more robust tracking under such conditions.

Also the seabed can have spatial layer changes. While one sediment layer and a semi-infinite bottom are adequate initially, a second sediment layer can form. Or the sediment type becomes sand, whereas the model given to the PF assumed that the region is clay, limiting the possible parameter values via priors. Such environments can be tracked using multiple environmental models, one for each possible scenario. This will require Gaussian sum filters such as the interactive multiple model EKF/UKF that involves a filter bank composed of multiple KFs running in parallel for each possible model.[29,36] Similarly, this can be accomplished using their PF counterpart, the MMPF.[37]

One interesting observation from the simulations is the KFs ability to continue to track some parameters while other parameters diverge and can only be tracked by the PF. This means that the marginal densities for these parameters are close to Gaussian and the measurement equations connecting those parameters to the acoustic field are close to being linear. This is unlike many other tracking problems such that when one parameter starts to diverge, so do all the others, usually resulting in a total divergence. However, there are many cases where such marginal Gaussian densities occur. In these cases, one common approach is to use a Rao–Blackwellized particle filter also known as the marginalized

particle filter that groups the state parameters into linear/Gaussian and nonlinear/non-Gaussian ones and uses a mixed EKF/PF approach, reducing the dimension of the state space that the PF has to sample, which, in return, reduces significantly the required number of particles for a desired accuracy.[29,38]

VI. SUMMARY

Tracking of geoacoustic environmental parameters has been addressed. Spatial and temporal evolutions of the water column and seabed parameters were estimated using EKFs, UKFs, and PFs with acoustic measurements as inputs. These tracking filters enabled providing real-time, continuously updated estimates of the geoacoustic parameters and their uncertainties, requiring far fewer forward model runs compared to alternatives such as successively running geoacoustic inversion algorithms.

This paper investigated how the three filters behaved for the nonlinear, non-Gaussian geoacoustic tracking problem using three examples with both the VLA and HLA simulated data. An efficient way of computing the local PCRLB to compute the filter efficiencies was shown. The results showed that all three filters performed well in geoacoustic applications. It was found that a PF with enough particles could typically achieve 80% filter efficiency in geoacoustic tracking while providing PPD evolutions for the environmental parameters. Even though KFs had less efficiency and high divergence rates and were unable to track some parameters while the PF was still able to maintain track, they also showed robust tracking in many cases. Since they are computationally very fast compared to the PF, they can be used in many applications where the performances are similar. The UKF outperformed the EKF in most of the simulations, but the improvement-over-EKF values of the UKF were modest compared to the PF. The PF was able to maintain track in environments that include sudden changes such as the sediment thickness. The two KFs used here showed mixed success in tracking sudden jumps in the parameter values.

PFs proved to be very promising in the nonlinear, non-Gaussian geoacoustic tracking problem. It was shown that the performance could degrade below that of the EKF if a small number of particles were used. However, in this paper, the PF with enough particles showed robust tracking in a number of cases involving different measurement configurations that use HLA and VLA data, slowly and quickly changing environments, and environmental parameters with relatively flat sensitivity curves. The limitations of all three filters were discussed using an example of tracking a quickly changing environment with parameters having medium to totally flat sensitivity curves.

ACKNOWLEDGMENT

This work was supported by the Office of Naval Research under Grant No. N00014-05-1-0264.

APPENDIX A: FILTER EQUATIONS

1. Extended Kalman filter

The first filter choice is the EKF (Ref. 22). Since KF is the best possible linear tracking filter, its extended version that can operate on nonlinear systems can still be near optimal. The EKF works by converting the system into a form over which the KF can operate. This is done by locally linearizing the equations using the first terms in the Taylor series expansions of the nonlinear transformations (such as the normal mode code in $\mathbf{h}$) and assuming that the nonlinearities are small so that EKF will perform well. Once the equations are linearized, starting with a Gaussian PDF for $\mathbf{x}_0$ will ensure that the evolving parameters will remain Gaussian, and it is necessary to propagate only the mean and covariance as in the KF. However, due to this approximation, the EKF cannot claim the optimality enjoyed by the KF for linear-Gaussian systems. The EKF has been implemented successfully in a large number of applications such as radar and sonar target tracking applications, and its speed and ease of implementation make the EKF the filter of choice.

In geoacoustic tracking, the complex source magnitude s_k is usually not known. Therefore, the EKF equations are modified by inserting a maximum likelihood (ML) estimator that estimates the source every time the forward normal $\mathbf{d}(\mathbf{x}_k)$ is run.[10] This is done by writing the likelihood function at step k as

$$\mathcal{L}(\mathbf{x}_k) = \frac{1}{\pi^{n_H}|\mathbf{R}|} \exp[-(\mathbf{y}_k - s_k\mathbf{d}(\mathbf{x}_k))^H \mathbf{R}^{-1}(\mathbf{y}_k - s_k\mathbf{d}(\mathbf{x}_k))], \tag{A1}$$

where n_H is the number of hydrophones. Assuming that the complex Gaussian noise $\mathbf{w}_k$ is uncorrelated with the same variance along the array ($\mathbf{R}=\nu\mathbf{I}$),

$$\mathcal{L}(\mathbf{x}_k) = \frac{1}{(\pi\nu)^{n_H}} \exp\left(-\frac{\|\mathbf{y}_k - s_k\mathbf{d}(\mathbf{x}_k)\|^2}{\nu}\right). \tag{A2}$$

The ML estimate for the source s_k is then obtained by solving for $\partial\mathcal{L}/\partial s_k = 0$, giving

$$\hat{\mathbf{s}}_k = \frac{\mathbf{d}(\mathbf{x}_k)^H \mathbf{y}_k}{\|\mathbf{d}(\mathbf{x}_k)\|^2}. \tag{A3}$$

This source estimate is used in the following EKF equations both for the calculation of $\mathbf{h}$ and during the linearization of $\mathbf{h}$ to obtain the matrix $\mathbf{H}$:

$$\hat{\mathbf{x}}_{k|k-1} = \mathbf{F}\hat{\mathbf{x}}_{k-1|k-1}, \tag{A4}$$

$$\mathbf{P}_{k|k-1} = \mathbf{Q}_{k-1} + \mathbf{F}\mathbf{P}_{k-1|k-1}\mathbf{F}^T, \tag{A5}$$

$$\hat{\mathbf{x}}_{k|k} = \hat{\mathbf{x}}_{k|k-1} + \mathbf{K}_k(\mathbf{y}_k - \mathbf{h}(\hat{\mathbf{x}}_{k|k-1})), \tag{A6}$$

$$\mathbf{P}_{k|k} = \mathbf{P}_{k|k-1} - \mathbf{K}_k\mathbf{S}_k\mathbf{K}_k^T, \tag{A7}$$

where

$$\mathbf{S}_k = \hat{\mathbf{H}}_k\mathbf{P}_{k|k-1}\hat{\mathbf{H}}_k^T + \mathbf{R}_k, \tag{A8}$$

$$\mathbf{K}_k = \mathbf{P}_{k|k-1}\hat{\mathbf{H}}_k^T\mathbf{S}_k^{-1}, \tag{A9}$$

$$\mathbf{h}(\hat{\mathbf{x}}_{k|k-1}) = \frac{\mathbf{d}(\hat{\mathbf{x}}_{k|k-1})^H \mathbf{y}_k \mathbf{d}(\hat{\mathbf{x}}_{k|k-1})}{\|\mathbf{d}(\hat{\mathbf{x}}_{k|k-1})\|^2}, \tag{A10}$$

$$\hat{\mathbf{H}}_k = [\nabla_{\mathbf{x}_k} \mathbf{h}^T(\hat{\mathbf{x}}_{k|k-1})]^T. \tag{A11}$$

Equations (A4) and (A5) are the prediction steps that give the environmental model estimate $\hat{\mathbf{x}}_k$ and its associated uncertainty in terms of the covariance matrix $\mathbf{P}_k$ at step index k given the previous history $\{\mathbf{x}_0, \mathbf{x}_1, \ldots, \mathbf{x}_{k-1}\}$, Eqs. (A6) and (A7) are the correction equations that give $\hat{\mathbf{x}}_k$ and $\mathbf{P}_k$ at step index k given its previous history $\{\mathbf{x}_0, \mathbf{x}_1, \ldots, \mathbf{x}_{k-1}\}$ and the set of measurements $\{\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_k\}$, and $\mathbf{K}$ is the Kalman gain.

Note that the insertion of the ML estimate of the source in Eq. (A3) into the Kalman update equation violates the Kalman formulation. This is true for Eq. (A6) where the Kalman gain is applied to the measured minus predicted data $(\mathbf{y}_k - \mathbf{y}_{\text{pred}})$ since the predicted data include $\mathbf{y}_k$ itself due to the ML source estimate in Eq. (A3). However, the ML estimator in Eq. (A3) simply normalizes the amplitude of the predicted data so that the acoustic field variation across the array is compared, not the actual amplitudes, eliminating the effects of the unknown source amplitude. Moreover, the averaging inherent in the inner product in Eq. (A3) over the array elements makes the source estimate less noisy and more robust relative to the environmental parameters. Finally, the performance calculations of the KFs used here are not affected since the synthetic data enable us to compute the true filter RMS error $E[(\hat{\mathbf{x}}_{k|k} - \mathbf{x}_{k|k})^2]$ instead of the conventional performance metric for the KF (covariance matrix $\mathbf{P}_{k|k}$). An alternative approach would be to include the unknown source term into the state model $\mathbf{x}_k$. However, this will increase the dimension of the state space for a nuisance parameter in which we are not interested.

2. Unscented Kalman filter

The analytical linearization used in the EKF results in poor estimates of the mean and covariance as the nonlinearity in the forward model increases. To mitigate this the UKF[23,39] has been introduced. Instead of analytical linearization, the UKF uses a concept called statistical linearization in which the filter enforces Gaussianity and keeps the nonlinearity. Enforcing Gaussian PDFs enables the filter to carry all the necessary information by propagating only the mean and covariance as does the KF. This is achieved by the UT that enables the propagation of the mean and variance through nonlinear functions. The UKF represents initial densities using only a few predetermined particles called sigma points. These points are chosen deterministically by the UT algorithm, and they describe accurately the mean and covariance of a PDF. As the random variable undergoes a nonlinear transformation, these points are propagated through the nonlinear function and used to reconstruct the new mean and covariance using the UT weights. Hence, unlike the EKF, they can compute accurately the mean and covariance to at least second order (third if the initial PDF is Gaussian) of the nonlinearity.

Similar to the EKF, the UKF algorithm used here incorporates a ML estimator for the unknown source term. The UKF uses the following recursive formulation where $2n_x + 1$ sigma points $\{\mathcal{X}^i\}_{i=0}^{2n_x}$ and their corresponding weights W^i are generated and used with the UT algorithm to perform the mean $(\hat{\mathbf{x}}_k)$ and covariance $(\mathbf{P}_k)$ calculations required in the Kalman framework. The UT weights are given in terms of the scaling parameter $\lambda = \alpha^2(n_x + \kappa) - n_x$ and prior knowledge parameter β, where α is used to control the spread of the sigma points around the mean and κ is the secondary scaling parameter. α, β, and κ are taken as 0.1, 2, and 0, respectively.

UT weights and sigma points are generated using

$$\mathcal{X}_{k-1}^0 = \hat{\mathbf{x}}_{k-1|k-1},$$

$$W_m^0 = \frac{\lambda}{n_x + \lambda}, \quad W_{\text{cov}}^0 = W_m^0 + \beta + 1 - \alpha^2, \tag{A12}$$

$$\mathcal{X}_{k-1}^i = \hat{\mathbf{x}}_{k-1|k-1} \pm (\sqrt{(n_x + \kappa)\mathbf{P}_{k-1|k-1}})_i,$$

$$W_m^i = W_{\text{cov}}^i = \frac{0.5}{n_x + \lambda}, \quad i = 1, 2, \ldots, 2n_x, \tag{A13}$$

where $(\sqrt{.})_i$ is the ith column of the matrix square root. The prediction step is composed of

$$\mathcal{X}_{k|k-1}^i = \mathbf{F}\mathcal{X}_{k-1}^i, \quad \mathcal{Y}_{k|k-1}^i = \frac{\mathbf{d}(\mathcal{X}_{k|k-1}^i)^H \mathbf{y}_k \mathbf{d}(\mathcal{X}_{k|k-1}^i)}{\|\mathbf{d}(\mathcal{X}_{k|k-1}^i)\|^2},$$

$$\hat{\mathbf{x}}_{k|k-1} = \sum_{i=0}^{2n_x} W_m^i \mathcal{X}_{k|k-1}^i, \quad \hat{\mathbf{y}}_{k|k-1} = \sum_{i=0}^{2n_x} W_m^i \mathcal{Y}_{k|k-1}^i,$$

$$\mathbf{P}_{k|k-1} = \mathbf{Q}_{k-1} + \sum_{i=0}^{2n_x} W_{\text{cov}}^i [\mathcal{X}_{k|k-1}^i - \hat{\mathbf{x}}_{k|k-1}][\mathcal{X}_{k|k-1}^i - \hat{\mathbf{x}}_{k|k-1}]^T, \tag{A14}$$

and the update step uses

$$\mathbf{P}_{xy} = \sum_{i=0}^{2n_x} W_{\text{cov}}^i [\mathcal{X}_{k|k-1}^i - \hat{\mathbf{x}}_{k|k-1}][\mathcal{Y}_{k|k-1}^i - \hat{\mathbf{y}}_{k|k-1}]^T,$$

$$\mathbf{P}_{yy} = \sum_{i=0}^{2n_x} W_{\text{cov}}^i [\mathcal{Y}_{k|k-1}^i - \hat{\mathbf{y}}_{k|k-1}][\mathcal{Y}_{k|k-1}^i - \hat{\mathbf{y}}_{k|k-1}]^T,$$

$$\mathbf{K}_k = \mathbf{P}_{xy}(\mathbf{P}_{yy} + \mathbf{R}_k)^{-1}, \tag{A15}$$

$$\hat{\mathbf{x}}_{k|k} = \hat{\mathbf{x}}_{k|k-1} + \mathbf{K}_k(\mathbf{y}_k - \hat{\mathbf{y}}_{k|k-1}), \tag{A16}$$

$$\mathbf{P}_{k|k} = \mathbf{P}_{k|k-1} - \mathbf{K}_k(\mathbf{P}_{yy} + \mathbf{R}_k)\mathbf{K}_k^T. \tag{A17}$$

Although it is fast relative to more advanced techniques, derivative-free, and an improvement over the EKF, there still are two weaknesses. The first is that the nonlinearity may be so severe that it may require an even higher order accuracy than the UKF can provide to correctly capture the mean and

covariance. The other is that the densities may be highly non-Gaussian so that the first two moments will not be sufficient even if they can be calculated correctly.

3. Particle filter

The third algorithm used in this paper is the SMC commonly known as the PF.[24] Rapid increases in the available computational power have made the PF very popular for many nonlinear, non-Gaussian tracking problems. Unlike the Kalman framework, neither Gaussian nor linearity assumptions are necessary for the PF. However, this means that propagating only the mean and covariance is not sufficient anymore. Instead, the PF propagates an ensemble of particles to represent the densities. These particles are selected randomly by MC runs. Compared with the sigma points of the UKF, a much larger number of particles are needed to represent the PDF. Therefore, the PF can perform much better than its KF variants, but it does this with an order of magnitude increase in the required computational resources. There are many different variants[29] of the PF such as the regularized particle filter, Markov chain MC step PF, and auxiliary and classical SIR PFs. The SIR (Ref. 40) algorithm is used throughout this work. Normally, degeneracy can be a problem for the SIR algorithm, especially for low process noise systems. However, due to the environmental uncertainty in the model, $\mathbf{Q}_k$ is selected to be relatively large, thus mostly eliminating the need for more complex PFs with improved sample diversity.

The SIR algorithm uses n_p particles $\{\chi_k^i\}_{i=1}^{n_p}$ to represent the PDF at each step k. The filter has the predict and update sections just as in a KF, but the SIR filter will use these sections to propagate the particles instead of mean and covariance calculations. The initial set of particles $\{\chi_0^i\}_{i=1}^{n_p}$ are sampled from the prior $p(\mathbf{x}_0)$. The SIR filter uses the importance sampling[41] density as the transitional prior $p(\mathbf{x}_k|\mathbf{x}_{k-1})$. Although this is a suboptimal choice, it is easy to sample from this density. This selection results in particle weights proportional to the likelihood $W_k \propto p(\mathbf{y}_k|\mathbf{x}_k)$.

The prediction step consists of sampling from the prior. Then the normalized weight W_k^i of each particle is calculated from its likelihood function. As with the KFs, the source term is estimated with a ML estimator during the likelihood calculation of each particle in the ensemble. The update step includes the resampling section where a new set of n_p particles is generated from the parent set according to the weights of the parent particles, with high likelihood particles generating more particles than the low likelihood ones. Hence, a single iteration of the recursive SIR algorithm can be summarized as

$$\{\mathcal{X}_{k|k-1}^i\}_{i=1}^{n_p} \sim p(\mathbf{x}_k|\mathbf{x}_{k-1}), \tag{A18}$$

$$W_k^i = \frac{p(\mathbf{y}_k|\mathcal{X}_{k|k-1}^i)}{\Sigma_{i=1}^{n_p} p(\mathbf{y}_k|\mathcal{X}_{k|k-1}^i)}, \tag{A19}$$

$$\{\mathcal{X}_{k|k}^i\}_{i=1}^{n_p} = \text{Resample}[W_k^j, \{\mathcal{X}_{k|k-1}^j\}_{j=1}^{n_p}],$$

$$\text{such that } \Pr\{\mathcal{X}_{k|k}^i = \mathcal{X}_{k|k-1}^j\} = W_k^j \tag{A20}$$

APPENDIX B: POSTERIOR CRAMÉR–RAO LOWER BOUND

One issue with the tracking problems is that the computation of the full PCRLB is not feasible. Unlike geoacoustic inversion where there is a fixed number (n_x) of random variables in the model vector $\mathbf{x}$, geoacoustic tracking introduces n_x new random variables with every new step k. Therefore, we will use a $(n_x \times n_x)$ matrix PCRLB_k instead of the full PCRLB matrix. PCRLB_k is defined as the inverse of the filtering information matrix $\mathbf{J}_k$ so that the MSE of any filter estimate at tracking step index k will be bounded as

$$E\{(\hat{\mathbf{x}}_{k|k} - \mathbf{x}_k)(\hat{\mathbf{x}}_{k|k} - \mathbf{x}_k)^T\} \geqslant \mathbf{J}_k^{-1}. \tag{B1}$$

A computationally efficient way of computing this PCRLB recursively for discrete-time nonlinear filtering problems is given in Ref. 42,

$$\mathbf{J}_k = \mathbf{D}_{k-1}^{22} - [\mathbf{D}_{k-1}^{12}]^T(\mathbf{J}_{k-1} + \mathbf{D}_{k-1}^{11})^{-1}\mathbf{D}_{k-1}^{12}, \tag{B2}$$

where

$$\mathbf{D}_{k-1}^{11} = -\mathbf{E}\{\nabla_{\mathbf{x}_{k-1}}[\nabla_{\mathbf{x}_{k-1}} \log p(\mathbf{x}_k|\mathbf{x}_{k-1})]^T\}, \tag{B3}$$

$$\mathbf{D}_{k-1}^{12} = -E\{\nabla_{\mathbf{x}_k}[\nabla_{\mathbf{x}_{k-1}} \log p(\mathbf{x}_k|\mathbf{x}_{k-1})]^T\}, \tag{B4}$$

$$\mathbf{D}_{k-1}^{22} = -E\{\nabla_{\mathbf{x}_k}[\nabla_{\mathbf{x}_k} \log p(\mathbf{x}_k|\mathbf{x}_{k-1})]^T\} - E\{\nabla_{\mathbf{x}_k}[\nabla_{\mathbf{x}_k} \log p(\mathbf{y}_k|\mathbf{x}_k)]^T\}. \tag{B5}$$

It is important to note that the computations only require $(n_x \times n_x)$ matrices, and the computation cost is independent of the step index k. The geoacoustic tracking problem with the system of equations defined in Eqs. (2) and (3) has a linear state equation, and both of the random noise sequences $\mathbf{v}$ and $\mathbf{w}$ are additive and Gaussian. Therefore, the above equations can be reduced to[43]

$$\mathbf{D}_{k-1}^{11} = \mathbf{F}^T\mathbf{Q}_{k-1}^{-1}\mathbf{F},$$

$$\mathbf{D}_{k-1}^{12} = -\mathbf{F}^T\mathbf{Q}_{k-1}^{-1}, \tag{B6}$$

$$\mathbf{D}_{k-1}^{22} = \mathbf{Q}_{k-1}^{-1} + E\{\mathbf{H}_k^T\mathbf{R}_k^{-1}\mathbf{H}_k\}, \tag{B7}$$

where $\mathbf{H}_k$ is the jacobian of $\mathbf{h}(\mathbf{x})$ computed similar to Eqs. (A10) and (A11) at its true value $\mathbf{x}_k$. Unfortunately the expectation in Eq. (B7) has to be evaluated numerically using a MC analysis. $\mathbf{D}_{k-1}^{22}$ is computed as

$$\mathbf{D}_{k-1}^{22} = \mathbf{Q}_{k-1}^{-1} + \frac{1}{n_{\text{MC}}}\sum_{j=1}^{n_{\text{MC}}} \nabla\mathbf{h}(\mathbf{x}_k^j)\mathbf{R}_k^{-1}[\nabla\mathbf{h}(\mathbf{x}_k^j)]^T, \tag{B8}$$

where n_{MC} is the number of MC trajectories, assuming a Gaussian prior PDF with a covariance matrix $\mathbf{P}_0$. The recursion in Eq. (B2) is initiated with

$$\mathbf{J}_0 = -E\{\nabla_{\mathbf{x}_0}[\nabla_{\mathbf{x}_0} \log p(\mathbf{x}_0)]^T\} = \mathbf{P}_0^{-1}. \tag{B9}$$

Once the PCRLB, the inverse of $\mathbf{J}_k$ in Eq. (B2), is computed, the filters can be compared with each other and the CRLB.

[1]M. Siderius, P. L. Nielsen, and P. Gerstoft, "Range-dependent seabed characterization by inversion of acoustic data from a towed receiver array," J. Acoust. Soc. Am. **112**, 1523–1535 (2002).

[2]D. P. Knobles, R. A. Koch, L. A. Thompson, K. C. Focke, and P. E. Eisman, "Broadband sound propagation in shallow water and geoacoustic inversion," J. Acoust. Soc. Am. **113**, 205–222 (2003).
[3]D. Battle, P. Gerstoft, W. A. Kuperman, W. S. Hodgkiss, and M. Siderius, "Geoacoustic inversion of tow-ship noise via near-field-matched-field processing," IEEE J. Ocean. Eng. **28**, 454–467 (2003).
[4]D. Battle, P. Gerstoft, W. S. Hodgkiss, W. A. Kuperman, and P. L. Nielsen, "Bayesian model selection applied to self-noise geoacoustic inversion," J. Acoust. Soc. Am. **116**, 2043–2056 (2004).
[5]M. R. Fallat, P. L. Nielsen, S. E. Dosso, and M. Siderius, "Geoacoustic characterization of a range-dependent ocean environment using towed array data," IEEE J. Ocean. Eng. **30**, 198–206 (2005).
[6]D. Tollefsen, S. E. Dosso, and M. J. Wilmut, "Matched-field geoacoustic inversion with a horizontal array and low-level source," J. Acoust. Soc. Am. **120**, 221–230 (2006).
[7]T. C. Yang, K. Yoo, and L. Y. Fialkowski, "Subbottom profiling using a ship towed line array and geoacoustic inversion," J. Acoust. Soc. Am. **122**, 3338–3352 (2007).
[8]P. Gerstoft, "Inversion of seismoacoustic data using genetic algorithms and *a posteriori* probability distributions," J. Acoust. Soc. Am. **95**, 770–782 (1994).
[9]M. D. Collins, W. A. Kuperman, and H. Schmidt, "Nonlinear inversion for ocean-bottom properties," J. Acoust. Soc. Am. **92**, 2770–2783 (1992).
[10]P. Gerstoft and C. F. Mecklenbräuker, "Ocean acoustic inversion with estimation of *a posteriori* probability distributions," J. Acoust. Soc. Am. **104**, 808–819 (1998).
[11]S. E. Dosso, "Quantifying uncertainty in geoacoustic inversion. I. A fast Gibbs sampler approach," J. Acoust. Soc. Am. **111**, 129–142 (2002).
[12]C.-F. Huang, P. Gerstoft, and W. S. Hodgkiss, "Validation of statistical estimation of transmission loss in the presence of geoacoustic inversion uncertainty," J. Acoust. Soc. Am. **120**, 1932–1941 (2006).
[13]J. V. Candy and E. J. Sullivan, "Sound velocity profile estimation: A system theoretic approach," IEEE J. Ocean. Eng. **18**, 240–252 (1993).
[14]O. Carrière, J.-P. Hermand, J.-C. L. Gac, and M. Rixen, "Full-field tomography and Kalman tracking of the range-dependent sound speed field in a coastal water environment," J. Mar. Syst. In press (2008).
[15]I. Zorych and Z.-H. Michalopoulou, "Particle filtering for dispersion curve tracking in ocean acoustics," J. Acoust. Soc. Am. **124**, EL45–EL50 (2008).
[16]A. A. Ganse and R. I. Odom, "Adapting results in filtering theory to inverse theory, to address the statistics of nonlinear geoacoustic inverse problems," J. Acoust. Soc. Am. **120**, 3357–3357 (2006).
[17]F. B. Jensen and M. C. Ferla, "SNAP: The SACLANTCEN normal-mode acoustic propagation model," SACLANT Undersea Research Centre, La Spezia, Italy, 1979.
[18]E. K. Westwood, C. T. Tindle, and N. R. Chapman, "A normal mode model for acoustoelastic ocean environments," J. Acoust. Soc. Am. **100**, 3631–3645 (1996).
[19]S. E. Dosso, P. L. Nielsen, and M. J. Wilmut, "Data error covariance in matched-field geoacoustic inversion," J. Acoust. Soc. Am. **119**, 208–219 (2006).
[20]C.-F. Huang, P. Gerstoft, and W. S. Hodgkiss, "Uncertainty analysis in matched-field geoacoustic inversions," J. Acoust. Soc. Am. **119**, 197–207 (2006).
[21]Y.-M. Jiang and N. R. Chapman, "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).
[22]S. M. Kay, *Fundamentals of Statistical Signal Processing—Volume I: Estimation Theory* (Prentice-Hall, NJ, 1993).
[23]S. Julier, J. Uhlmann, and H. F. Durrant-White, "A new method for nonlinear transformation of means and covariances in filters and estimators," IEEE Trans. Autom. Control **45**, 477–482 (2000).
[24]A. Doucet, N. de Freitas, and N. Gordon, *Sequential Monte Carlo Methods in Practice* (Springer, New York, 2001).
[25]K. Becker, S. Rajan, and G. Frisk, "Results from the geoacoustic inversion techniques workshop using modal inverse methods," IEEE J. Ocean. Eng. **28**, 331–341 (2003).
[26]J. Dettmer, S. E. Dosso, and C. W. Holland, "Uncertainty estimation in seismo-acoustic reflection travel time inversion," J. Acoust. Soc. Am. **122**, 161–176 (2007).
[27]M. Siderius, C. H. Harrison, and M. B. Porter, "A passive fathometer technique for imaging seabed layering using ambient noise," J. Acoust. Soc. Am. **120**, 1315–1323 (2006).
[28]J.-P. Hermand and P. Gerstoft, "Inversion of broad-band multitone acoustic data from the yellow shark summer experiments," IEEE J. Ocean. Eng. **21**, 324–346 (1996).
[29]B. Ristic, S. Arulampalam, and N. Gordon, *Beyond the Kalman Filter, Particle Filters for Tracking Applications* (Artech House, Boston, 2004).
[30]H. L. van Trees, *Detection, Estimation and Modulation Theory* (J Wiley, New York, 1968).
[31]M. Zanolin, I. Ingram, A. Thode, and N. C. Makris, "Asymptotic accuracy of geoacoustic inversions," J. Acoust. Soc. Am. **116**, 2031–2042 (2004).
[32]W. Xu, A. B. Baggeroer, and C. D. Richmond, "Bayesian bounds for matched-field parameter estimation," IEEE Trans. Signal Process. **52**, 3293–3305 (2004).
[33]C. Yardim, P. Gerstoft, and W. S. Hodgkiss, "Estimation of radio refractivity from radar clutter using Bayesian Monte Carlo analysis," IEEE Trans. Antennas Propag. **54**, 1318–1327 (2006).
[34]C. Yardim, P. Gerstoft, and W. S. Hodgkiss, "Statistical maritime radar duct estimation using a hybrid genetic algorithm—Markov chain Monte Carlo method," Radio Sci. **42**, 1–15 (2007).
[35]D. Tollefsen and S. E. Dosso, "Geoacoustic information content of horizontal line array data," IEEE J. Ocean. Eng. **32**, 651–662 (2007).
[36]Y. Bar-Shalom, X. R. Li, and T. Kirubarajan, *Estimation with Applications to Tracking and Navigation* (Wiley, New York, 2001).
[37]S. McGinnity and G. W. Irwin, "Multiple model bootstrap filter for maneuvering target tracking," IEEE Trans. Aerosp. Electron. Syst. **36**, 1006–1012 (2000).
[38]T. Schön, F. Gustafsson, and P.-J. Nordlund, "Marginalized particle filters for mixed linear/nonlinear state-space models," IEEE Trans. Signal Process. **53**, 2279–2289 (2005).
[39]E. A. Wan and R. van der Merve, "The unscented Kalman filter," in *Kalman Filtering and Neural Networks*, edited by S. Haykin (Wiley, New York, 2001).
[40]N. J. Gordon, D. J. Salmond, and A. F. M. Smith, "Novel approach to nonlinear/non-Gaussian Bayesian state estimation," IEE Proc. F, Radar Signal Process. **140**, 107–113 (1993).
[41]J. J. K. Ó Ruanaidh and W. J. Fitzgerald, *Numerical Bayesian Methods Applied to Signal Processing*, Statistics and Computing Series (Springer-Verlag, New York, 1996).
[42]P. Tichavský, C. H. Muravchik, and A. Nehorai, "Posterior Cramér-Rao bounds for discrete-time nonlinear filtering," IEEE Trans. Signal Process. **46**, 1386–1396 (1998).
[43]C. Yardim, P. Gerstoft, and W. S. Hodgkiss, "Tracking refractivity from clutter using Kalman and particle filters," IEEE Trans. Antennas Propag. **56**, 1058–1070 (2008).

Blind inversion method using Lamb waves for the complete elastic property characterization of anisotropic plates

J. Vishnuvardhan, C. V. Krishnamurthy, and Krishnan Balasubramaniam[a)]
Center for Nondestructive Evaluation and Department of Mechanical Engineering, Indian Institute of Technology, Chennai 600 036, India

(Received 14 February 2008; revised 27 October 2008; accepted 12 November 2008)

A novel blind inversion method using Lamb wave S_0 and A_0 mode velocities is proposed for the complete determination of elastic moduli, material symmetries, as well as principal plane orientations of anisotropic plates. The approach takes advantage of genetic algorithm, introduces the notion of "statistically significant" elastic moduli, and utilizes their sensitivities to velocity data to reconstruct the elastic moduli. The unknown material symmetry and the principal planes are then evaluated using the method proposed by Cowin and Mehrabadi [Q. J. Mech. Appl. Math. **40**, 451–476 (1987)]. The blind inversion procedure was verified using simulated ultrasonic velocity data sets on materials with transversely isotropic, orthotropic, and monoclinic symmetries. A modified double ring configuration of the single transmitter and multiple receiver compact array was developed to experimentally validate the blind inversion approach on a quasi-isotropic graphite-epoxy composite plate. This technique finds application in the area of material characterization and structural health monitoring of anisotropic platelike structures.
 [DOI: 10.1121/1.3050253]

PACS number(s): 43.35.Cg, 43.38.Hz, 43.40.Le, 43.40.Sk [PEB] Pages: 761–771

I. INTRODUCTION

Structural health monitoring (SHM) can be defined as the continuous process of monitoring the condition of a structure such as wing box in an aircraft to detect and image the defects. The importance of health monitoring of aerospace structures has gained considerable interest over the past two decades. There are different array based techniques[1–10] such as single transmitter and multiple receiver (STMR) array and multiple transmitter and multiple receiver (MTMR) array for the health monitoring of isotropic and anisotropic structures. The applicability of an array based system for rapid inspection of isotropic[1–4] and anisotropic[5–8] platelike structures using Lamb waves[11,12] was demonstrated using a beam steering algorithm. MTMR array along with tomographic[9,10] reconstruction technique was used for SHM of anisotropic platelike structures. Baseline subtraction methods[13–15] were also used for health monitoring of platelike and pipe structures.

All the above algorithms require prior knowledge of elastic moduli in order to compute the phase velocities particularly in anisotropic materials such as fiber reinforced composites. But, in all the above works, the elastic moduli used for the calculation correspond to the virgin sample. It is well known that elastic moduli of the sample may vary over time due to the introduction of defects or due to heat damage. Hence, there is a need to measure elastic moduli as well as image defects as part of SHM. In addition to the elastic moduli, availability of information regarding the material symmetry and orientation of principal planes will help in arranging sensor arrays on the structure. The availability of the above two will reduce the number of calculations in algorithms such as phase reconstruction algorithms used in SHM.

Recently it was shown that a genetic algorithm (GA) based blind inversion approach[16] could be used for the determination of elastic moduli, material symmetry, and orientation of principal planes of orthotropic material using bulk wave velocity data. However, only seven out of the nine elastic moduli could be reconstructed as the remaining two moduli were found to be insensitive to the bulk wave velocity data. Furthermore, as the bulk wave velocity based reconstruction of elastic moduli involves goniometry, it is unsuitable for *in situ* material characterization and SHM.

Lamb wave based methods to reconstruct elastic moduli have also been explored for isotropic,[17,18] transversely isotropic,[19] and orthotropic[20–22] materials. Inversion of elastic moduli of an orthotropic composite plate was carried out using experimental data on the zeros of the reflection coefficient.[20,21] Reconstruction of elastic moduli was carried out for transversely isotropic[19] and orthotropic[22] materials by reducing the error between theoretical and experimental Lamb wave dispersion curves. In the case of orthotropic material, four elastic moduli (C_{11}, C_{13}, C_{33}, and C_{55}) were reconstructed using dispersion data along the fibers, three elastic moduli (C_{22}, C_{23}, and C_{44}) were reconstructed using dispersion data across the fibers, and the remaining two elastic moduli were reconstructed using the dispersion data along a nonsymmetry direction (45° direction). However, the error was found to be high in the reconstruction of the two elastic moduli (C_{12} and C_{66}) using dispersion data in the nonsymmetry direction. It was also found that the measurement of dispersion in the nonsymmetry direction was very difficult.[22]

Lamb wave velocity based sensitivity studies on different orthotropic materials were carried out recently.[23] It was

[a)]Author to whom correspondence should be addressed. Electronic mail: balas@iitm.ac.in

shown that two elastic moduli (C_{44} and C_{55}) are more sensitive to the A_0 mode and the rest of the seven elastic moduli are more sensitive to the S_0 mode. Furthermore, it was observed that different elastic moduli were sensitive to the S_0 mode and A_0 mode velocity profiles in different angular sectors. Based on these observations, a method to reconstruct elastic moduli of orthotropic material based on the slowness curves of S_0 and A_0 modes had been proposed and validated.[23] As the slowness based method covers the sensitive regions of all elastic moduli, the reconstruction resulted in good estimates for the elastic moduli.

In the present work, a Lamb wave based blind inversion method is proposed for the reconstruction of elastic moduli, material symmetry, and orientation of principal planes of monoclinic and higher symmetry systems. The proposed method combines the advantages of blind inversion method[16] and Lamb wave slowness based reconstruction method.[23] Advantages in reconstructing the elastic moduli from the Lamb wave phase velocity data include (a) contact mode (sensor array patches placed on the structure) or noncontact mode (e.g., laser and air coupled ultrasonics) operation, (b) reconstruction with higher accuracy as all nonzero components of elastic moduli matrix are sensitive to ultrasonic Lamb wave phase velocities,[22] and (c) self-calibration for SHM work. Being *in situ*, this technique finds application in the area of material characterization and SHM of anisotropic platelike structures.

This paper is organized as follows. Section II outlines the approach to determine statistically significant elastic moduli, to provide Lamb wave sensitivity based selection of effective elastic moduli, and to determine principal plane orientations. Section III presents examples of the blind inversion approach using simulated S_0 and A_0 Lamb wave mode velocities of materials with different material symmetries. Validation of the blind inversion method using measured velocities of S_0 and A_0 modes on a 3.15 mm quasi-isotropic graphite-epoxy composite plate cut at 45° to the material symmetry direction is described in Sec. IV. Summary and conclusions are presented in Sec. V.

II. PROPOSED METHOD

A. Blind inversion

In the context of elastic moduli determination from measured ultrasonic velocity data, blind inversion is, in general, characterized by (i) reconstructed elastic moduli having ill-defined error margins, (ii) nonzero values associated with certain elastic moduli that ought to be zero due to material symmetry, and (iii) nonuniqueness of the resulting data set. In other words, different initial conditions in the multidimensional search space would lead to different solution sets with different error margins. The approach proposed in this work seeks to use Lamb wave mode velocity data for blind inversion by extracting "statistically significant" elastic moduli and identifying effective elastic moduli from their sensitivities to velocity data. Throughout this work, the term velocity refers to the phase velocity and not the group velocity.

The blind inversion approach employed in the present work uses the GA (Ref. 24) and proceeds by minimizing the error function, err(C), where

$$\text{minimize } \text{err}(C) = \sum_{i=0.1}^{i=90.1} [V_i^c(\phi) - V_i^m(\phi)]^2$$

$$\text{subject to } C_{\min} < C < C_{\max}, \qquad (1)$$

where V^c is the calculated velocity, V^m is the measured velocity, and i is the index specifying the angle ϕ (in degrees). As shown in Eq. (1), err(C) is defined as the sum of squares of the difference between measured and calculated velocities. The calculated velocities refer to the velocities of S_0 and A_0 modes calculated over several angles in a quadrant by solving the Rayleigh–Lamb equation[11,12] employing each of the elastic moduli sets generated by the GA. Using each of the 12 different sets of GA parameters listed in Tables XII and XIII, ten trials are carried out to generate ten reconstructed elastic moduli data sets. In each trial, 200 generations are allowed to evolve a possible solution set. The search space is defined by the bounds on the values of the elastic moduli. While a priori information helps in defining these bounds, the creep option[24,25] used during inversion allows for random excursions from the search space to ensure that the sampling is extensive and unbiased. Simulations were carried out with "measured" wave velocities using available elastic moduli pertaining to various material symmetry classes ranging from the monoclinic class to the isotropic class. The "calculated" velocities were obtained from trial elastic moduli generated by the GA. The fluctuations in the reconstructed elastic moduli exhibited a distinct trend. Most or all the elastic moduli of a symmetry class exhibited small fluctuations while a few of the elastic moduli pertaining to that symmetry class showed much larger fluctuations through all the trials. Among those that exhibited large fluctuations, those elastic moduli that were relevant to the given symmetry class had very small values and were found to have very little impact on the ultrasonic velocities.

This trend led to the use of a simple statistical measure of the fluctuations, the coefficient of variation[26] in a particular elastic constant C_{ij}, defined as

$$C_v[C_{ij}] = \frac{\sigma[C_{ij}]}{\mu[C_{ij}]} \times 100, \qquad (2)$$

where $\mu[C_{ij}]$ is the mean and $\sigma[C_{ij}]$ is the standard deviation of that elastic constant C_{ij}. The brackets refer to the 12 different sets of C_{ij} obtained using the 12 sets of GA parameters given in Tables XII and XIII. From the simulations carried out on all the material symmetries ranging from the monoclinic case to the isotropic case, the C_v of the corresponding elastic moduli was empirically found not to exceed 20%. Accordingly, an elastic modulus was considered to be statistically significant if its C_v is less than 20% and was treated as "noise" otherwise. For example, if a material is orthotropic (defined by 9 unknown elastic constants) but treated as monoclinic (defined by 13 unknown elastic constants), the GA based blind inversion described above would lead to 4 of the 13 elastic constants exhibiting large fluctuations (termed as noise in the present context), while the fluctuations of the

TABLE I. Maximum percent variation in S_0 and A_0 mode velocities when elastic moduli of a monoclinic crystal (M) and a quasi-isotropic composite (Q) were varied by 10% and 20%.

	Maximum % variation in S_0 mode velocity when $\boldsymbol{C}_{ij}$ is varied by				Maximum % variation in A_0 mode velocity when $\boldsymbol{C}_{ij}$ is varied by			
	Monoclinic crystal		Rotated quasi-isotropic composite		Monoclinic crystal		Rotated quasi-isotropic composite	
$\{C_{ij}\}$	10%	20%	10%	20%	10%	20%	10%	20%
C_{11}	**5.76**	**11.23**	**5.21**	**10.18**	2.09	3.96	0.39	0.71
C_{12}	**0.74**	**1.48**	**0.76**	**1.51**	0.34	0.55	0.07	0.12
C_{13}	**2.03**	**4.29**	**0.92**	**1.95**	0.64	1.37	0.03	0.06
C_{16}	**0.25**	**0.49**	**0.47**	**0.93**	0.10	0.19	0.05	0.08
C_{22}	**4.96**	**9.73**	**5.21**	**10.20**	1.95	3.68	0.39	0.71
C_{23}	**0.98**	**2.07**	0.93	**1.96**	0.32	0.67	0.03	0.06
C_{26}	**1.01**	**2.01**	**0.47**	**0.93**	0.54	0.87	0.05	0.09
C_{33}	**0.93**	**1.69**	**0.70**	**1.20**	0.32	0.45	0.04	0.07
C_{36}	**0.12**	**0.24**	**0.01**	**0.02**	0.19	0.16	0.00	0.00
C_{44}	0.00	0.00	0.00	0.00	**1.39**	**2.54**	**4.09**	**7.89**
C_{45}	0.00	0.00	0.00	0.00	**0.28**	**0.36**	**0.12**	**0.22**
C_{55}	0.00	0.00	0.00	0.00	**1.33**	**2.44**	**4.09**	**7.87**
C_{66}	**1.48**	**2.94**	**1.41**	**2.80**	0.55	1.06	0.12	0.21

remaining 9 elastic constants would remain within a narrow range (termed as statistically significant in the present context). In this sense, the GA, due to the quasirandom nature of the search through the multidimensional solution space, appears to distinguish certain elastic moduli as statistically significant and certain elastic moduli as noise for a given symmetry class.

B. Lamb wave mode sensitivity study in monoclinic systems

Studies have been carried out to determine Lamb wave sensitivities in the monoclinic systems following earlier such studies on orthotropic systems.[23] Two monoclinic systems, a monoclinic crystal[27] and a rotated quasi-isotropic composite plate,[23] were chosen here for presenting the results. Varying each elastic constant of these two monoclinic materials by 10% and 20%, respectively, the sensitivities of the 13 monoclinic elastic moduli to the S_0 and A_0 mode velocities have been calculated at a fixed frequency-thickness product (fd = 1), and the results are given in Table I. For example, from Table I it can be seen that, when C_{44} of the monoclinic crystal was varied by 10%, the variation in velocity is negligibly small for the S_0 mode but was 1.39% for the A_0 mode. Similarly, when C_{55} was varied by 10%, the variation in velocity is negligibly small for the S_0 mode but was 1.33% for the A_0 mode. In all the cases, it is observed that C_{44}, C_{45}, and C_{55} are more sensitive to A_0 mode velocities and the rest of ten elastic moduli are more sensitive to S_0 mode velocities. These extensive studies have led to the following selection "rule:" in monoclinic symmetric materials, C_{44}, C_{45}, and C_{55} are to be chosen from the elastic moduli set reconstructed using A_0 mode velocities and the rest of the ten elastic moduli are to be chosen from the elastic moduli set reconstructed using S_0 mode velocities to define the "effective elastic moduli."

C. Determination of principal plane orientations

Once the effective elastic moduli have been determined, the method proposed by Cowin and co-workers[28,29] is used to find the orientation of principal planes of an anisotropic material. Using the elastic stiffness matrix (C_{ij}), two second order tensors $\boldsymbol{A}_{ij}$ (the Voigt tensor) and $\boldsymbol{B}_{ij}$ (the dilatational modulus) are defined.[28–30] Eigenvectors corresponding to each tensor $\boldsymbol{A}$ and $\boldsymbol{B}$ are calculated. Table II provides the classification of the material symmetry based on the number of eigenvectors of $\boldsymbol{A}$ that match with that of $\boldsymbol{B}$. Orientation of the principal plane, R_p, with respect to the laboratory system can be identified through Euler angles made by the normal to this principal plane which is an eigenvector of tensors $\boldsymbol{A}$ and $\boldsymbol{B}$.

III. CASE STUDIES USING SIMULATED VELOCITY DATA

Extensive simulations have been carried out to reconstruct elastic moduli, identify material symmetry, and determine principal plane orientations using simulated velocities of materials having monoclinic, orthotropic, and transversely isotropic material symmetries. Two case studies are pre-

TABLE II. Various symmetries obtained by comparing eigenvectors of tensors A and B.

If three eigenvectors of $\boldsymbol{A}$ and $\boldsymbol{B}$ do not match	The material has no symmetric planes, i.e., triclinic symmetry
If one eigenvector of $\boldsymbol{A}$ is equal to one eigenvector of $\boldsymbol{B}$	Monoclinic symmetry
If all three eigenvectors of $\boldsymbol{A}$ and $\boldsymbol{B}$ are equal	Orthotropic
If all three eigenvectors of $\boldsymbol{A}$ and $\boldsymbol{B}$ are equal and two eigenvalues are also equal	Transversely isotropic

TABLE III. Theoretical and reconstructed moduli of 3.15 mm graphite-epoxy composite.

	Elastic moduli (GPa) of 3.15 mm quasi-isotropic graphite-epoxy composite					
		Reconstructed moduli				
		From S_0 mode	From A_0 mode	Effective moduli $\{C_{ij}\}_{\text{eff}}$		
$\{C_{ij}\}$	Theoretical moduli	$\{C_{ij}\}_{S0}$	$\{C_{ij}\}_{A0}$	$C_{ij\ \text{eff}}$	(%error)	$(\sigma)_{\text{eff}}$
C_{11}	64.48	**63.13**	61.55	63.13	2.09	1.23
C_{12}	19.51	**18.81**	18.69	18.81	3.59	0.95
C_{13}	5.84	**5.68**	5.70	5.68	2.74	0.51
C_{16}	2.00	**1.99**	2.00	1.99	0.50	0.32
C_{22}	55.77	**54.24**	56.29	54.24	2.74	1.14
C_{23}	5.78	**5.90**	5.83	5.90	2.08	0.34
C_{26}	5.54	**5.58**	5.70	5.58	0.72	0.37
C_{33}	12.43	**12.24**	12.55	12.24	1.53	0.93
C_{36}	0.05	**0.54**	0.48	0.54	980.00	0.16
C_{44}	4.20	4.15	**4.13**	4.13	1.67	0.04
C_{45}	0.09	0.49	**0.19**	0.19	111.11	0.06
C_{55}	4.31	4.27	**4.26**	4.26	1.16	0.04
C_{66}	18.27	**17.72**	17.67	17.72	3.01	0.62

sented here dealing with orthotropic and transversely isotropic symmetric materials. In the two case studies, simulations are carried out using the elastic moduli data obtained from literature.[31] Using the known elastic moduli data, phase velocities of S_0 and A_0 modes at various angles in the X_1-X_3 plane are calculated using the Rayleigh–Lamb equations. Simulations have been carried out for a 3.15 mm quasi-isotropic graphite-epoxy composite rotated by 30° about X_3-axis and a 2.16 mm unidirectional graphite-epoxy composite.

In all the simulations considered, reconstruction starts with the assumption that the specimen belongs to the monoclinic symmetry class for which the symmetric and asymmetric mode equations are decoupled.

The "blind inversion" procedure is implemented as follows. Ultrasonic phase velocity data of S_0 and A_0 modes evaluated using the reference elastic moduli at 1° intervals in a quadrant are used as input or measured data to the inversion method. The calculated velocity data are obtained from elastic moduli generated by GA.

The C_v is evaluated for several sets of GA parameters with each set consisting of many repeated runs using different parameters of the quasirandom GA search process (mutation chance, creep chance, and creep amount) and is compared with the threshold value. If the C_v of all 13 elastic moduli is less than the threshold limit, then the material is considered to have a monoclinic symmetry in the laboratory coordinate system; otherwise the reconstruction will be carried out assuming that the material is having next higher symmetry, i.e., orthotropic symmetry. For the orthotropic assumption, C_v is calculated and compared with the threshold value. If the C_v is less than the threshold limit, then the material is considered to have an orthotropic symmetry; otherwise the reconstruction will be carried out assuming that the material is having next higher symmetry, i.e., transversely isotropic symmetry. The same procedure can be followed for the next higher symmetries such as cubic and isotropic. From the statistically significant elastic moduli, the unknown material symmetry and the principal planes (angles between the geometrical coordinates and the material symmetry coordinates) are then evaluated using the method proposed by Cowin and co-workers.[28–30]

A. Case study 1: Determination of orthotropic symmetry (3.15 mm graphite-epoxy composite material rotated by 30° about X_3-axis)

The first case study discusses about a quasi-isotropic composite plate rotated 30° about X_3-axis. The elastic moduli used for the simulation are listed in Table III under the column "Theoretical" and are calculated using rule of mixtures[32] with the elastic moduli of unidirectional graphite-epoxy composite given in literature.[31] Using the theoretical moduli, the simulated phase velocities of S_0 and A_0 modes for this configuration are calculated. As the axes of symmetry are assumed unknown, the material is initially considered to have the monoclinic symmetry, requiring 13 elastic moduli as input. Reconstruction has been carried out using the proposed GA based blind inversion approach.

Figure 1 presents the C_v evaluated from the mean and standard deviation of each elastic modulus reconstructed using the monoclinic assumption. From the trend in Fig. 1, it is evident that the C_v of 11 out of 13 elastic moduli are statistically significant. Two off-diagonal moduli, C_{36} and C_{45}, which are relevant to the monoclinic symmetry, are found to be numerically very small and not to affect the propagation velocities in any significant manner. Hence, they resulted in poor estimatesm, and the corresponding C_v values are above the threshold limit indicating that they are statistically insignificant. However, since these two elastic moduli have very low impact on the velocity data, these are combined with the 11 statistically significant elastic moduli to form effective elastic moduli representing monoclinic symmetry in the laboratory coordinate system.

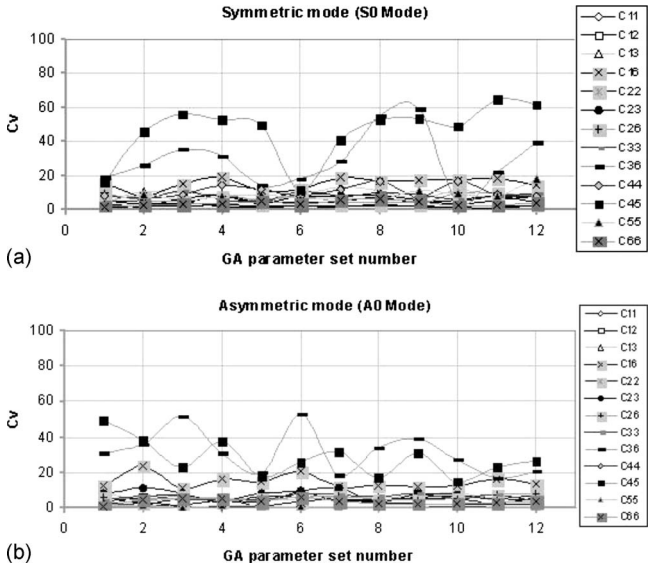


FIG. 1. C_v of 13 elastic moduli of quasi-isotropic composite reconstructed using (a) symmetric (S_0) mode and (b) asymmetric (A_0) mode using simulated velocity data.

The reconstructed elastic moduli using individual S_0 and A_0 mode velocities have been listed in Table III. Boldfaced regions in Table III represent the elastic moduli sensitive to Lamb wave phase velocities of S_0 and A_0 modes, respectively. The effective elastic moduli ($\{C_{ij}\}_{\text{eff}}$) are also given in Table III. The maximum error in the effective elastic moduli other than C_{36} and C_{45} with respect to the theoretical elastic moduli is found to be 3.59%. C_{36} and C_{45} are low in magnitude, insensitive to velocity data, and gave rise to large errors.

From Table III, it can also be observed that the elastic moduli reconstructed from a single mode either S_0 or A_0 is in good agreement with the theoretical moduli. This can have useful practical consequences in situations where the signal to noise ratio of any one of these modes is inadequate to make measurements.

Table IV shows the Voigt tensor ($\boldsymbol{A}$) and dilatational tensor ($\boldsymbol{B}$) constructed from the effective elastic moduli ($\{C_{ij}\}_{\text{eff}}$) listed in Table III. The eigenvectors of these tensors are calculated and are given in Table V. Neglecting the small variations (in the second decimal) in the eigenvectors, it is found that three eigenvectors of $\boldsymbol{A}$ match with three eigenvectors of $\boldsymbol{B}$, implying that there are three orthogonal principal planes and that the material symmetry is orthotropic. Two of the three symmetry axes were found to be coinciding with the laboratory axes. The third symmetry axis was reconstructed to be 30.42° (against 30°) with respect to the X_3-axis. The error in the angle reconstruction was found to be less than 1.5%.

TABLE IV. The two second order tensors: (a) Voigt tensor ($\boldsymbol{A}_{ij}$) and (b) dilatational tensor ($\boldsymbol{B}_{ij}$).

(a)	87.62	8.11	0
	8.11	78.95	0
	0	0	23.82
(b)	85.11	7.76	0
	7.76	76.09	0
	0	0	20.63

TABLE V. Eigenvectors of (a) Voigt tensor ($\boldsymbol{A}_{ij}$) and (b) dilatational tensor ($\boldsymbol{B}_{ij}$).

(a)	0	0.5141	−0.8577
	0	−0.8577	−0.5141
	1	0	0
(b)	0	0.4988	−0.8667
	0	−0.8667	−0.4988
	1	0	0

To account for the experimental error in velocity assessment, 2% random noise was added to the simulated velocities and the reconstruction was carried out. The 2% noise is based on error estimates[18,23] expected in such experiments. The reconstructed elastic moduli are given in Table VI along with the error and standard deviation in the reconstructed moduli. The maximum error in the effective elastic moduli is found to be less than 6.44% while the maximum error in the effective elastic moduli reconstructed using noise-free velocity data was found to be 3.59%. Euler angles are calculated for the reconstructed elastic moduli. The angle reconstructed is 30.87° with an error of 2.9% while the reconstructed angle for the noise-free velocity data was 30.42° having an error of 1.5%.

B. Case study 2: Determination of transversely isotropic symmetry (2.16 mm graphite-epoxy composite)

This case study discusses a transversely isotropic symmetry (2.16 mm graphite-epoxy) composite plate with the elastic moduli listed in Table VII under the column Theoretical. Simulated phase velocities of S_0 and A_0 modes are calculated using the "theoretical" moduli. As the symmetry axes are assumed unknown, the material is initially considered as monoclinic material requiring 13 elastic moduli as input. Simulated velocity data were used in reconstructing the 13 elastic moduli using the GA based blind inversion approach.

Figure 2 shows the C_v evaluated from the mean and standard deviation of the reconstructed monoclinic elastic moduli. From Fig. 2, it can be observed that 10 out of 13 elastic moduli are statistically significant while 3 are found to be statistically insignificant. Unlike in the first case study, the material in the present case is not regarded as monoclinic as the three statistically insignificant moduli were found to be sensitive to velocity data. Going to the next higher symmetry, namely, orthotropic, it is noted that the three statistically insignificant moduli are irrelevant to the orthotropic symmetry class and can be discarded. As an orthotropic material requires only nine elastic moduli and all these are found to be statistically significant, the material in this case study is taken to be orthotropic.

Figure 3 shows the C_v of all the nine elastic moduli. From Fig. 3, it can be observed that the C_v of all nine orthotropic elastic moduli are within the threshold limit. Hence, the material is deduced to be orthotropic. Based on sensitiv-

TABLE VI. Theoretical and reconstructed moduli of 3.15 mm graphite-epoxy composite when 2% random noise was added to the theoretical velocities.

Elastic moduli (GPa) of 3.15 mm quasi-isotropic graphite-epoxy composite						
		Reconstructed moduli				
		From S_0 mode	From A_0 mode	Effective moduli $\{C_{ij}\}_{\text{eff}}$		
$\{C_{ij}\}$	Theoretical moduli	$\{C_{ij}\}_{S0}$	$\{C_{ij}\}_{A0}$	$C_{ij\ \text{eff}}$	(%error)	$(\sigma)_{\text{eff}}$
C_{11}	64.48	**62.78**	61.87	62.78	2.64	1.71
C_{12}	19.51	**18.94**	18.63	18.94	2.92	1.06
C_{13}	5.84	**5.77**	5.70	5.77	1.19	0.48
C_{16}	2.00	**1.99**	1.93	1.99	0.50	0.43
C_{22}	55.77	**54.07**	55.79	54.07	3.05	1.64
C_{23}	5.78	**5.98**	5.70	5.98	3.46	0.43
C_{26}	5.54	**5.65**	5.68	5.65	1.98	0.53
C_{33}	12.43	**11.63**	12.54	11.63	6.44	1.09
C_{36}	0.05	**0.60**	0.52	0.60	1100.00	0.19
C_{44}	4.20	4.19	**4.13**	4.13	1.66	0.05
C_{45}	0.09	0.36	**0.24**	0.24	166.66	0.07
C_{55}	4.31	4.34	**4.24**	4.24	1.62	0.05
C_{66}	18.27	**17.67**	17.83	17.67	3.28	0.79

ity analysis C_{44} and C_{55} are selected from the elastic moduli set reconstructed using A_0 mode velocities and the rest of the seven moduli are selected from the elastic moduli set reconstructed using S_0 mode velocities. Table VII contains elastic moduli reconstructed using S_0 and A_0 mode velocity data separately and the effective elastic moduli ($\{C_{ij}\}_{\text{eff}}$). The maximum error in the effective elastic moduli from the theoretical elastic moduli is found to be 1.29%.

Table VIII shows the Voigt tensor ($\boldsymbol{A}$) and dilatational tensor ($\boldsymbol{B}$) constructed from the $\{C_{ij}\}_{\text{eff}}$ provided in Table VII. As three eigenvectors of each of these two tensors ($\boldsymbol{A}$ and $\boldsymbol{B}$) match, the material is considered to have three planes of mirror symmetry, and the three planes are found to be mutually orthogonal to each other. Furthermore, because of the degeneracy of two eigenvalues of each of the tensors $\boldsymbol{A}$ and $\boldsymbol{B}$, the material symmetry is determined to be transversely isotropic and the material symmetry planes are found to match exactly with the laboratory coordinate system.

IV. EXPERIMENTAL RESULTS

To validate the present work, two different experiments were carried out on a plate cut from a 3.15 mm graphite-epoxy composite at 45° to the symmetry planes. A modified double ring configuration of the STMR compact SHM array was developed to measure Lamb wave velocities; the schematic of which is presented in Fig. 4. The 3.15 mm graphite-epoxy composite has been made from 21 layers of unidirectional graphite-epoxy layers in the sequence of $(+45,-45,0,90,0,-45,+45)_{3s}$. In the first experiment, a 500 kHz center frequency (V101 Panametrics Inc., Waltham, MA) Lead Zirconate Titanate (PZT) transducer based modified double ring STMR array was used for transmission and reception and in the second experiment, transmission was same as above and POLYTEC OFV-5000 laser vibrometer with DD-300 displacement decoder was used for reception. Experiments were carried out using MATEC pulser-receiver,

TABLE VII. Theoretical and reconstructed moduli of transversely isotropic graphite-epoxy composite.

Elastic moduli (GPa) of transversely isotropic graphite-epoxy composite						
		Reconstructed moduli				
		From S_0 mode	From A_0 mode	Effective moduli $\{C_{ij}\}_{\text{eff}}$		
$\{C_{ij}\}$	Theoretical moduli	$\{C_{ij}\}_{S0}$	$\{C_{ij}\}_{A0}$	$C_{ij\ \text{eff}}$	(%error)	$(\sigma)_{\text{eff}}$
C_{11}	134.36	**134.23**	130.89	134.23	0.10	0.56
C_{12}	6.24	**6.24**	6.21	6.24	0.00	0.34
C_{13}	6.24	**6.27**	6.33	6.27	0.48	0.30
C_{22}	12.43	**12.49**	12.38	12.49	0.48	0.40
C_{23}	5.39	**5.46**	5.55	5.46	1.29	0.22
C_{33}	12.43	**12.56**	12.42	12.56	1.04	0.38
C_{44}	3.52	3.53	**3.54**	3.54	0.57	0.05
C_{55}	5.00	5.05	**5.01**	5.01	0.20	0.01
C_{66}	5.00	**5.04**	5.07	5.04	0.80	0.09

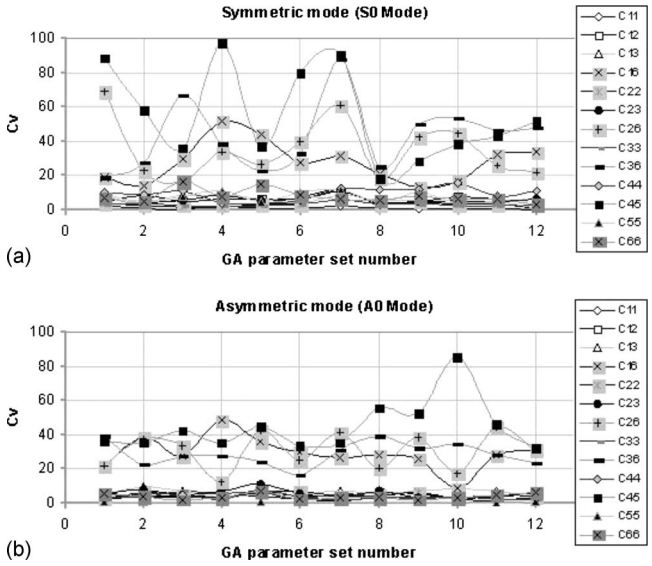


FIG. 2. C_v of 13 elastic moduli of quasi-isotropic composite reconstructed using (a) symmetric (S_0) mode and (b) asymmetric (A_0) mode using simulated velocity data.

and the signals were digitized by using Agilent DSO6032A oscilloscope. A two-cycle tone burst signal at an excitation frequency of 200 kHz was used for excitation and 256 signals were acquired and averaged using an oscilloscope to reduce noise. Experimental signals have been taken at an interval of 10° for PZT based modified double ring STMR array system and at an interval of 5° for laser based reception system. In both the experiments, to ensure that only S_0 and A_0 modes get generated, a 500 kHz probe was excited at 200 kHz such that the low fd(fd=0.63 MHz mm) product was maintained.

A. Velocity data from a 3.15 mm quasi-isotropic graphite-epoxy composite using PZT based STMR array for transmission and reception

A 500 kHz PZT based double ring STMR array was used to generate and receive the signals at two different radii (60 and 90 mm). From the experimental signals, extraction of the A_0 mode was difficult as the amplitude of the signal

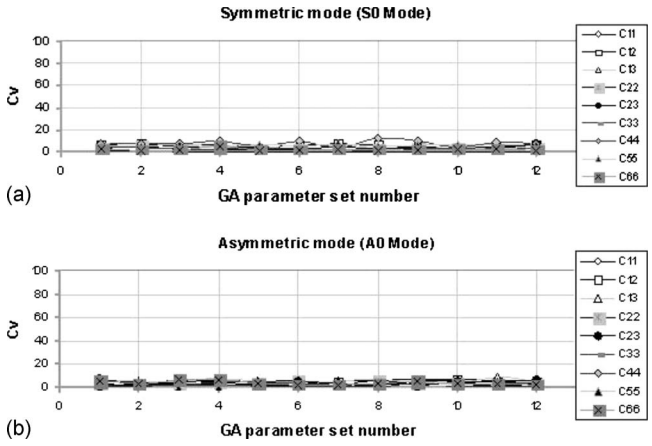


FIG. 3. C_v of nine elastic moduli of orthotropic composite reconstructed using (a) symmetric (S_0) mode and (b) asymmetric (A_0) mode using simulated velocity data.

TABLE VIII. The two second order tensors: (a) Voigt tensor ($\boldsymbol{A}_{ij}$) and (b) dilatational tensor ($\boldsymbol{B}_{ij}$).

(a)	146.74	0	0
	0	24.19	0
	0	0	24.29
(b)	144.28	0	0
	0	21.07	0
	0	0	21.11

was very small. Hence only the time of flights of propagating S_0 mode were measured. These measured time of flights were then used to calculate phase velocities of S_0 mode in different directions. The difference in phase velocity and group velocity at this particular fd was much less and hence the velocities obtained using experiments are considered as phase velocities.

As the axes of symmetry are assumed unknown, the material is considered as monoclinic material which requires 13 elastic moduli as input to the proposed blind GA approach. Experimentally measured S_0 mode velocity data in the X_1-X_3 plane were used in reconstructing the 13 elastic moduli. From Fig. 5, it can be observed that 11 out of 13 elastic moduli are statistically significant while 2 elastic moduli (C_{36} and C_{45}) are found to be statistically insignificant. These two off-diagonal moduli, however do not affect the velocities of the S_0 mode in different propagation directions. Hence, the material is considered to have a monoclinic symmetry in the laboratory coordinate system.

The elastic moduli reconstructed by inverting the experimentally measured S_0 mode velocities are tabulated in Table IX. Theoretical elastic moduli listed in Table IX are calculated using rule of mixtures[32] with the elastic moduli of unidirectional graphite-epoxy composite given in literature.[31] As observed in simulations, it is found that C_{36} and C_{45} are low in magnitude and produce large errors. The maximum error in the effective elastic moduli (other than C_{36} and C_{45}) from the theoretical elastic moduli is found to be 6.78% and the maximum standard deviation is 6.54% from the mean of corresponding elastic moduli.

Table X shows the Voigt tensor (**A**) and dilatational tensor (**B**) constructed from the $\{C_{ij}\}_{S0}$ given in Table IX. As the three eigenvectors of these two tensors (***A*** and ***B***) match, the material is determined to have an orthotropic symmetry. The angle reconstructed is 45.46° while the actual angle is 45°.

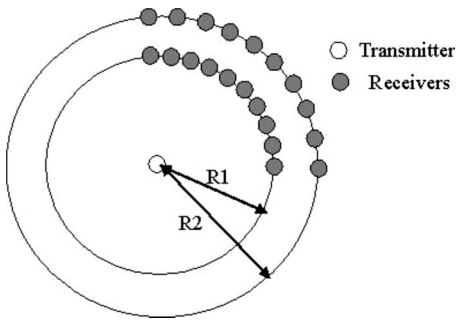


FIG. 4. Schematic of modified double ring STMR array: R1—inner radius and R2—outer radius of the ring.

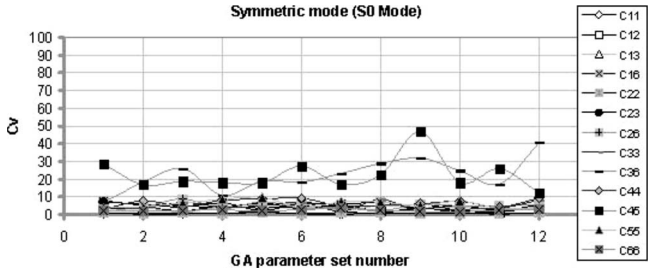


FIG. 5. C_v of 13 elastic moduli of quasi-isotropic composite reconstructed using measured velocity data of fundamental symmetric (S_0) mode.

Hence, the error in angle calculation is only 1%. It is to be noted that, for the same material, performance of the inversion algorithm obtained for the experimental data is in close agreement with that obtained using simulated velocity data in Sec. III A. Percentage error in the angle estimation is found to be 1% for the experimental data whereas it is found to be 1.5% for the simulated data.

B. Velocity data from a 3.15 mm quasi-isotropic graphite-epoxy composite using commercial PZT probe for transmission and laser Doppler velocimetry for reception

Experiments are also carried out with conventional PZT transducer for generation and laser Doppler velocimetry with out-of-plane displacement decoder for reception. From the experimental signals collected at two different radii (60 and 90 mm) it was possible to identify the S_0 mode and A_0 mode signals separately and deduce their velocities in different propagation directions.

As the symmetry axes are assumed as unknown, the material is initially considered as monoclinic material, requiring 13 elastic moduli as input. Experimental S_0 and A_0 mode velocity data were used in reconstructing the 13 elastic moduli.

TABLE IX. Theoretical and reconstructed moduli of 3.15 mm quasi-isotropic graphite-epoxy composite.

	Elastic moduli (GPa) of 3.15 mm quasi-isotropic graphite-epoxy composite			
		Reconstructed from S_0 mode		
$\{C_{ij}\}$	Theoretical moduli	$\{C_{ij}\}_{S0}$	(%error)	(σ)
C_{11}	61.15	58.05	5.07	0.41
C_{12}	18.49	18.00	2.65	0.62
C_{13}	5.81	6.09	4.82	0.23
C_{16}	4.36	4.39	0.69	0.23
C_{22}	61.15	58.32	4.63	0.53
C_{23}	5.81	6.11	5.16	0.28
C_{26}	4.36	4.30	1.37	0.22
C_{33}	12.43	11.99	3.54	0.59
C_{36}	0.06	0.57	850.00	0.13
C_{44}	4.26	4.28	0.47	0.28
C_{45}	0.11	0.51	363.63	0.11
C_{55}	4.26	4.25	0.23	0.26
C_{66}	17.25	16.08	6.78	0.40

TABLE X. The two second order tensors: (a) Voigt tensor ($\boldsymbol{A}_{ij}$) and (b) dilatational tensor ($\boldsymbol{B}_{ij}$).

(a)	82.14	9.26	0
	9.26	82.43	0
	0	0	24.19
(b)	78.38	9.20	0
	9.20	78.68	0
	0	0	20.52

The C_v, evaluated from the mean and standard deviation of the elastic moduli reconstructed using S_0 and A_0 modes, is shown in Fig. 6. Similar to Sec. IV A, it can be observed from Fig. 6 that 11 out of 13 elastic moduli are statistically significant while 2 elastic moduli (C_{36} and C_{45}) are found to be statistically insignificant. Hence, the material is considered to have a monoclinic symmetry. Out of 13 effective elastic moduli, 3 moduli (C_{44}, C_{45}, and C_{55}) are taken from the elastic moduli set reconstructed using A_0 mode velocities and the remaining 10 moduli are taken from the elastic moduli set reconstructed using S_0 mode velocities. The effective elastic moduli ($\{C_{ij}\}_{\text{eff}}$), reconstructed by inverting the experimentally determined S_0 and A_0 mode velocities, are tabulated in Table XI.

The maximum error in the effective elastic moduli (other than C_{36} and C_{45}) is found to be 6.62% and the maximum standard deviation is 8% from the mean of corresponding elastic modulus. It may be noted that the percent error given in Tables IX and XI does not imply actual error as the theoretical moduli are calculated using the rule of mixtures.[32]

As the three eigenvectors of the Voigt tensor ($\mathbf{A}$) and dilatational tensor ($\mathbf{B}$) match, the material is considered to have an orthotropic symmetry. The angle reconstructed is 46.69° while the actual angle is 45°. Hence, the error in angle calculation is found to be only 3.7%.

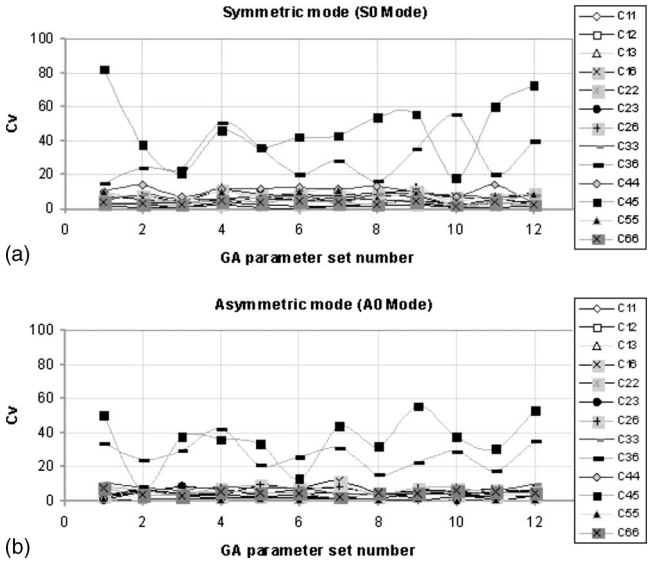


FIG. 6. C_v of 13 elastic moduli of quasi-isotropic composite reconstructed using measured velocity data of (a) symmetric (S_0) mode and (b) asymmetric (A_0) mode.

TABLE XI. Theoretical and reconstructed moduli of 3.15 mm quasi-isotropic graphite-epoxy composite.

	Elastic moduli (GPa) of 3.15 mm quasi-isotropic graphite-epoxy composite					
		Reconstructed moduli				
		From S_0 mode	From A_0 mode	Effective moduli $\{C_{ij}\}_{\text{eff}}$		
$\{C_{ij}\}$	Theoretical moduli	$\{C_{ij}\}_{S0}$	$\{C_{ij}\}_{A0}$	$C_{ij\ \text{eff}}$	(%error)	$(\sigma)_{\text{eff}}$
C_{11}	61.15	**57.10**	57.94	57.10	6.62	1.44
C_{12}	18.49	**17.89**	18.24	17.89	3.25	0.80
C_{13}	5.81	**5.93**	6.04	5.93	2.06	0.38
C_{16}	4.36	**4.36**	4.37	4.36	0.00	0.35
C_{22}	61.15	**58.06**	61.34	58.06	5.05	0.72
C_{23}	5.81	**6.19**	6.03	6.19	6.54	0.33
C_{26}	4.36	**4.26**	4.48	4.26	2.29	0.26
C_{33}	12.43	**11.81**	12.46	11.81	4.99	0.84
C_{36}	0.06	**0.61**	0.51	0.61	916.66	0.18
C_{44}	4.26	4.09	**4.31**	4.31	1.17	0.05
C_{45}	0.11	0.42	**0.19**	0.19	72.72	0.07
C_{55}	4.26	4.48	**4.35**	4.35	2.11	0.04
C_{66}	17.25	**16.58**	17.32	16.58	3.88	0.57

V. SUMMARY AND CONCLUSIONS

In the present work, a GA based blind inversion method has been proposed for the complete reconstruction of elastic moduli, material symmetry, and orientation of principal planes of monoclinic and higher symmetry plate structures using ultrasonic Lamb wave velocities of S_0 and A_0 modes.

The blind inversion approach is based on the use of a simple statistical measure of the variation in the reconstructed elastic moduli. A statistical parameter, coefficient of variation (C_v), was used in identifying the statistically significant elastic moduli. From the simulation studies on different symmetries, the threshold limit was set 20%. A sensitivity analysis has been carried out for several known monoclinic materials. Similar trends were observed for different monoclinic materials. Sensitivity studies for two monoclinic materials have been listed in Table I. In all cases, it was observed that C_{44}, C_{45}, and C_{55} are more sensitive to A_0 mode velocity compared to the S_0 mode velocity, and the rest of the ten moduli are more sensitive to S_0 mode velocities. Further studies have to be carried out to understand this important and interesting feature. Hence, in selecting effective elastic moduli of monoclinic symmetric materials, C_{44}, C_{45}, and C_{55} are selected from the elastic moduli reconstructed using A_0 mode velocities and the rest of the ten elastic moduli are selected from the elastic moduli set reconstructed from S_0 mode velocities. After calculating the effective elastic moduli, the method proposed by Cowin and co-workers[28,29] was used to find the symmetry and orientation of principal planes of the material.

From the simulations, it was observed that the maximum error in the reconstruction of elastic moduli was 3.6% and the error in the estimation of Euler angle was less than 1.5%. In principle, the error in the reconstruction can be reduced by reducing the velocity step size in the inversion algorithm but at the cost of high computation. From the inversions of experimental measured velocity, it was observed that the maximum error in the reconstruction of elastic moduli was less than 7% and the error in the angle reconstruction was 3.7%. The error in the experimental data reconstruction includes the error from the GA inversion method and the error in measuring phase velocity. Small errors were observed in the experimental velocities that may be attributed to the error in measuring the STMR array radii. An error of 1 mm in measuring the difference in the STMR array radii causes approximately 3% error in the measurement of the velocity. The above error can be minimized by going in for larger radii but a bigger footprint of STMR array may not be feasible on all structures.

The empirical nature of the threshold used in the inversion method may be refined using a more rigorous approach such as Bayesian framework. The determination of regions of sensitivity, for the elastic constant inversion, is based on similar approach by many previous works[33–38] for ultrasonic based elastic moduli reconstruction. However, an improved approach to the analytical determination of the regions of sensitivity, based on the Rayleigh–Lamb equations for monoclinic or orthotropic media, should be attempted using methods such as adjoint field approach. However, these are the areas of future development of the blind inversion approach for the determination of material symmetry and elastic moduli of arbitrarily oriented anisotropic composite plates.

The compact STMR array patch can be mounted onto the structure for the *in situ* measurement of changes in the

TABLE XII. Part of GA parameters used in the inversion procedure.

Part of the GA parameters used in the inversion procedure	
No. of generations	200
Population size	25
Best population	10
Cross over rate	1
Cross over type	Single point
No. of elites	1

TABLE XIII. Low mutation with creep amount as 0.01.

GA parameters	1	2	3	4	5	6	7	8	9	10	11	12
Mutation chance	0.10	0.10	0.10	0.15	0.15	0.15	0.20	0.20	0.20	0.25	0.25	0.25
Creep chance	0.40	0.50	0.60	0.40	0.50	0.60	0.40	0.50	0.60	0.40	0.50	0.60
Creep amount	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01

elastic moduli during the lifetime monitoring of structures. The blind inversion technique using the STMR array patch finds application in the area of (i) material characterization, where the sensor patch can be used as an effective sensor which can be mounted onto the structure to find elastic moduli, material symmetry, and orientation of principal planes, and (ii) SHM, where the sensor patch can provide orientation of principal planes, which can be used in placing the sensor patch onto the structure, along with the intrinsic symmetry of the material and corresponding elastic moduli. The above applications can be implemented on anisotropic platelike structures used in aerospace and automobile components made using fiber reinforced composites.

ACKNOWLEDGMENTS

The authors thank the Advanced Composite Division of the National Aeronautical Laboratory, Bangalore for providing the composite samples and Department of Science and Technology (DST), New Delhi for funding this research work.

APPENDIX

Part of the real coded GA parameters has been listed in Table XII and the rest of the GA parameters can be found in Table XIII. More details on the use of GA can be obtained in the book by Goldberg.[24]

[1]P. D. Wilcox, M. Lowe, and P. Cawley, "Lamb and SH wave transducer arrays for the inspection of large areas of thick plates," Rev. Prog. Quant. Nondestr. Eval. **19A**, 1049–1056 (1999).
[2]P. D. Wilcox, "Guided wave beam steering from omnidirectional transducer arrays," Rev. Prog. Quant. Nondestr. Eval. **22A**, 761–768 (2002).
[3]P. D. Wilcox, "Omnidirectional guided wave transducer arrays for the rapid inspection of large areas of plate structures," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **50**, 699–709 (2003).
[4]R. Jagannathan, K. Balasubramaniam, and C. V. Krishnamurthy, "A single transmitter multi-receiver (STMR) PZT array for guided ultrasonic wave based structural health monitoring of large isotropic plate structures," Smart Mater. Struct. **15**, 1190–1196 (2006).
[5]H. Sohn, G. Park, J. R. Wait, N. P. Limback, and C. R. Farrar, "Wavelet-based active sensing for delamination detection in composite structures," Smart Mater. Struct. **13**, 153–160 (2004).
[6]R. Jagannathan, K. Balasubramaniam, and C. V. Krishnamurthy, "A phase reconstruction algorithm for Lamb wave based structural health monitoring of anisotropic multilayered composite plates," J. Acoust. Soc. Am. **119**, 872–878 (2005).
[7]J. Vishnuvardhan, A. Muralidharan, C. V. Krishnamurthy, and K. Balasubramaniam, "SHM of orthotropic plates through an ultrasonic guided wave STMR array patch," Rev. Prog. Quant. Nondestr. Eval. **27**, 1445–1452 (2007).
[8]A. Muralidharan, K. Balasubramaniam, and C. V. Krishnamurthy, "A migration based reconstruction algorithm for the imaging of defects in a plate using a compact array," Smart Structures and Systems **4**, 400–415 (2008).
[9]P. M. V. Subbarao, P. Munshi, and K. Muralidhar, "Performance of iterative tomographic algorithms applied to non-destructive evaluation with limited data," NDT & E Int. **30**, 359–370 (1996).
[10]S. Mahadev Prasad, K. Balasubramaniam, and C. V. Krishnamurthy, "Structural health monitoring of composite structures using Lamb wave tomography," Smart Mater. Struct. **13**, N73-N79 (2004).
[11]J. L. Rose, *Ultrasonic Waves in Solid Media* (Cambridge University Press, Cambridge, 1999).
[12]A. H. Nayfeh, *Wave Propagation in Layered Anisotropic Media With Application to Composites* (Elsevier, Amsterdam, 1995).
[13]J. E. Michaels, A. J. Croxford, and P. D. Wilcox, "Imaging algorithms for locating damage via in situ ultrasonic sensors," IEEE Sensors Application Symposium (2008), pp. 63–67.
[14]T. E. Michaels and J. E. Michaels, "Sparse ultrasonic transducer array for structural health monitoring," Rev. Prog. Quant. Nondestr. Eval. **23**, 1468–1475 (2004).
[15]P. D. Wilcox, G. Konstantinidis, A. J. Croxford, and B. W. Drinkwater, "Strategies for guided wave structural health monitoring," Rev. Prog. Quant. Nondestr. Eval. **26**, pp. 1469–1476 (2006).
[16]J. Vishnuvardhan, C. V. Krishnamurthy, and K. Balasubramaniam, "Determination of material symmetries from ultrasonic velocity measurements: A genetic algorithm based blind inversion method," Compos. Sci. Technol. **68**, 862–871 (2008).
[17]C. K. Jen, J. F. Bussiere, G. W. Farnell, E. L. Adler, and M. Esonu, "Elastic constants evaluation using the dispersive property of acoustic waves," Rev. Prog. Quant. Nondestr. Eval. **4**, 889–900 (1984).
[18]W. P. Rogers, "Elastic property measurement using Rayleigh-Lamb waves," Res. Nondestruct. Eval. **6**, 185–208 (1995).
[19]M. R. Karim, A. K. Mal, and Y. Bar-Cohen, "Inversion of leaky Lamb wave data by simplex algorithm," J. Acoust. Soc. Am. **88**, 482–491 (1990).
[20]S. I. Rokhlin and D. E. Chimenti, "Reconstruction of elastic constants from ultrasonic reflectivity data in a fluid coupled composite plate," Rev. Prog. Quant. Nondestr. Eval. **9B**, 1411–1418 (1990).
[21]S. I. Rokhlin, C. Y. Wu, and L. Wang, "Application of coupled ultrasonic plate modes for elastic constant reconstruction of anisotropic composites," Rev. Prog. Quant. Nondestr. Eval. **9B**, 1403–1410 (1990).
[22]N. S. Rao, "Inverse problems in ultrasonic non-destructive characterization of composite materials using genetic algorithm," MS thesis, Mississippi State University, Mississippi State, MS (1997).
[23]J. Vishnuvardhan, C. V. Krishnamurthy, and K. Balasubramaniam, "Genetic algorithm based reconstruction of the elastic moduli of orthotropic plates using an ultrasonic guided wave single-transmitter-multiple-receiver SHM array," Smart Mater. Struct. **16**, 1639–1650 (2007).
[24]D. E. Goldberg, *Genetic Algorithms in Search, Optimization, and Machine Learning* (Addison-Wesley, Reading, MA, 1989).
[25]W. P. Ralph, "Optimization for engineering systems," http://www.mpri.lsu.edu/bookindex.html (Last viewed October, 2008).
[26]P. A. W. Lewis and E. J. Orav, *Simulation Methodology for Statisticians, Operations Analysts, and Engineers* (CRC, Boca Raton, FL, 1989).
[27]D. G. Isaak, I. Ohno, and P. C. Lee, "The elastic constants of monoclinic single-crystal chrome-diopside to 1300K," Phys. Chem. Miner. **32**, 691–699 (2006).
[28]S. C. Cowin and M. M. Mehrabadi, "On the identification of material symmetry for anisotropic elastic materials," Q. J. Mech. Appl. Math. **40**, 451–476 (1987).
[29]S. C. Cowin, "Properties of the anisotropic elasticity tensor," Q. J. Mech. Appl. Math. **42**, 249–266 (1989).
[30]M. Sun, "Optical recovery of elastic properties for a general anisotropic material through ultrasonic measurements," MS thesis, University of Maine, Orono, ME (2002).
[31]J. Vishnuvardhan, C. V. Krishnamurthy, and K. Balasubramaniam, "Genetic algorithm based reconstruction of elastic constants of orthotropic fibre-reinforced composite plates from ultrasonic velocity data from a single non-symmetry plane," Composites, Part B **38**, 216–227 (2007).
[32]T. D. Lhermitte and B. Perrin, "Anisotropy of the elastic properties of crossply fiber-reinforced composite materials," Proc.-IEEE Ultrason. Symp. **2**, 825–830 (1991).
[33]S. I. Rokhlin and W. Wang, "Double through-transmission bulk wave

method for ultrasonic phase velocity measurement and determination of elastic constants of composite materials," J. Acoust. Soc. Am. **91**, 3303–3312 (1992).
[34]A. G. Every and W. Sachse, "Sensitivity of inversion algorithms for recovering elastic constants of anisotropic solids from longitudinal wave speed data," Ultrasonics **30**, 43–48 (1992).
[35]Y. C. Chu and S. I. Rokhlin, "Stability of determination of composite moduli from velocity data in planes of symmetry for weak and strong anisotropies," J. Acoust. Soc. Am. **95**, 213–225 (1994).
[36]Y. C. Chu, A. D. Degtyar, and S. I. Rokhlin, "On determination of orthotropic material moduli from ultrasonic velocity data in non-symmetry planes," J. Acoust. Soc. Am. **95**, 3191–3202 (1994).
[37]C. Aristegui and S. Baste, "Optimal recovery of the elasticity tensor of general anisotropic material from ultrasonic velocity data," J. Acoust. Soc. Am. **101**, 813–833 (1997).
[38]K. Balasubramaniam and N. S. Rao, "Inversion of composite material elastic constants from ultrasonic bulk wave phase velocity data using genetic algorithms," Composites, Part B **29B**, 171–180 (1998).

Reflection of plane elastic waves in tetragonal crystals with strong anisotropy

Vitaly B. Voloshinov[a] and Nataliya V. Polikarpova[b]
Department of Physics, M. V. Lomonosov Moscow State University, 119991 Moscow, Russia

Nico F. Declercq[c]
George W. Woodruff School of Mechanical Engineering, Georgia Institute of Technology, 801 Ferst Drive, Atlanta, Georgia 30332-0405 and Laboratory for Ultrasonic Nondestructive Evaluation Office 304, UMI Georgia Tech-CNRS 2958, Georgia Tech Lorraine, 2 rue Marconi, 57070 Metz-Technopole, France

(Received 24 April 2008; revised 20 November 2008; accepted 22 November 2008)

Propagation and reflection of plane elastic waves in the acousto-optic crystals tellurium dioxide, rutile, barium titanate, and mercury halides are examined in the paper. The reflection from a free and flat boundary separating the crystals and the vacuum is investigated in the (001) planes in the case of glancing acoustic incidence on the boundary. The analysis shows that two bulk elastic waves may be reflected from the crystal surface. The energy flow of one of the reflected waves in paratellurite and in the mercury compounds propagates in a quasi-back-direction with respect to the incident energy flow. It is proved that energy flows of the incident and reflected elastic waves are separated by a narrow angle of only a few degrees. It is also found that the relative intensity of the unusually reflected waves is close to a unit in a wide variety of crystal cuts. General conclusions related to acoustic propagation and reflection in crystals have been made based on the examined phenomena in the materials. © *2009 Acoustical Society of America.* [DOI: 10.1121/1.3050307]

PACS number(s): 43.35.Sx, 43.20.El [OAS] Pages: 772–779

I. INTRODUCTION

Even though it is known that there exists a strong acoustic energy walk-off in certain crystalline materials, the influence of the walk-off on the process of reflection in terms of direction and intensity has not been profoundly studied before. This paper examines regular trends of propagation and unusual reflection of harmonic homogeneous plane elastic waves in a family of tetragonal dielectric crystals. The crystalline materials chosen for the analysis are widely used in modern acousto-optic devices intended for applications in optical engineering and laser technology.[1–3,17] In particular, the crystals considered in this paper are tellurium dioxide (TeO_2), mercury halides (Hg_2Cl_2, Hg_2Br_2, and Hg_2I_2), rutile (TiO_2), and barium titanate ($BaTiO_3$). It was found that the crystals based on mercury and tellurium demonstrate extremely strong anisotropy of their elastic properties. On the contrary, rutile and barium titanate, similar to the majority of acoustic crystalline materials, are characterized by a moderate and low grade of anisotropy.[1–7]

It is known that the phase velocity of bulk acoustic waves V strongly depends on the direction of propagation in crystals possessing pronounced elastic anisotropy.[4–9] For example, a slow-shear acoustic wave propagating in the XY plane of tellurium dioxide at the angle $\theta=45°$ relative to the axis X has an extraordinary low magnitude of the phase velocity $V=616$ m/s. However, the same acoustic mode sent along the axis X of the material is characterized by a velocity up to five times higher $V=3050$ m/s.[1–3] As for the crystals with a low grade of anisotropy, the ratio r of maximal and minimal velocity magnitudes does not exceed a factor of 2 yielding the square of this ratio $A=4.0$.[1,2]

Elastic waves propagating in strongly anisotropic crystals are characterized by large walk-off angles ψ between the vectors of phase $\boldsymbol{V}$ and group $\boldsymbol{V}_g$ velocities of ultrasound.[1–10] It was found that the walk-off angle ψ between the acoustic wave vector $\boldsymbol{K}$ and the energy flow (Poynting) vector in these materials might reach magnitudes $\psi>70°$. On the other hand, in crystals with moderate and low anisotropy, e.g., in $BaTiO_3$, the walk-off angle is limited to $\psi=50°$.[1–16] As recently predicted theoretically [13,14] and confirmed experimentally in the crystals of tellurium dioxide,[15] propagation of waves with large acoustic walk-off angles may give rise to a significant unusual reflection of the elastic energy from a free boundary separating the material and the vacuum.[15,16]

In this paper, the peculiar reflection of bulk and plane elastic waves is studied in the XY plane of the acousto-optic materials though the analysis may easily be extended to crystals of other classes. In order to examine regular trends of unusual reflection and to reveal the influence of elastic anisotropy on the reflection process, the investigation was carried out not in a single but in a family of crystalline materials characterized by a different grade of elastic anisotropy.

II. GLANCING INCIDENCE AND REFLECTION IN CRYSTALS

The analysis carried out in this paper is related to a rather peculiar case of acoustic reflection in crystals when a crystalline specimen is cut in the form of a rectangular prism. Figure 1 shows the rectangular specimen of a tetragonal

[a]Electronic mail: volosh@phys.msu.ru
[b]Electronic mail: polikarp@phys.msu.ru
[c]Author to whom correspondence should be addressed. Electronic mail: nico.declercq@me.gatech.edu

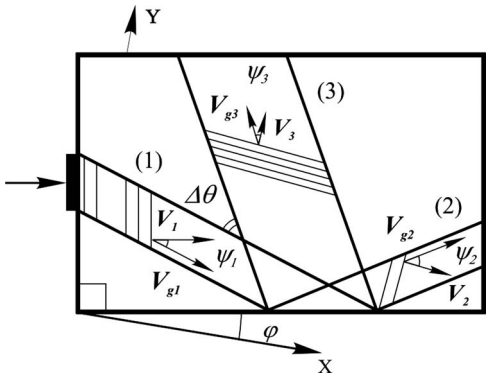


FIG. 1. General scheme of reflection in crystals with strong anisotropy.

crystal and the piezoelectric transducer bonded to the left side facet of the crystal. It is considered that the transducer generates a quasilongitudinal or quasishear acoustic wave (1) in the crystal. It is also assumed that the incident wave possesses a phase velocity vector $\boldsymbol{V}_1$ and a corresponding wave vector $\boldsymbol{K}_1$ directed at the angle φ relative to the X-axis [100]. As illustrated, the acoustic beam (1) propagates in the material at a walk-off angle ψ_1 evaluated between the vectors of phase velocity $\boldsymbol{V}_1$ and group velocity $\boldsymbol{V}_{g1}$ of ultrasound. After propagation through the crystal, the acoustic energy is incident on the bottom facet of the specimen. This facet of the crystal serves as a solid-vacuum interface. Therefore the acoustic energy of the (bulk) harmonic homogeneous plane elastic wave is incident on a free and smooth, homogeneous surface.

According to everyday experience, one would be tempted to intuitively predict that the reflected beam (2) propagates from the bottom surface of the crystal away from the direction of the acoustic incidence. However, as seen in the figure, the beam (2) is characterized by the walk-off angle ψ_2. This angle is evaluated between the wave vector $\boldsymbol{K}_2$, or the phase velocity vector $\boldsymbol{V}_2$, on the one hand, and the group velocity vector $\boldsymbol{V}_{g2}$, on the other hand. However, it was recently found [13–16] that, in addition to the traditionally reflected beam (2), there appears an unusually reflected beam (3). It possesses a phase velocity vector $\boldsymbol{V}_3$ corresponding to the wave vector $\boldsymbol{K}_3$. The extraordinary reflected beam (3) and the incident ray (1) are separated in space by the space separation angle $\Delta\theta$. As predicted in Refs. 6, 13, and 14, one of the two reflected waves, in particular, the wave (3), may propagate approximately toward the incident acoustic ray, i.e., in back direction relative to the direction of the incidence.

III. PHASE VELOCITIES OF ULTRASOUND IN THE *XY*-PLANE OF CRYSTALS

In order to understand the origin of the unusual reflection in the crystals, the dependences of magnitudes $V(\theta)$ of the acoustic phase velocities on the direction of propagation in the XY-plane of tetragonal materials were calculated. The velocity vectors of the quasishear acoustic waves $\boldsymbol{V}_s$ and the quasilongitudinal acoustic waves $\boldsymbol{V}_l$ propagating at the angle θ relative to the axis [100] have been obtained from Christoffel's equation. [1,2] The elastic coefficients of the materials used in the calculations are summarized in Table I. Data on the coefficients were found in the literature. [1–3,5]

It is shown [1,2,11–14] that the value of the acoustic phase velocity of waves in the XY-plane of the crystals can be expressed as

$$V_{1,2}^2 = (1/2\rho)(c_{11} + c_{66} \pm \sqrt{(c_{11} - c_{66})^2 \cos^2 2\theta + (c_{12} + c_{66})^2 \sin^2 2\theta}),$$
$$V_3^2 = c_{44}/\rho. \tag{1}$$

The acoustic slowness curves[1] $1/V_1$ and $1/V_2$ in the XY-plane of $BaTiO_3$ and TiO_2 are plotted in Figs. 2(a) and 2(b) while data in Figs. 2(c) and 2(d) illustrate the acoustic slowness curves corresponding to Hg_2Cl_2 and TeO_2. It can also be seen that paratellurite and calomel definitely demonstrate a very strong anisotropy of their elastic properties compared to TiO_2 and $BaTiO_3$. The ratios A of the second power of the maximum and minimum magnitudes describing the shear velocities V_2 in the XY-plane of calomel and tellurium dioxide are equal to $A=22.1$ and $A=24.0$. In comparison, the single crystal of barium titanate possesses a velocity ratio $A=2.3$. As for the crystals of TiO_2, they are characterized by the coefficient $A=4.0$. It is quite evident that the

TABLE I. Density and elastic coefficients of tetragonal crystals.

Crystal	Density (10^3 kg/m^3) ρ	Elastic coefficients (10^{10} N/m^2) c_{11}	c_{12}	c_{13}	c_{33}	c_{44}	c_{66}
Barium titanate ($BaTiO_3$)	6.00	27.5	17.9	15.2	16.5	5.43	11.3
Rutile (TiO_2)	4.30	27.3	17.6	14.9	48.4	12.5	19.4
Paratellurite (TeO_2)	6.00	5.60	5.15	2.20	10.6	2.65	6.60
Calomel (Hg_2Cl_2)	7.18	1.89	1.71	1.56	8.03	0.84	1.22
Mercury bromide (Hg_2Br_2)	7.31	1.61	1.50	1.88	8.88	0.74	1.11
Mercury iodide (Hg_2I_2)	7.70	1.42	1.32	2.20	10.70	0.58	1.11

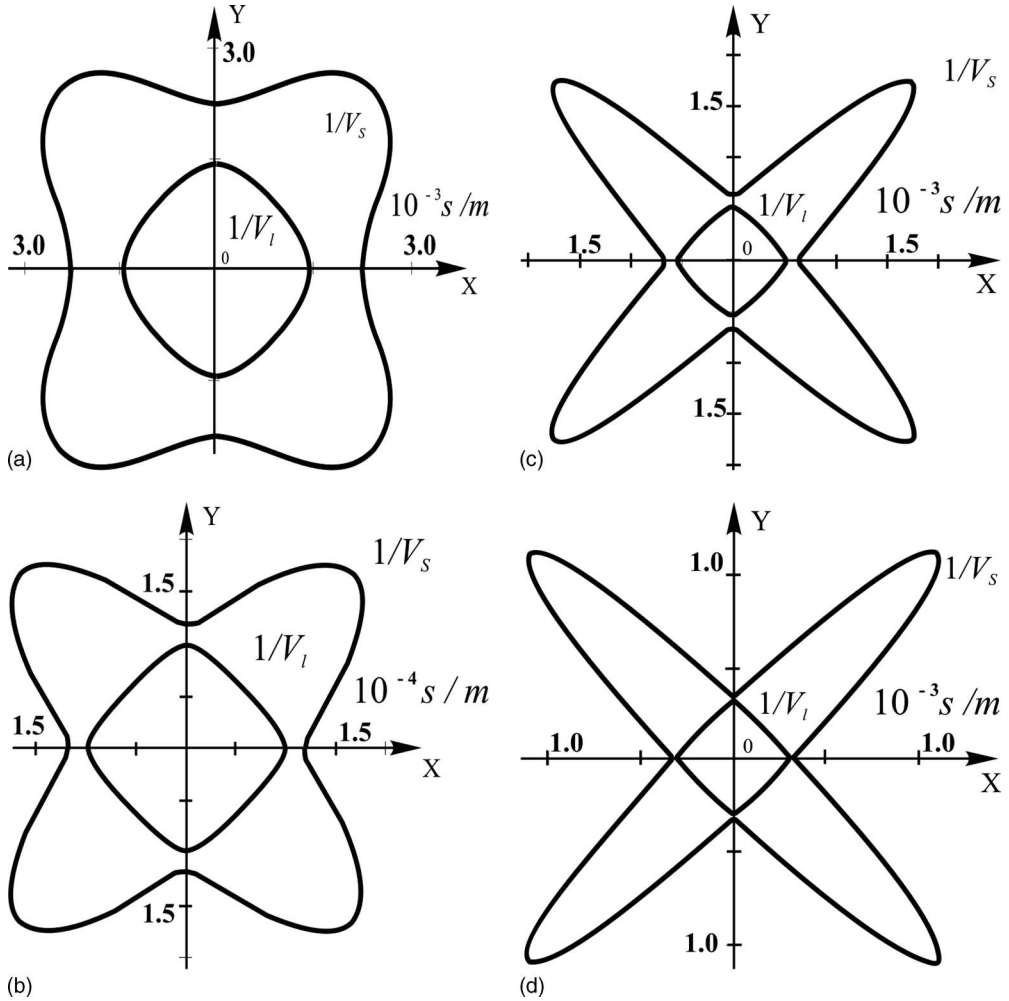


FIG. 2. Acoustic slowness curves in the *XY* plane of (a) $BaTiO_3$, (b) TiO_2, (c) Hg_2Cl_2, and (d) TeO_2.

elastic anisotropy of a crystal may be qualitatively evaluated just from a general view of the acoustic slowness curves.

IV. ACOUSTIC WALK-OFF ANGLE IN THE CRYSTALS

In addition to the magnitudes of the phase velocity V, the acoustic walk-off angles ψ in the *XY*-plane of the investigated crystals were calculated following the traditional procedure.[1,2] As proved by the calculations, the maximum acoustic walk-off angle for the slow-shear wave in this plane of tellurium dioxide may be amazingly wide, $\psi=74°$.[4,8–10] On the other hand, the quasilongitudinal mode with a velocity value V_1 possesses a smaller walk-off angle $\psi=35°$. As for the mercury halides, the behavior of the acoustic waves, in many aspects, is similar to paratellurite. For example, the maximum walk-off angle in the *XY*-plane of Hg_2Cl_2 is equal to $\psi=70°$. In the other two mercury materials, i.e., mercury bromide and mercury iodide, the magnitudes of the maximum walk-off angles are between the limits $\psi=70°$ and $\psi=74°$. In barium titanate and rutile, the acoustic walk-off angles are less than $\psi=50°$. This regular trend is of importance in understanding the phenomenon of acoustic reflection in crystals.

V. DIRECTIONS OF PROPAGATION OF REFLECTED WAVES

It is seen in Fig. 1 that the energy flow of the acoustic wave (1) generated by the transducer is incident on the bottom facet of the sample. The schematic presented in Fig. 3 shows the directions of the wave vectors $\boldsymbol{K}_2$ and $\boldsymbol{K}_3$ corresponding to the two reflected beams [(2) and (3)] in Fig. 1. It is indicated in Fig. 1 that the transducer launches the incident

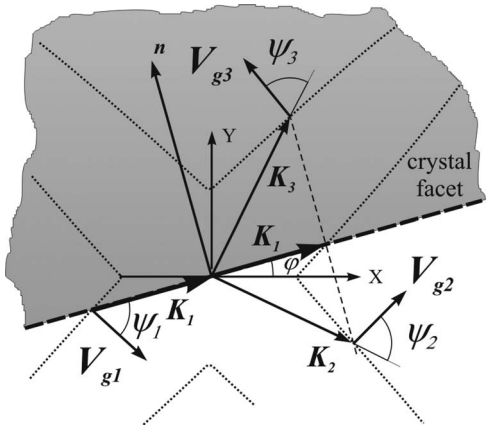


FIG. 3. Directions of wave vectors and group velocities of ultrasound.

wave (1) at the angle φ relative to the axis X. It is also clear that the wave vector $\boldsymbol{K}_1$ of the incident beam is directed parallel to the bottom facet of the crystal. This facet is shown in Fig. 3 by the bold dashed line that forms an angle φ with the X-axis. The body of the crystal is shown in Fig. 3 by the gray background, while the boundary is indicated by the label "crystal facet." The vector $\boldsymbol{n}$ in the figure is normal to the boundary separating the crystal and the vacuum. The length of the projection of the wave vector $\boldsymbol{K}_1$ on the boundary for the incident wave is equal just to the length of the wave vector itself because the specimen is rectangular and the investigated case of acoustic incidence may be defined as glancing or grazing. The projection of this vector on the boundary is shown to the left of the normal.

A. Directions of wave vectors of the reflected beams

According to the general laws of wave theory, directions of the two reflected wave vectors $\boldsymbol{K}_2$ and $\boldsymbol{K}_3$ in Fig. 3 may be found from the known condition, usually referred to as the law of Snell–Descartes, of equal projections of the incident and reflected wave vectors on the crystal facet.[1,2,14] Detailed explanations of the method to find the directions of the wave vectors and group velocity vectors are presented in Ref. 14. In order to satisfy the condition of equal projections, a supplementary dashed line is plotted in the figure at the extremity of the vector $\boldsymbol{K}_1$. The vector $\boldsymbol{K}_1$ is plotted to the right of the normal $\boldsymbol{n}$. This vector is equal to the vector $\boldsymbol{K}_1$ shown to the left of the normal. The supplementary line is orthogonal to the boundary and parallel to the normal n. A dotted line representing the cross section of the acoustic wave vector surface in the XY plane of the crystal may be seen in Fig. 3. This curve directly follows from the slowness curve for the velocity V_S depicted in Fig. 2(d). Data in Fig. 3 prove that the supplementary line intersects with the dotted curve. Moreover, there are as much as two intersections of the supplementary line with the dotted line. As a result, the directions of the two acoustic wave vectors of the reflected waves $\boldsymbol{K}_2$ and $\boldsymbol{K}_3$ may be found if the intersection points are known. It is common to define the reflected wave (2) corresponding to the wave vector $\boldsymbol{K}_2$ as the "ordinary" reflected wave. The wave (3) described by the vector $\boldsymbol{K}_3$ may be defined as the "extraordinary" reflected wave. It should be emphasized that the tangential projections, on the boundary, of the wave vectors $\boldsymbol{K}_2$ and $\boldsymbol{K}_3$ are equal to the length of the wave vector $\boldsymbol{K}_1$ plotted to the right of the normal $\boldsymbol{n}$. As mentioned, the same vector $\boldsymbol{K}_1$, to the left of the normal, describes the incident wave.

B. Directions of energy flows of the reflected beams

Based on the presented considerations, it is possible to determine in Fig. 3 the directions of the acoustic energy flow before and after the reflection. As seen in Fig. 3, the angle between the acoustic wave vector $\boldsymbol{K}_1$ and the Poynting vector of the incident beam (1) is equal to ψ_1. The picture also shows that the group velocity vector $\boldsymbol{V}_{g1}$ of the beam (1) is orthogonal to the wave surface in the point where the vector $\boldsymbol{K}_1$ touches the surface of the wave vectors to the right of the normal $\boldsymbol{n}$. As for the reflected beam (2), the drawing proves

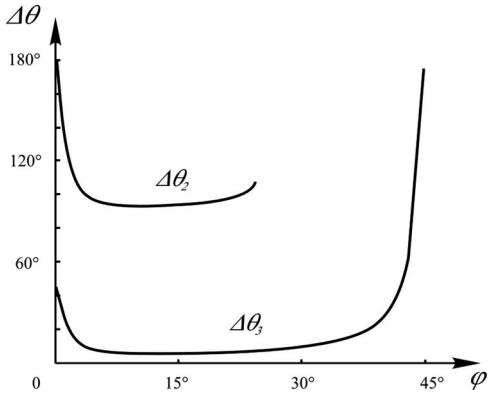


FIG. 4. Separation angle for the slow acoustic mode in TeO_2.

that the wave vector $\boldsymbol{K}_2$ is directed outside the crystal. Therefore, acoustic wave fronts of the wave (2) in Figs. 3 and 1 are tilted clockwise relative to the boundary. On the other hand, the energy flow of the reflected beam (2) propagates at the angle ψ_2 inside the crystal so that the reflected beam is directed away from the boundary and from the incident beam (1).

In order to understand the origin of the peculiar reflection at the boundary, the direction of the group velocity $\boldsymbol{V}_{g3}$ of the reflected wave (3) must be examined. The unusually reflected wave (3) is represented in Fig. 3 by the wave vector $\boldsymbol{K}_3$ and the acoustic walk-off angle ψ_3. As proved by Fig. 3, the wave vector $\boldsymbol{K}_3$ is directed inside the crystal and away from the bottom facet of the specimen. In this respect, the reflection takes place in accordance with the expectations. However, the energy flow of the reflected wave (3) propagates backward with respect to the energy flow of the incident wave (1) because the group velocity vectors $\boldsymbol{V}_{g3}$ and $\boldsymbol{V}_{g1}$ are practically directed opposite to each other.

It has been shown recently that the revealed peculiarity originates from the extremely large value of the acoustic walk-off angle ψ_3 while the large walk-off angle is the consequence of the strong elastic anisotropy of the acousto-optic materials under consideration.[13–16]

VI. THE SPATIAL SEPARATION ANGLE BETWEEN TWO REFLECTED WAVES

The magnitudes of the acoustic walk-off angles ψ_1, ψ_2, and ψ_3 in the XY plane of the crystals for the three acoustic waves were calculated. The calculation was carried out for directions of the incident wave propagation corresponding to the cut angle $0<\varphi<45°$. In other words, due to the symmetry of the tetragonal crystals, the analysis included practically all orientations of the rectangular sample with respect to the crystalline axes. As for the directions with $\varphi=0$ and $\varphi=45°$, they were not considered because the acoustic waves propagate along these directions without energy walk-off and hence without any reflection from the bottom facet of the specimen.

During the carried out analysis, similar to that presented in this paper,[14] the angles $\Delta\theta_2$ and $\Delta\theta_3$ between the energy flows of the incident (1), the ordinary (2), and the extraordinary (3) reflected slow-shear waves were determined for

TABLE II. Trends of the materials depending on the grade of the elastic anisotropy in tetragonal crystals during the incidence of slow incident acoustic mode.

Crystals	Walk-off angle ψ (°)	Anisotropy parameter $A=(V_{max}/V_{min})^2$	Separation angle $\Delta\theta_{min}$ (°)	Critical angle φ_C (°)	Maximum reflection coefficient R_3
Barium titanate ($BaTiO_3$)	$\psi_1=40$	2.3	72.8	9.7	0.05
Rutile (TiO_2)	$\psi_1=51$	4.0	38.6	14.3	0.43
Mercury bromide (Hg_2Br_2)	$\psi_1=72$	19.4	7.2	23.7	0.96
Calomel (Hg_2Cl_2)	$\psi_1=70$	22.1	9.7	22.324	0.93
Mercury iodide (Hg_2I_2)	$\psi_1=74$	24.0	5.6	24.060	0.97
Paratellurite (TeO_2)	$\psi_1=74$	24.0	5.3	24.131	1.00

each value of the cut angle φ. The data in Fig. 4 demonstrate the obtained dependences of the separation angles $\Delta\theta_2$ and $\Delta\theta_3$ on the cut angle of the crystals. It may be seen in Fig. 4 that the angle $\Delta\theta_3$ in tellurium dioxide is amazingly small $\Delta\theta_3 \leqslant 10°$ over the wide range of the cut angles $4° < \varphi < 32°$. It means that there are no strict requirements on the orientation of a specimen in order to observe the peculiar reflection in the crystal. The data in Fig. 4 also represent the behavior of the spatial separation angle $\Delta\theta_2$ for the ordinary reflected wave (2). As proved by the analysis, this wave exists in TeO_2 samples cut at the angle $0 < \varphi < 24°$.

Similar calculations were made for the fast acoustic mode in paratellurite and also in the single crystals of calomel, mercury bromide, and mercury iodide. As found, the separation angle for the extraordinary reflected wave in the crystals also does not exceed the value $\Delta\theta_3 \leqslant 10°$ in a wide range of cut angles. However, the separation angle between the incident and reflected waves in rutile and barium titanate is equal to dozens of degrees, thus indicating that there is no backreflection of the elastic waves in the crystals with moderate and low anisotropy.

The performed analysis proves that the minimum value of the separation angle in tellurium dioxide occurs as low as $\Delta\theta_3=5.3°$, while in the mercury halides the angle is only slightly wider: in calomel it is limited to $\Delta\theta_3=9.7°$. Therefore, the research revealed that the regular trend of "near backreflection" of acoustic waves is typical for all crystalline materials possessing strong anisotropy of the elastic properties. For example, in the crystal $BaTiO_3$ the minimum value of the separation angle is as large as $\Delta\theta_3=72.8°$. In the crystal rutile, the angle is equal to $\Delta\theta_3=38.6°$. It means that, in the crystals with moderate anisotropy, the acoustic reflection takes place in the forward and not in the backward direction with respect to the incident energy flow.

The analysis was extended to the cases of glancing incidence and reflection of the fast, i.e., the quasilongitudinal, waves in the tetragonal crystals. Similar to the reflection of the quasishear waves, the separation angles in the materials with the pronounced elastic anisotropy are narrow. It was also found that one of the two reflected waves in paratellurite is a quasishear acoustic mode while the other wave is a quasilongitudinal mode. The minimal separation angle in TeO_2 is equal to $\Delta\theta_3=5.9°$. In the crystal Hg_2Cl_2, this angle is slightly wider $\Delta\theta_3=12.2°$. However, in the materials with low anisotropy, for example, in $BaTiO_3$, the separation angle is equal to $\Delta\theta_3=73.9°$. Data on results of calculations of the separation angles for the slow incident acoustic mode are included in Table II. Therefore, based on the carried out analysis, it is possible to conclude that the quasi-backreflection of the acoustic waves is a phenomenon typical of the quasishear and the quasilongitudinal mode in strongly anisotropic media.

VII. THE DISTRIBUTION OF ELASTIC ENERGY OVER REFLECTED WAVES

Analysis of the phenomenon of the unusual reflection of waves required evaluation of the amount of energy flow reflected from the free and flat boundary in the form of the extraordinary (3) wave. In order to fulfill the analysis, mutual distribution of the incident elastic energy over the two reflected waves was determined. The reflection coefficients R_2 and R_3 describing energy flows of the reflected beams (2) and (3) were calculated for this purpose. The method of calculation of the reflection coefficients was proposed in Ref. 6. Application of the method to the tetragonal crystals is discussed in detail in Refs. 13 and 14. In this paper, we mainly concentrate on discussion of the calculation results related to the chosen family of crystals. Each coefficient was defined as a ratio of the normal projections of the energy flows in the corresponding reflected and incident waves.[1,2,6] Therefore, it was stated that the following relation $R_2+R_3=1$ was valid in the crystal. It means that the energy of the incident bulk acoustic wave is considered equal to unity. This energy is totally distributed over the energy flows of the two bulk reflected waves. For simplicity, a possibility of appearance of surface and inhomogeneous (evanescent) waves is ignored. The analysis also ignores piezoelectric and other effects that

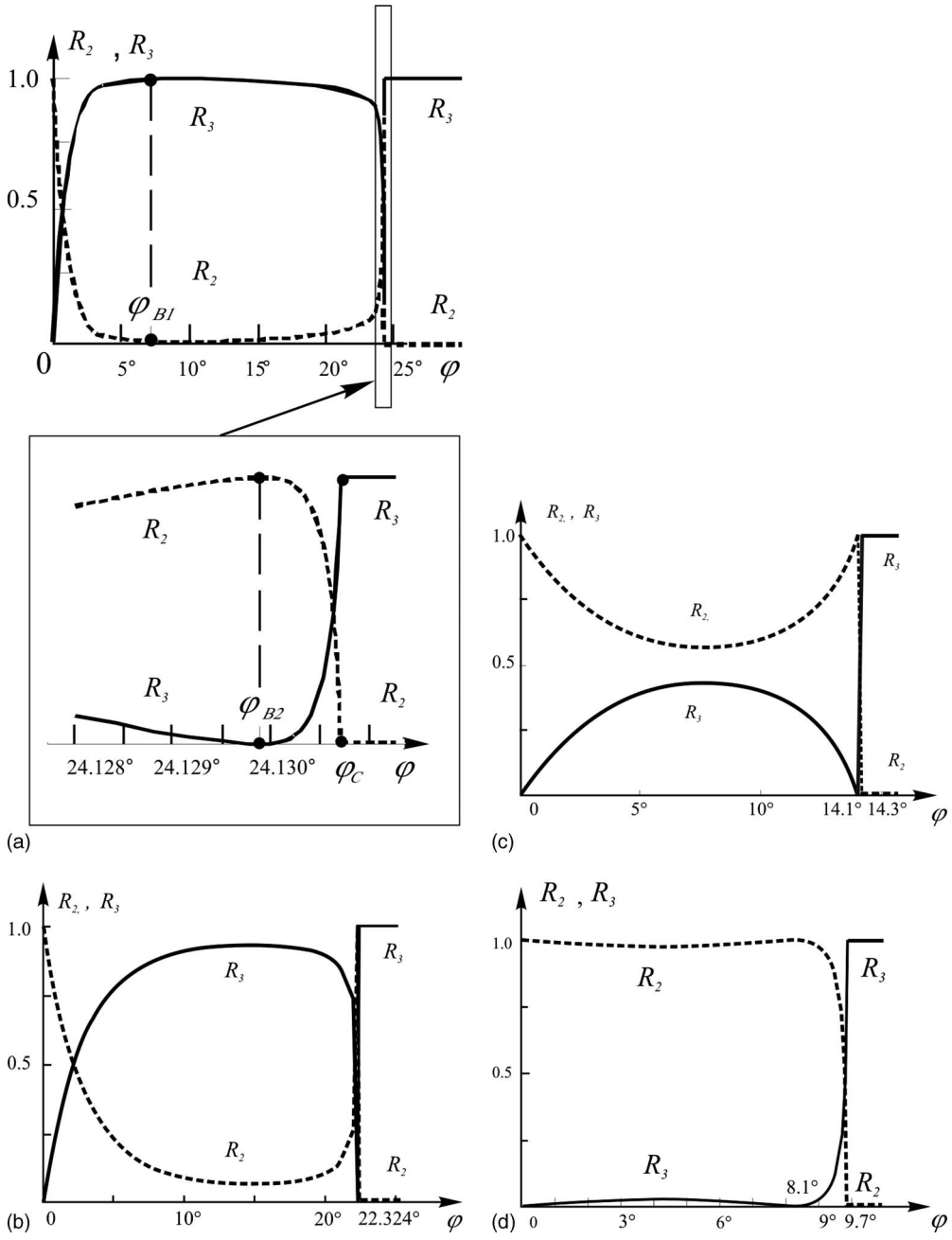


FIG. 5. Reflection coefficients of the slow acoustic mode in the *XY* plane of (a) TeO_2, (b) Hg_2Cl_2, (c) TiO_2, and (d) $BaTiO_3$.

might influence the process of reflection. This is justifiable because in a crystal like tellurium dioxide the piezoelectric effect is relatively small, almost negligible in fact. The complicated calculation required the evaluation of components of the stress tensor and of the acoustic displacement vectors in the examined crystals. For such purposes a relation between the stress tensor represented in the system of coordinates of the boundary and crystalline axes is found.

The data in Fig. 5(a) show calculated dependences of the reflection coefficients R_2 and R_3 on the cut angle φ in tellurium dioxide. It is seen that the range of changes of the cut angles $0<\varphi<45°$ for convenience may be divided into four different intervals. These intervals correspond to different types of dependences of the reflection coefficients on the cut angle. The analysis proves that if the crystal is cut at the angle φ slightly different from zero,the major amount of the incident elastic energy is reflected in the form of the ordinary elastic wave, i.e., the wave (2) in Fig. 1. Growth of the angle φ is accompanied by an increase in the energy flow of the extraordinary reflected wave (3), while the intensity of the forward reflected wave (2) vanishes.

If the propagation angle φ is limited to the range $3<\varphi \leqslant 25°$ then the major amount of the reflected energy is concentrated in the extraordinary reflected ray. Calculation proves that at the angle $\varphi_{B1}=7.5°$ all incident elastic energy is reflected in the quasi-back-direction because the coefficient $R_3=1$. Consequently, the cut angle φ_{B1}, similar to optics, may be defined as the Brewster angle. This definition of the angle seems possible because one of the reflected waves, i.e., wave (2), at this particular angle possesses zero

intensity.[14] A comparable phenomenon was found earlier in acoustics for diffraction on a corrugated surface.[18]

Analysis of the data in Fig. 5(a) proves that the interval of angles $23° < \varphi \leqslant \varphi_C$ is characterized by abrupt changes of the reflection coefficients. This interval includes two characteristic points, one of which is the Brewster angle φ_{B2}, while the other one is the critical angle φ_C. As known, at $\varphi > \varphi_C$ there exists, in a crystal, only a single reflected acoustic wave characterized by the reflection coefficient equal to unity.[1] The performed calculations prove that, in tellurium dioxide, the magnitudes of the second Brewster and the critical angle are very close to each other $\varphi_{B2} \approx 24.130°$ and $\varphi_C \approx 24.131°$. Consequently, it is unlikely to distinguish between the two angles experimentally. As for the distribution of the elastic energy over the two reflected waves, it is reasonable to predict that, in the interval of the angles from $\varphi = 23°$ to $\varphi_{B2} \approx 24.130°$, the reflection coefficient R_2 rapidly approaches unity, while the energy flow of the extraordinary reflected beam vanishes. It means that the total reflection of the incident energy into the ordinary wave (2) may be expected in the specimen at $\varphi = \varphi_{B2}$. Finally, a further increase in the cut angle from φ_{B2} to the critical angle φ_C is accompanied by the abrupt drop of the energy of the ordinary wave ($R_2 = 0$). It results in inevitable jump of the energy in the backreflected wave (3), for which the coefficient R_3 changes from zero to $R_3 = 1$.

As for the critical angle φ_C, it can be predicted that all incident elastic energy is reflected from the boundary as the quasibackward wave (3). Moreover, the investigation confirms that the forward wave (2) disappears at $\varphi > \varphi_C$ because the supplementary dashed line in Fig. 3 intersects with the acoustic wave surface only at one point. It is also quite likely that the ordinary reflected wave at the critical angle $\varphi = \varphi_C$ becomes inhomogeneous (evanescent). The wave propagates with the energy flow directed along the border of the crystal. It is obvious that all cases of the reflection at the cut angle $\varphi > \varphi_C$ correspond to a single extraordinary reflected acoustic wave (3). The reflection coefficient R_3 in this case is equal to $R_3 = 1$. The energy flow of this wave practically meets the energy flow of the incident beam because the separation angle $\Delta\theta_3$ in Fig. 4 is limited to $\Delta\theta_3 \leqslant 10^\circ$ over a wide range of the cut angles.

Calculations prove that the behavior of the waves in mercury halides remains similar to that in tellurium dioxide. Corresponding dependences of the reflection coefficient in the compounds of mercury are plotted in Fig. 5(b). It is seen that absolute magnitudes of the critical angles are only slightly different from those in TeO_2. For example, in calomel, the critical angle equals $\varphi_C = 22.324^\circ$. On the other hand, it was found that, in calomel, the maximal value of the reflection coefficient of the backreflected wave is equal to $R_3 = 0.93$ and not to $R_3 = 1$, as in paratellurite. It means that it is not reasonable to predict the existence of the first Brewster angle in mercury crystals because their elastic anisotropy is not as strong as in TeO_2. Nevertheless, the elastic anisotropy of the mercury materials is still sounding because the difference between the second Brewster angle φ_{B2} and the critical angle φ_C is negligibly small.

The carried out analysis proves that the unusual reflection of acoustic waves in the materials possessing moderate and low elastic anisotropy, for example, in TiO_2 and $BaTiO_3$, may not be observed at all. This statement is confirmed by data presented in Figs. 5(c) and 5(d) and in Table II. As seen, the maximal magnitude of the reflection coefficient for the reflected wave (3) in TiO_2 and $BaTiO_3$ is very low, $R_3 = 0.43$ and $R_3 = 0.05$, respectively. The analysis also demonstrates that the magnitudes of the Brewster and the critical angles in the ferroelectric material appear smaller than in paratellurite and calomel: they are equal to $\varphi_{B2} = 14.1^\circ$ and $\varphi_C = 14.3^\circ$ in TiO_2 while $\varphi_{B2} = 8.1^\circ$ and $\varphi_C = 9.7^\circ$ in $BaTiO_3$. Consequently, the difference between these two angles in TiO_2 and $BaTiO_3$ is quite noticeable in comparison with tellurium dioxide or calomel. Moreover, the reflected wave (3) in barium titanate propagates at the angle $\Delta\theta_3$ exceeding 70°, i.e., far away from the energy flow of the incident wave. That is why the quasi-back-reflection of elastic waves in the commonly used acoustic crystals does not exist.

A similar analysis was carried out for the case of the fast acoustic modes in the crystals. It was found that, in a wide range of the cut angles in paratellurite, a major amount of the incident elastic energy is transformed into a quasishear wave. The maximal value of the reflection coefficient for this wave approaches a unit $R_3 = 0.94$. The reflection in TeO_2 samples cut at the angles close to $\varphi = 90^\circ$ is characterized by growth of the energy of the quasilongitudinal wave and by total decrease in the energy in the quasishear wave. In general, this reflection, contrary to the reflection of the quasishear waves, takes place in accordance with the expectations. It means that the Brewster and the critical angles may not be observed in the crystals. It also means that mutual distribution of elastic energy over the reflected waves changes continuously, i.e., without abrupt drops.

VIII. CONCLUSIONS

Basic results of the carried out analysis in tetragonal crystals are summarized in Table II. The examined crystalline materials are listed in the table in accordance with the anisotropy of their elastic properties. The ratio of the maximal and minimal acoustic phase velocity values r was used to describe the anisotropy quantitatively. It is seen in the table that the walk-off angle between the phase and group velocities of ultrasound increases with the growth of the parameter A. As found, the propagation of bulk acoustic waves in the crystalline materials possessing strong anisotropy of elastic properties may be accompanied by a peculiar quasi-back-reflection of the acoustic energy flux from a free surface separating the crystals and the surrounding vacuum. The angle between the incident and backreflected energy flows in the crystals may be as narrow as a few degrees. In commonly used materials, the angle is wider.

The major peculiarity of the examined case of acoustic reflection consists of the fact that the unusual reflection follows the process of glancing incidence of elastic waves on a free and flat boundary separating a crystal and the vacuum. This type of reflection was, up until now, practically not investigated in acoustic crystals and other anisotropic media.

As proved by the investigation and seen in Table II, the relative intensity of the extraordinary reflected waves depends on the elastic anisotropy of the crystals. In acousto-optic crystals with strong anisotropy, e.g., in tellurium dioxide and mercury halides, the efficiency of the unusual reflection may be as high as 1.00, i.e., 100%. On the other hand, in traditional acoustic materials with moderate anisotropy, the effect of backreflection is relatively weak. In crystals with a low grade of anisotropy, the unusual reflection of elastic energy is absent. Therefore, it may be stated that the greater the anisotropy the stronger the effect of backacoustic reflection.

It may also be concluded that the unusual cases of acoustic reflection are interesting not only from the point of view of physical acoustics but also of other fundamental sciences, optics, magnetism, and the theory of waves. It is reasonable to expect the existence of similar reflection phenomena in anisotropic media of another physical nature, e.g., in polymers, magnetic films, ionosphere, etc.[19] Observation of the examined effects may also be expected in nanomaterials. Finally, it is clear that new types of acoustoelectronic and acousto-optic instruments may be designed based on the examined effects.[8–12,20] Tunable acousto-optic filters with collinear and noncollinear propagations of beams as well as acoustoelectronic delay lines with low consumption of expensive crystalline materials are the most evident examples of the possible applications.

ACKNOWLEDGMENTS

The work of N.V.P. is supported in part by the grant of the President of the Russian Federation for young scientists MK-1358.2008.8.

[1]B. A. Auld, *Acoustic Fields and Waves in Solids* (Krieger, New York, 1990).
[2]E. Dieulesaint and D. Royer, *Ondes Elastiques dans les Solides (Elastic Waves in Solids)* (Masson, Paris, 1974).
[3]A. Goutzoulis and D. Pape, *Design and Fabrication of Acousto-Optic Devices* (Dekker, New York, 1994).
[4]J. C. Kastelik, M. G. Gazalet, C. Bruneel, and E. Bridoux, "Acoustic shear wave propagation in paratellurite with reduced spreading," J. Appl. Phys. **74**, 2813–2817 (1993).
[5]M. Gottlieb, A. Goutzoulis, and N. Singh, "High-performance acousto-optic materials: Hg_2Cl_2 and $PbBr_2$," Opt. Eng. (Bellingham) **31**, 2110–2117 (1992).
[6]M. J. P. Musgrave, "Refraction and reflection of plane elastic waves at a plane boundary between aeolotropic media," Geophys. J. R. Astron. Soc. **3**, 406–418 (1960).
[7]E. G. Lean and W. H. Chen, "Large angle beam sterring in acoustically anisotropic crystal," Appl. Phys. Lett. **35**, 101–103 (1979).
[8]V. B. Voloshinov, "Close to collinear acousto-optical interaction in paratellurite," Opt. Eng. (Bellingham) **31**, 2089—2094 (1992).
[9]V. B. Voloshinov, "Anisotropic light diffraction on ultrasound in a tellurium dioxide single crystal," Ultrasonics **31**, 333–338 (1993).
[10]J. Sapriel, D. Charissoux, V. B. Voloshinov, and V. Molchanov, "Tunable acousto-optic filters and equalizers for WDM applications," J. Lightwave Technol. **20**, 888–896 (2002).
[11]V. B. Voloshinov and N. V. Polikarpova, "Application of acousto-optic interactions in anisotropic media for control of light radiation," Acust. Acta Acust. **89**, 930–935 (2003).
[12]V. B. Voloshinov and N. V. Polikarpova, "Collinear tunable acousto-optic filters applying acoustically anisotropic material tellurium dioxide," Molecular and Quantum Acoustics, **24**, 225–235 (2003).
[13]N. V. Polikarpova and V. B. Voloshinov, "Intensity of reflected acoustic waves in acousto-optic crystal tellurium dioxide," Proc. SPIE **5828**, 25–36 (2004).
[14]V. B. Voloshinov, N. V. Polikarpova, and V. G. Mozhaev, "Nearly backward reflection of bulk acoustic waves at grazing incidence in a TeO_2 crystal," Acoust. Phys. **52**, 297–305 (2006); "Близкое к обратному отражение объемных акустических волн при скользящем падении в кристалле парателлурита", Akust. Zh. **52**, 245–251 (2006).
[15]V. B. Voloshinov, O. Yu. Makarov, and N. V. Polikarpova, "Nearly backward reflection of elastic waves in an acousto-optic crystal of paratellurite," Tech. Phys. Lett. **31**, 79–87 (2005); "Близкое к обратному отражение волн в акустооптическом кристалле парателлурита," Pis'ma Zh. Tekh. Fiz. **31**, 352–355 (2005).
[16]N. V. Polikarpova and V. B. Voloshinov, "Glancing incidence and back reflection of elastic waves in tetragonal crystals" Proc. SPIE **5953**, 0C1–0C12 (2005).
[17]A. V. Zakharov, N. V. Polikarpova, and E. Blomme, "Intermediate regime of light diffraction in media with strong elastic anisotropy" Proc. SPIE **5953**, 0D1–0D10 (2005).
[18]N. F. Declercq, R. Briers J. Degrieck, and O. Leroy "Diffraction of horizontally polarized ultrasonic plane waves on a periodically corrugated solid-liquid interface for normal incidence and Brewster angle incidence," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **49**, 1516–1521 (2002).
[19]N. F. Declercq, J. Degrieck, and O. Leroy, "Sound in biased piezoelectric materials of general anisotropy," Ann. Phys. **14**, 705–722 (2005).
[20]N. F. Declercq, N. V. Polikarpova, V. B. Voloshinov, O. Leroy, and J. Degrieck, "Enhanced anisotropy in paratellurite for inhomogeneous waves and its possible importance in the future development of acousto-optic devices," Ultrasonics **44**, 833–837 (2006).

Experimental evaluation of the acoustic properties of stacked-screen regenerators

Yuki Ueda[a)]
Department of Bio-Applications and Systems Engineering, Tokyo University of Agriculture and Technology, Nakacho 2-24-16, Koganei, Tokyo 187-8588, Japan

Toshihito Kato and Chisachi Kato
Institute of Industrial Science, The University of Tokyo, Komaba, Meguroku, Tokyo 153-8505, Japan

(Received 13 July 2008; revised 17 November 2008; accepted 2 December 2008)

The experimental evaluation of the wave number and characteristic impedance of stacked-screen regenerators is described. First, a two-by-two transfer matrix of a stacked-screen regenerator was estimated from pressure measurements performed at four different positions; then, the wave number and characteristic impedance of the regenerator were evaluated using a "capillary-tube-based" theory that models a stacked-screen regenerator as an array of pores having a uniform cross section. The evaluation was applied to seven types of stacked-screen regenerators. The experimental results show that these stacked-screen regenerators can be modeled as arrays of circular-cross-section tubes. Moreover, an empirical equation used to estimate the radius of the circular cross section of the tubes comprising the modeled stacked-screen regenerators was addressed.
 [DOI: 10.1121/1.3056552]

PACS number(s): 43.35.Ud, 43.20.Mv [RR] Pages: 780–786

I. INTRODUCTION

The thermal interaction between solid walls and oscillating gas causes a rich variety of thermoacoustic phenomena. By harnessing thermoacoustic phenomena, one can construct thermoacoustic engines and coolers.[1] Thermoacoustic engines and coolers have high reliability because they have no or a few moving parts. Moreover, recently developed thermoacoustic engines and coolers employ a thermodynamic cycle similar to the Stirling cycle; therefore, they have the potential to achieve high efficiency comparable to that of conventional Stirling engines and coolers.[2–5] Due to their high reliability and potential, thermoacoustic devices have attracted much attention.

In thermoacoustic devices based on the Stirling cycle, regenerators comprising stacked screens are usually adopted as an energy conversion component. Hence, in order to design such devices, one must know the characteristics of stacked-screen regenerators. These regenerators can be acoustically characterized by the wave number and the characteristic impedance in them.[6–8] Although the wave number and the characteristic impedance for simple geometries such as arrays of circular- or square-cross-section tubes can be analytically calculated, it is difficult to analytically calculate the wave number and characteristic impedance for stacked-screen regenerators. This is because the flow channels in stacked-screen regenerators are very complex.

In this study, we consider seven types of stacked-screen regenerators and experimentally evaluate their wave number and characteristic impedance. The evaluation is based on a capillary-tube-based theory,[9] which models a stacked-screen regenerator as an array of tubes having a simple geometrical cross section. The experimental results show that a stacked-screen regenerator can be modeled as an array of circular-cross-section tubes and that the wave number and characteristic impedance of a stacked-screen regenerator can be estimated by using the theoretical results of these tubes. Further, a method is discussed that can be used to estimate the radius of the circular cross section of the tubes constituting the model of a stacked-screen regenerator.

In the next section, the theory for the experimental evaluation of the wave number and characteristic impedance is described. In Sec. III, the experimental setup and procedure are explained. In Sec. IV, the preliminary measurements that demonstrate the validity of our evaluation method are shown; then, the experiments performed on stacked-screen regenerators are presented. In Sec. V, the experimental results are discussed. The study is summarized in Sec. VI.

II. THEORY

A. Wave number and characteristic impedance

With Rott's acoustic approximation,[1,10] the momentum and continuity equations for a tube can be written as

$$\frac{dP}{dx} = -\frac{1}{S}\frac{i\omega\rho_m}{1-\chi_\nu}U, \tag{1}$$

$$\frac{dU}{dx} = -\frac{i\omega S[1+(\gamma-1)\chi_\alpha]}{\gamma P_m}P + \frac{\chi_\alpha-\chi_\nu}{(1-\chi_\nu)(1-\sigma)}\frac{1}{T_m}\frac{dT_m}{dx}U, \tag{2}$$

where P is the oscillatory pressure, U is the volume velocity, ω is the angular frequency of pressure oscillations, and S is the cross-sectional area of the tube. ρ_m, P_m, T_m, γ, and σ are

a)Electronic mail: uedayuki@cc.tuat.ac.jp

the mean density, the mean pressure, the mean temperature, the ratio of specific heats, and the Prandtl number of the working gas, respectively. χ_α and χ_ν are the thermoacoustic functions[10,11] that allow us to describe the three-dimensional phenomena in the acoustical channel using the two one-dimensional equations.

It is possible to analytically determine the thermoacoustic functions in a tube having a simple geometrical cross section such as circular and square. In order to express the thermoacoustic functions, we use two parameters: thermal relaxation time τ_α and viscous relaxation time τ_ν.[11] They are defined as

$$\tau_\alpha = r^2/(2\alpha), \tag{3a}$$

$$\tau_\nu = r^2/(2\nu), \tag{3b}$$

where r is the characteristic length in a tube, α is the thermal diffusivity of the working gas, and ν is its kinematic viscosity. For the case of a circular-cross-section tube, r is the radius of its cross section, and for the case of a square-cross-section tube, r is half the side length of its cross section. For the case of a circular-cross-section tube, the thermoacoustic functions χ_α and χ_ν are expressed as[1,10]

$$\chi_\alpha = \frac{2J_1(Y_\alpha)}{Y_\alpha J_0(Y_\alpha)}, \tag{4a}$$

$$\chi_\nu = \frac{2J_1(Y_\nu)}{Y_\nu J_0(Y_\nu)}, \tag{4b}$$

where

$$Y_\alpha = (i-1)\sqrt{\omega\tau_\alpha}, \tag{5a}$$

$$Y_\nu = (i-1)\sqrt{\omega\tau_\nu}. \tag{5b}$$

For the case of a square-cross-section tube,[1,12]

$$\chi_\alpha = 1 - \frac{64}{\pi^4}\sum_{m,n\ \mathrm{odd}}\frac{1}{m^2n^2C_{\alpha,mn}}, \tag{6a}$$

$$\chi_\nu = 1 - \frac{64}{\pi^4}\sum_{m,n\ \mathrm{odd}}\frac{1}{m^2n^2C_{\nu,mn}}, \tag{6b}$$

where

$$C_{\alpha,mn} = 1 - i\frac{\pi^2}{8\omega\tau_\alpha}(m^2+n^2), \tag{7a}$$

$$C_{\nu,mn} = 1 - i\frac{\pi^2}{8\omega\tau_\nu}(m^2+n^2). \tag{7b}$$

Equations (1) and (2) can be solved analytically for a tube with a uniform cross section and with $dT_m/dx=0$, and their solution is expressed by using the wave number k and the characteristic impedance Z_c as

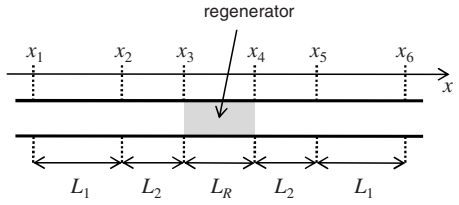


FIG. 1. Coordinate system for the experimental evaluation of the wave number and characterstic impedance for a stacked-screen regenerator.

$$\begin{pmatrix} P_1 \\ U_1 \end{pmatrix} = M(x_1,x_0)\begin{pmatrix} P_0 \\ U_0 \end{pmatrix},$$

$$M(x_1,x_0) \equiv \begin{pmatrix} \cos(k(x_1-x_0)) & \dfrac{Z_c}{iS}\sin(k(x_1-x_0)) \\ \dfrac{S}{iZ_c}\sin(k(x_1-x_0)) & \cos(k(x_1-x_0)) \end{pmatrix}, \tag{8}$$

where P_j and U_j represent the oscillatory pressure and volume velocity at x_j, respectively. The wave number k and the characteristic impedance Z_c are written by using the thermoacoustic functions as

$$k = k_0\sqrt{\frac{1+(\gamma-1)\chi_\alpha}{1-\chi_\nu}} \tag{9}$$

and

$$Z_c = Z_0\frac{k_0}{k(1-\chi_\nu)}, \tag{10}$$

where $k_0=\omega/a$ and $Z_0=\rho_m a$, respectively; a is the adiabatic sound speed.

B. Theory for measurements

In this subsection, we describe the theory used for the experimental evaluation of the wave number k_{exp} and the characteristic impedance $Z_{c\mathrm{exp}}$ for a stacked-screen regenerator. In this theory, the transfer-matrix method[7,8,13] is used, and the capillary-tube-based theory[9] that models a stacked-screen regenerator as an array of pores having a uniform cross section is employed.

We consider the case wherein the regenerator of a length L_R is sandwiched between two circular-cross-section tubes, as shown in Fig. 1. The k and Z_c values in these tubes can be calculated from Eqs. (4), (9), and (10) and are denoted as k_T and $Z_{c,T}$, respectively. The axial coordinate x and the lengths L_1, L_2, and L_R are set as shown in Fig. 1.

By using Eq. (8), the oscillatory pressure and volume velocity at $x=x_1$, (P_1,U_1), can be related to those at x_2, (P_2,U_2),

$$\begin{pmatrix} A_a & B_a \\ C_a & A_a \end{pmatrix}\begin{pmatrix} P_1 \\ U_1 \end{pmatrix} = \begin{pmatrix} P_2 \\ U_2 \end{pmatrix}, \tag{11}$$

where

$$A_a = \cos k_T L_1, \tag{12}$$

$$B_a = \frac{Z_{c,T}}{iS_T} \sin k_T L_1, \tag{13}$$

$$C_a = \frac{S_T}{iZ_{c,T}} \sin k_T L_1. \tag{14}$$

Here, $L_1 = x_2 - x_1 (= x_6 - x_5)$ and S_T is the cross-sectional area of the tubes. The pressure P_2 and volume velocity U_2 at $x = x_2$ can be related to the pressure P_3 and volume velocity U_3 at one end of the regenerator, $x = x_3$,

$$\begin{pmatrix} A_b & B_b \\ C_b & A_b \end{pmatrix} \begin{pmatrix} P_2 \\ U_2 \end{pmatrix} = \begin{pmatrix} P_3 \\ U_3 \end{pmatrix}, \tag{15}$$

where

$$A_b = \cos k_T L_2, \tag{16}$$

$$B_a = \frac{Z_{c,T}}{iS_T} \sin k_T L_2, \tag{17}$$

$$C_a = \frac{S_T}{iZ_{c,T}} \sin k_T L_2. \tag{18}$$

Here, $L_2 = x_3 - x_2 (= x_5 - x_4)$. Equation (11) yields

$$U_1 = (P_2 - A_a P_1)/B_a, \tag{19}$$

$$U_2 = C_a P_1 + A_a U_1. \tag{20}$$

By using Eqs. (15), (19), and (20), we obtain

$$P_3 = A_b P_2 + B_b \left(C_a P_1 + \frac{A_a}{B_a} (P_2 - A_a P_1) \right), \tag{21}$$

$$U_3 = C_b P_2 + A_b \left(C_a P_1 + \frac{A_a}{B_a} (P_2 - A_a P_1) \right). \tag{22}$$

Following the above approach, the following two equations are obtained:

$$P_4 = A_b P_5 + B_b \left(C_a P_6 + \frac{A_a}{B_a} (P_5 - A_a P_6) \right), \tag{23}$$

$$U_4 = -C_b P_5 - A_b \left(C_a P_6 + \frac{A_a}{B_a} (P_5 - P_a P_6) \right), \tag{24}$$

where P_4 and U_4 are the pressure and volume velocity at the other end of the regenerator, $x = x_4$.

Although the wave number k_{exp} and the characteristic impedance $Z_{c,\mathrm{exp}}$ for the regenerator are still unknown, the relation between (P_3, U_3) and (P_4, U_4) can be written as

$$\begin{pmatrix} A_R & B_R \\ C_R & A_R \end{pmatrix} \begin{pmatrix} P_3 \\ U_3 \end{pmatrix} = \begin{pmatrix} P_4 \\ U_4 \end{pmatrix}, \tag{25}$$

where

$$A_R = \cos k_{\mathrm{exp}} L_R, \tag{26}$$

$$B_R = \frac{Z_{c,\mathrm{exp}}}{iS_R} \sin k_{\mathrm{exp}} L_R, \tag{27}$$

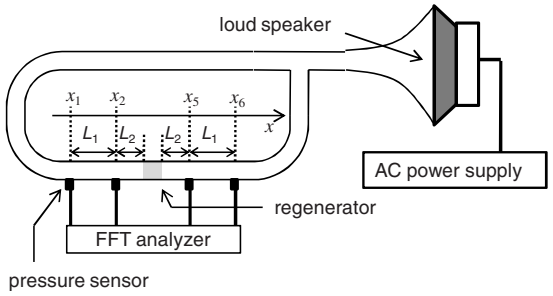


FIG. 2. Experimental setup to evaluate the wave number and characteristic impedance for a stacked-screen regenerator.

$$C_R = \frac{S_R}{iZ_{c,\mathrm{exp}}} \sin k_{\mathrm{exp}} L_R, \tag{28}$$

and S_R is the cross-sectional area of the regenerator. This is because the regenerator is modeled as an array of pores having a uniform cross section. Since the determinant of the matrix in Eq. (25) is unity, i.e., $A_R^2 - B_R C_R = 1$,

$$\begin{pmatrix} A_R & -B_R \\ -C_R & A_R \end{pmatrix} \begin{pmatrix} P_4 \\ U_4 \end{pmatrix} = \begin{pmatrix} P_3 \\ U_3 \end{pmatrix}. \tag{29}$$

Equations (25) and (29) yield

$$A_R = \frac{P_3 U_3 + P_4 U_4}{P_3 U_4 + P_4 U_3}, \tag{30}$$

$$B_R = \frac{P_4^2 - P_3^2}{P_4 U_3 + P_3 U_4}, \tag{31}$$

$$C_R = \frac{U_4^2 - U_3^2}{P_4 U_3 + P_3 U_4}. \tag{32}$$

From Eqs. (26) and (30), we obtain

$$k_{\mathrm{exp}} = \left(\arccos \left(\frac{P_3 U_3 + P_4 U_4}{P_3 U_4 + P_4 U_3} \right) \right) \Big/ L_R, \tag{33}$$

and from Eqs. (27), (28), (31), and (32), we obtain

$$Z_{c,\mathrm{exp}} = S_R \sqrt{\frac{P_4^2 - P_3^2}{U_4^2 - U_3^2}}. \tag{34}$$

Equations (33) and (34) indicate the following: when a stacked-screen regenerator is sandwiched by the circular tubes for which Z_c and k are analytically calculated, the pressure measurements at four positions allow us to evaluate the wave number and characteristic impedance for the regenerator. This is because (P_3, U_3) and (P_4, U_4) in Eq. (33) can be expressed by the matrix elements $(A_a, B_a, C_a, A_b, B_b, C_b)$ and the four values of pressure (P_1, P_2, P_5, P_6), as can be seen in Eqs. (21)–(24).

III. EXPERIMENTAL SETUP AND PROCEDURE

The constructed experimental setup is shown in Fig. 2. It was composed of a loudspeaker, branching resonator, looped tube, and regenerator. The looped tube was used to make the phase difference between measured pressures large enough to be measured. A regenerator having a length of L_R

=20 mm was inserted into the looped tube. The lengths of the looped tube and branching resonator were 2.5 m and between 0.5 and 1.0 m, respectively, and their inner diameter was 24 mm. The tube and resonator were filled with atmospheric air.

We constructed the stacked-screen regenerators by randomly stacking stainless-steel wire mesh screens. The diameter of the screen wire is denoted as d. Half the hydraulic diameter $(D_h/2)$ is employed as the characteristic radius r of a stacked-screen regenerator. D_h is defined as the value that is four times the ratio of the gas volume V_{gas} to the gas-solid contact surface area $S_{g\text{-}s}$ in a stacked-screen regenerator, i.e.,

$$D_h = \frac{4V_{\mathrm{gas}}}{S_{g\text{-}s}}. \tag{35}$$

V_{gas} is calculated from the equation $V_{\mathrm{gas}}=V_{\mathrm{holder}}-V_{\mathrm{solid}}$, where V_{holder} denotes the volume of the regenerator holder and V_{solid} denotes the volume of the wires of a stacked-screen regenerator. $S_{g\text{-}s}$ is estimated from the surface area of the wires and the inner-side surface area of the regenerator holder. For the evaluation of $Z_{c,\mathrm{exp}}$, the cross-sectional area S_R of a stacked-screen regenerator is required [see Eq. (34)]. However, in a stacked-screen regenerator, flow channels are not uniform along the axial direction, and hence, S_R is also nonuniform. Hence, instead of S_R, we use the porosity ϵ of a regenerator for the evaluation; ϵ is defined as $\epsilon=V_{\mathrm{gas}}/V_{\mathrm{holder}}$.

Four pressure sensors (Toyodakoki DD-102 whose resonant frequency is 5 kHz) were mounted on the wall of the looped tube. The distance between the positions of the mounted pressure sensors, L_1 (see Fig. 2), was set to 0.30 m, and the distance of the mounted pressure sensor from the regenerator, L_2 (see Fig. 2), was 0.10 m. We used a 24 bit fast Fourier transform analyzer (Ono sokki DS-2000 whose maximum sampling frequency is 102.4 kHz) to analyze the signal received from the pressure sensors. When oscillatory pressure was measured with the four pressure sensors located at the same axial position and the measured signals were input to the fast Fourier transform analyzer, the obtained phase and amplitude differences between the measured pressure signals were found to be smaller than 0.1° and 0.5%, respectively.

Alternating-current power was continuously supplied to the loudspeaker, and the pressure oscillation was measured at four positions x_1, x_2, x_5, and x_6 (see Fig. 2). By substituting the measured pressures P_1, P_2, P_5, and P_6 into Eqs. (21)–(24), the pressure and volume velocity at both ends of the regenerator were calculated. By substituting the calculated pressure and volume velocity into Eqs. (33) and (34), we evaluated the wave number k_{exp} and characteristic impedance $Z_{c,\mathrm{exp}}$. The regenerator was air cooled so that the temperature gradient along the regenerator was not caused by the thermoacoustic effect.[4] The measurements were made under the condition that the pressure amplitude is sufficiently small to avoid nonlinear effects.

TABLE I. Geometrical properties of the ceramic honeycombs. D_h denotes hydraulic diameter.

Type	SA	SB	SC	SD
$D_h=2r$ (mm)	0.66	0.75	0.90	1.36
Porosity ϵ (%)	87	85	83	69

IV. EXPERIMENTAL RESULTS

A. Preliminary measurements

Preliminary measurements were performed to check our implementation of the present evaluation method of the wave number and characteristic impedance. In these measurements, ceramic honeycombs having many square-cross-section pores were used as regenerators.

We used four types of ceramic honeycombs. Their properties are listed in Table I. The values of D_h were calculated from Eq. (35). Since the ceramic honeycombs can be regarded as arrays of tubes having a uniform square cross section, the wave number k and the characteristic impedance Z_c can be analytically obtained from Eqs. (6), (9), and (10): The theoretically obtained wave number and the characteristic impedance are denoted as k_{theo} and $Z_{c,\mathrm{theo}}$, respectively.

The measurements were made in a frequency range from 40 to 490 Hz and were repeated four times at a given frequency for each ceramic honeycomb. Figures 3 and 4 show the experimentally obtained wave number k_{exp} and characteristic impedance $Z_{c,\mathrm{exp}}$, respectively; the symbols and error bars indicate the mean value and standard deviation of the four measurements, respectively. Note that k_{exp} and $Z_{c,\mathrm{exp}}$ are divided by $k_0=\omega/a$ and $Z_0=\rho_m a$, respectively, where a is the adiabatic sound speed. The theoretically obtained dimensionless wave number k_{theo}/k_0 and characteristic impedance $Z_{c,\mathrm{theo}}/Z_0$ are also shown by dotted lines in Figs. 3 and 4, respectively.

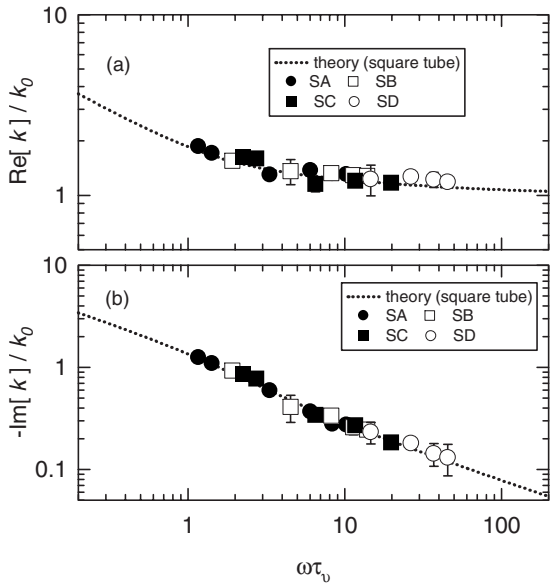


FIG. 3. The experimental results of the dimensionless wave number k/k_0 of the ceramic honeycombs. The experimental results are shown by symbols, and the theoretical results for a square-cross-section tube are shown by dotted lines.

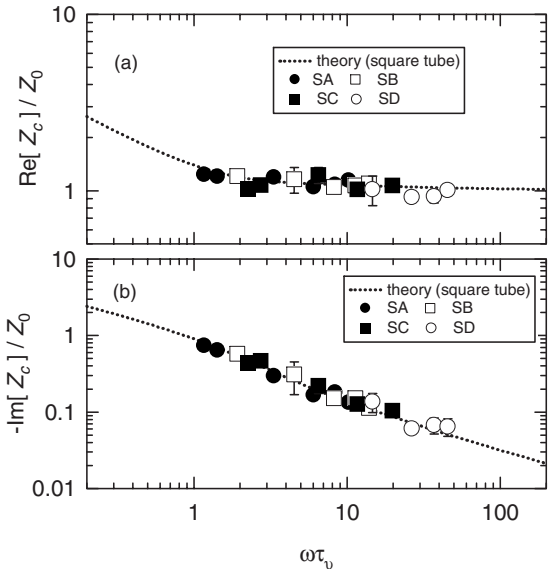


FIG. 4. The experimental results of the dimensionless characteristic impedance Z_c/Z_0 of the ceramic honeycombs. The experimental results are shown by symbols, and the theoretical results for a square-cross-section tube are shown by dotted lines.

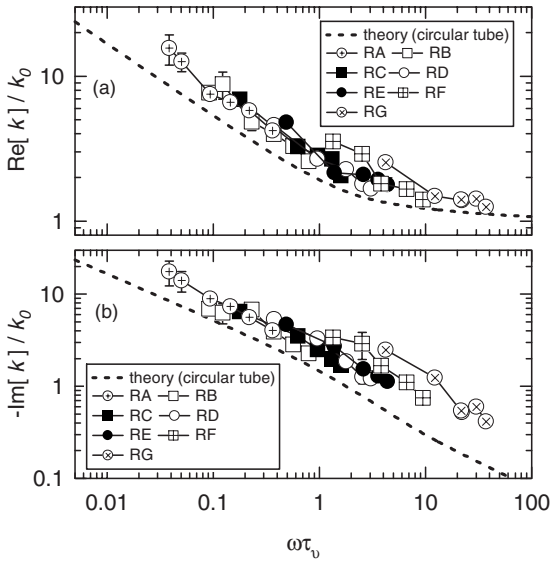


FIG. 5. The experimental results of the wave number k for the stacked-screen regenerators. The dimensionless wave number k/k_0 is shown as a function of $\omega\tau_\nu$.

As can be seen in Fig. 3, the values of the experimentally obtained wave number agree with the theoretical values; the maximum discrepancies between the theoretical and measured mean values of $\mathrm{Re}[k/k_0]$ and those of $\mathrm{Im}[k/k_0]$ are 11% and 15% of the theoretical values, respectively. Figure 4 shows that $Z_{c,\mathrm{exp}}$ agrees with the analytically obtained thermoacoustic function $Z_{c,\mathrm{theo}}$ within a discrepancy of 25% of the theoretical values. On the basis of these results, we consider that the present method can be used for evaluating the wave number k and characteristic impedance Z_c in a stack-screen regenerator.

B. Experimental results with staked screen regenerators

For a stacked-screen regenerator, k_{exp} and Z_c were experimentally evaluated using the method described in Sec. IV A. Seven types of stacked-screen regenerators were used. The geometrical properties of the regenerators are listed in Table II. Note that f_1^2 in this table will be described later.

The experimentally obtained k_{exp}/k_0 of the stacked-screen regenerators is shown as a function of $\omega\tau_\nu$ in Fig. 5. The theoretically obtained wave number k_{theo}/k_0 of a circular-cross-section tube is also shown as a reference. As shown in Fig. 5, $\mathrm{Re}[k_{\mathrm{exp}}/k_0]$ and $-\mathrm{Im}[k_{\mathrm{exp}}/k_0]$ are larger than $\mathrm{Re}[k_{\mathrm{theo}}/k_0]$ and $-\mathrm{Im}[k_{\mathrm{theo}}/k_0]$, respectively, and at a given value of $\omega\tau_\nu$, the values of the real and imaginary parts of k_{exp}/k_0 depend on the type of regenerator. In other words, the wave number of a stacked-screen regenerator depends not only on $\omega\tau_\nu$ but also on the type of regenerator. However, the plots of the real and imaginary parts of k_{exp}/k_0 of each stacked-screen regenerator appear to be parallel to those of k_{theo}/k_0. This implies that by using fitting factors that shift the $\omega\tau_\nu$ versus k_{exp}/k_0 curves toward the $\omega\tau_\nu$ versus k_{theo}/k_0 curves in Fig. 5, the stacked-screen regenerators can be modeled as arrays of circular-cross-section tubes.

The two fitting factors f_1 and f_2 can be used to fit the $\omega\tau_\nu$ versus k_{exp}/k_0 curves to the $\omega\tau_\nu$ versus k_{theo}/k_0 curves.[9,14–17] f_1 and f_2 shift the $\omega\tau_\nu$ versus k_{exp}/k_0 curves toward the left and down in Fig. 5, respectively. However, we introduce only f_1 because we have found that the goodness of fit between the shifted $\omega\tau_\nu$ versus k_{exp}/k_0 curves and the $\omega\tau_\nu$ versus k_{theo}/k_0 curves for each stacked-screen regenerator obtained using only f_1 is almost the same as that obtained using both f_1 and f_2: The goodness of fit for each stacked-screen regenerator was evaluated from the minimum value of the quantity,

TABLE II. Geometrical properties of the stacked-screen regenerators. D_h and f_1^2 denote hydraulic diameter and square of the fitting factor, respectively.

Type	RA	RB	RC	RD	RE	RF	RG
Mesh No.	200	100	80	60		40	24
Wire diameter d (mm)	0.04	0.10	0.10	0.12		0.14	0.22
Porosity (%)	76	66	73	75	78	83	86
$D_h=2r$ (mm)	0.13	0.19	0.26	0.36	0.43	0.65	1.23
f_1^2	2.6	2.1	2.4	2.5	3.2	4.8	6.8

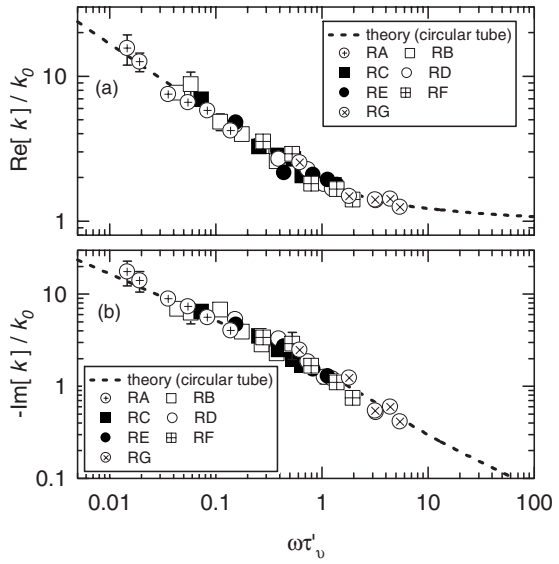


FIG. 6. The experimental results of the wave number k for the stacked-screen regenerators. The dimensionless wave number k/k_0 is shown as a function of $\omega\tau'_\nu=\omega\tau_\nu/f_1^2$.

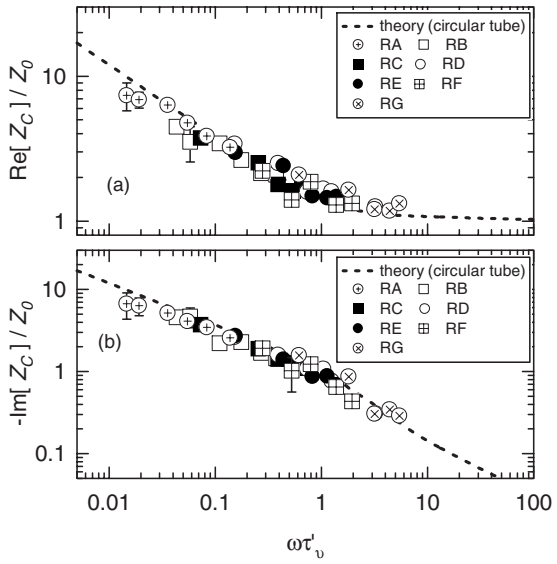


FIG. 7. The experimental results of the characteristic impedance Z for the stacked-screen regenerators. The dimensionless characteristic impedance Z/Z_0 is shown as a function of $\omega\tau'_\nu=\omega\tau_\nu/f_1^2$.

$$F(f_1,f_2)=\sum_{i=1}^{N}\left(\log_{10}\frac{\mathrm{Re}[k_{i,\mathrm{exp}}(\omega\tau'_\nu)]}{f_2\,\mathrm{Re}[k_{i,\mathrm{theo}}(\omega\tau_\nu)]}\right)^2+\sum_{i=1}^{N}\left(\log_{10}\frac{\mathrm{Im}[k_{i,\mathrm{exp}}(\omega\tau'_\nu)]}{f_2\,\mathrm{Im}[k_{i,\mathrm{theo}}(\omega\tau_\nu)]}\right)^2, \tag{36}$$

where

$$\tau'_\nu=\frac{\tau_\nu}{f_1^2} \tag{37}$$

and N is the number of measured data for each stacked-screen regenerator. This would be attributed to the fact that the present measurements were made in the low-$\omega\tau_\nu$ region where both $\mathrm{Re}[k/k_0]$ and $\mathrm{Im}[k/k_0]$ largely depend on $\omega\tau_\nu$, as shown by the dashed lines in Fig. 5.

For each type of the regenerators listed in Table II, f_1^2 was determined so that the value of the goodness of fit $F(f_1)$ is as small as possible. The obtained values of f_1^2 are shown in Table II, and k_{exp}/k_0 and k_{theo}/k_0 are shown as functions of $\omega\tau'_\nu$ in Fig. 6. Note that for the case of k_{theo}, f_1^2 is set to unity, i.e., $\tau'_\nu=\tau_\nu$. Figure 6 shows that the real and imaginary parts of the experimentally obtained wave number k_{exp}/k_0 agree well with those of k_{theo}/k_0. The experimentally determined characteristic impedance $Z_{c,\mathrm{exp}}$ is plotted as a function of $\omega\tau'_\nu$ in Fig. 7. As shown in this figure, the real and imaginary parts of $Z_{c,\mathrm{exp}}/Z_0$ are in acceptable agreement with those of $Z_{c,\mathrm{theo}}/Z_0$; a discrepancy between the measured and theoretical values is within 35% of the theoretical values. From these results, we conclude that stacked-screen regenerators having complex flow channels can be modeled as an array of circular-cross-section tubes by using the obtained values of the fitting factor f_1.

V. DISCUSSION

Table II shows that the fitting factor f_1 largely depends on the type of the regenerator. This indicates that the effective circular radius of the stacked-screen regenerator depends not only on $D_h/2$ but also on its type; the effective circular radius can be defined as

$$r_{\mathrm{eff}}=\sqrt{2\nu\tau'_\nu}=\frac{D_h/2}{f_1}. \tag{38}$$

In this section, the method to estimate the value of f_1 and that of r_{eff} with the characteristics of the stacked-screen regenerators is discussed.

To calculate τ_ν, we employed half the hydraulic diameter $D_h/2$ as the characteristic length r of a stacked-screen regenerator. This is because for the theoretical case of a circular-cross-section tube, the radius of its cross section is used as the characteristic length r and the radius; i.e., r is equal to $D_h/2$. However, for a stacked-screen regenerator, in addition to $D_h/2$, half the wire diameter and half the spacing between the wires that constitute its screens can be regarded as its characteristic lengths; half the wire diameter is approximately equal to half the wire spacing along the axial direction. Hence, we attempt to express f_1 and r_{eff} by using these characteristic lengths.

We focus on the values of the geometric average of 2 or 3 out of $D_h/2$, $d/2$, and $D_s/2$, where d is the wire diameter and D_s is the spacing between wires. D_s can be written as

$$D_s=\frac{25.4\times10^{-3}}{\text{mesh number}}-d. \tag{39}$$

We defined

$$r_1=\frac{\sqrt{D_h d}}{2}, \tag{40}$$

$$r_2=\frac{\sqrt{D_h D_s}}{2}, \tag{41}$$

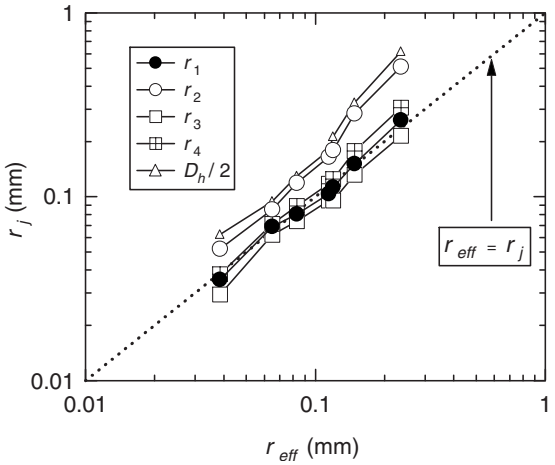


FIG. 8. The relation between r_j and r_{eff} $(j=1,2,3,4)$.

$$r_3 = \frac{\sqrt{D_s d}}{2}, \tag{42}$$

$$r_4 = \frac{\sqrt[3]{D_h D_s d}}{2} \tag{43}$$

and calculated r_j $(j=1,2,3,4)$.

In Fig. 8, r_j is plotted as a function of the experimentally evaluated value of r_{eff} the $r_j=r_{\mathrm{eff}}$ line and D_h versus r_{eff} plot are also shown. As can be seen in this figure, the calculated values of r_1 are closest to the experimentally evaluated values of r_{eff}, and it was found that the difference between r_1 and r_{eff} is smaller than 0.10 of r_{eff}. This indicates that the effective circular radius r_{eff} is approximately expressed by r_1, i.e.,

$$r_{\mathrm{eff}} \approx \frac{\sqrt{D_h d}}{2}, \tag{44}$$

and therefore, the fitting factor f_1 of the stacked-screen regenerators is approximately denoted as

$$f_1 = \frac{D_h/2}{r_{\mathrm{eff}}} \approx \frac{D_h/2}{r_1} = \sqrt{\frac{D_h}{d}}. \tag{45}$$

In other words, the important characteristic length of stacked-screen regenerators is given by the value of the geometric average of its half the hydraulic diameter and half the wire diameter, and the stacked-screen regenerator is approximately modeled as an array of circular-cross-section tubes with the radius of cross section equal to the above characteristic length.

VI. SUMMARY

We have shown the experimental evaluations of the wave number and characteristic impedance for ceramic honeycombs and stacked-screen regenerators. The experimental results of the ceramic honeycombs demonstrated the validity of the present evaluation method. The results of the stacked-screen regenerators indicated that a stacked-screen regenerator can be modeled as an array of circular-cross-section tubes by using one fitting factor f_1, which depends on the type of the regenerator. Further, it was demonstrated that the value of f_1 is approximately estimated by using the hydraulic diameter and the wire diameter of the regenerators.

ACKNOWLEDGMENTS

This research was partially supported by the Ministry of Education, Science, Sports, and Culture in Japan under the Grant-in-Aid for Scientific Research (19860030, 2008) and the Grant-in-Aid for Division of Young Researchers. The authors would like to thank the reviewers for very useful comments and suggestions.

[1]G. W. Swift, *Thermoacoustics: A Unifying Perspective for Some Engines and Refrigerators* (Acoustical Society of America, Melville, NY, 2002).
[2]P. Ceperley, "A piston-less Stirling engine," J. Acoust. Soc. Am. **65**, 1508–1513 (1979).
[3]S. Backhaus and G. W. Swift, "A thermoacoustic Stirling engine," Nature (London) **399**, 335–338 (1999).
[4]G. W. Swift, D. L. Gardner, and S. Backhaus, "Acoustic recovery of lost power in pulse tube refrigerators," J. Acoust. Soc. Am. **105**, 711–724 (1999).
[5]M. Tijani and S. Spoelstra, "Study of a coaxial thermoacoustic-Stirling cooler," Cryogenics **48**, 77–82 (2008).
[6]K. Attenborough, "Acoustical characteristics of rigid fibrous absorbents and granular materials," J. Acoust. Soc. Am. **73**, 785–799 (1983).
[7]R. Muehleisen, C. W. Beamer, IV, and B. Tinianov, "Measurements and empirical model of the acoustic properties of reticulated vitreous carbon," J. Acoust. Soc. Am. **117**, 536–544 (2005).
[8]B. H. Song and J. S. Bolton, "A transfer-matrix approach for estimating the characteristic impedance and wave numbers of limp and rigid porous materials," J. Acoust. Soc. Am. **107**, 1131–1152 (2000).
[9]K. Attenborough, "Acoustical characteristics of porous materials," Phys. Rep. **82**, 179–227 (1982).
[10]N. Rott, "Damped and thermally driven acoustic oscillations," Z. Angew. Math. Phys. **20**, 230–243 (1969).
[11]A. Tominaga, "Thermodynamic aspect of thermoacoustic phenomena," Cryogenics **35**, 427–440 (1995).
[12]W. Arnott, H. Bass, and R. Raspet, "General formulation of thermoacoustics for stacks having arbitrarily shaped pore cross sections," J. Acoust. Soc. Am. **90**, 3228–3237 (1991).
[13]Y. Ueda and C. Kato, "Stability analysis for spontaneous gas oscillations thermally induced in straight and looped tubes," J. Acoust. Soc. Am. **124**, 851–858 (2008).
[14]L. A. Wilen, "Measurements of thermoacoustic functions for single pores," J. Acoust. Soc. Am. **103**, 1406–1412 (1998).
[15]A. Petculescu and L. A. Wilen, "Lumped-element technique for the measurement of complex density," J. Acoust. Soc. Am. **110**, 1950–1957 (2001).
[16]W. Arnott, J. Belcher, R. Raspet, and H. Bass, "Stability analysis of a helium-filled thermoacoustic engine," J. Acoust. Soc. Am. **96**, 370–375 (1994).
[17]H. Roh, R. Raspet, and H. Bass, "Parallel capillary-tube-based extension of thermoacoustic theory for random porous media," J. Acoust. Soc. Am. **121**, 1413–1422 (2007).

Helmholtz-like resonators for thermoacoustic prime movers

Bonnie J. Andersen and Orest G. Symko
Department of Physics, University of Utah, 115 South 1400 East, Salt Lake City, Utah 84112-0830

(Received 16 April 2008; revised 6 November 2008; accepted 16 November 2008)

In a thermoacoustic prime mover, high acoustic output power can be achieved with a large-diameter stack and with a cavity with a large volume attached at the open end of the resonator containing the stack. The combination of resonator and cavity makes the device Helmholtz-like, with special characteristics of the resonant frequencies and quality factor, Q. Analysis of its acoustic behavior based on a model of a closed bottle presents features that are useful for the development of such prime movers for energy conversion from heat to sound. In particular, the arrangement produces in the cavity a high sound level, which is determined by the Q of the system. Comparison with a half-wave resonator type of prime mover, closed at both ends, shows the advantages of the Helmholtz-like device. © *2009 Acoustical Society of America.* [DOI: 10.1121/1.3050263]

PACS number(s): 43.35.Ud, 43.25.Gf, 43.40.At [RR] Pages: 787–792

I. INTRODUCTION

An interesting and promising application of thermoacoustic prime movers is in the conversion of heat, particularly waste heat, to electricity. The development of such technology is based on the availability of new materials, the extension of the operating frequency range for prime movers, and the quest for renewable sources of energy. A thermoacoustic prime mover is an ideal device for converting waste heat to electricity since it has no moving parts, it is simple, and it is capable of a large power density. Studies of thermoacoustic devices have shown that their performance can be optimized by a careful design of geometric factors and operating frequency and with an appropriate match of the various components. This paper deals with the development of resonators, which can provide a good match to sound-to-electricity converters, such as piezoelectric devices.

A simple approach to achieve high acoustic power output from a thermoacoustic device[1] is to use a large-diameter stack located in a correspondingly wide resonator. The generated acoustic power intensity varies directly as the cross-sectional area, A, of the stack perpendicular to the sound flow. For a resonator, closed at one end and open at the other end, i.e., a quarter-wave resonator, such an approach affects the standing wave in the resonator, decreasing the standing wave ratio (SWR). A consequence of this is an increase in the temperature gradient for an onset of acoustic oscillations in the device, becoming so large with an increase in A that the device has difficulties in initiating oscillations. The solution to this problem consists of attaching a cavity to the open end of the prime mover resonator; this will reflect sound waves back into the resonator, providing positive feedback and thus lowering the threshold for acoustic oscillations. Moreover such an approach provides a place for locating the piezoelectric element at the end of the cavity; it also provides the option for pressurizing this unit while maintaining the quarter-wavelength resonator configuration.[2] A study of a resonator-cavity combination is presented here. It will be referred to as a "Helmholtz-like" resonator.

II. HELMHOLTZ-LIKE RESONATOR

By attaching a cavity to the quarter-wavelength prime mover resonator, the system is closed at each end; the cavity provides positive feedback. It is quite different from a half-wave resonator as its quality factor, Q, is higher when the cavity radius, R, is larger than the resonator radius, R_0.

Figure 1 shows the basic geometry of a Helmholtz-like resonator. The resonator-cavity system shown above is more complex than a simple quarter-wave resonator or even a Helmholtz resonator, which was used to characterize Sondhauss tubes.[3] A method developed by Crawford[4] to study the modes of a bottle is applied to the thermoacoustic prime movers of this study. A Helmholtz-like model is developed for characterizing the resonant frequency and the acoustic pressure profile within the system. Its performance is then compared to a half-wave resonator prime mover, closed at both ends.

In a resonator viscous losses play a major role in its performance. They occur over the entire inner surface area within the viscous penetration depth, δ_v. To reduce such losses, Hofler[2] developed a thermoacoustic refrigerator, which consisted of a quarter-wavelength resonator with a sphere attached to the open end of the resonator. One purpose for having the sphere is that it allowed the working gas to be pressurized or substituted by a different gas. In the case presented here, the quarter-wavelength resonator has a large diameter for greater sound power output, and this leads to a reduced SWR. A resonator's SWR is roughly proportional to the ratio of the length, L, and the radius, R, of the resonator. The air effectively shorts out the prime mover to the extent that sound is not generated at all or the threshold for oscillations becomes very large. Loading the resonator with a cavity enhances the overall quality factor, Q, of the system.

It is interesting to compare the Q of the quarter-wavelength resonator and cavity system to that of a half-wave resonator closed at both ends; both resonators have the same radius R. Q is proportional to the mass of air within a system and the resistance of a system.[5] The resistance of a resonator is proportional to its surface area, where thermoviscous losses occur. The mass of air in the resonator is

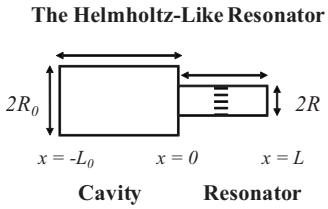


FIG. 1. Prime mover with a Helmholtz-like resonator. The dashed lines within the resonator region represent the stack material placement.

proportional to its volume. Thus Q will be proportional to the volume of air to the surface area of the resonator. The ratio of the Q values of the two resonators is

$$\frac{Q_{\mathrm{qc}}}{Q_{\mathrm{hw}}} = \frac{A_{\mathrm{hw}} V_{\mathrm{qc}}}{A_{\mathrm{qc}} V_{\mathrm{hw}}}, \tag{1}$$

where qc refers to the quarter-wave cavity system and hw refers to the half-wave system. The above expression shows that Q_{qc} will be larger than Q_{hw} when $R_0 > R$, where R_0 is the radius of the cavity, if both have the same total length.

When a tube, open at both ends, is joined to a cavity, the system is a Helmholtz resonator and the resonant frequency, f, obeys the relation[5]

$$f = \frac{c}{2\pi} \sqrt{\frac{S}{L'V}}, \tag{2}$$

where V is the volume of the cavity, L' is the effective length of the tube, S is the cross-sectional area of the tube, and c is the speed of sound. The air in the tube essentially moves in and out of the cavity without being compressed, while the air in the cavity compresses and rarifies.

Lord Rayleigh[3] identified the oscillations of the Sondhauss tube as a mass-spring system. There is another regime between the Helmholtz resonator and the quarter-wavelength resonator with a large cavity. The latter is like the unit in Hofler's[2] refrigerator where the cavity is so large compared to the resonator that most of the wave is contained in the resonator with a high enough SWR.

In the present paper the resonator and cavity are considered as a combined acoustic system. Gaitan and Atchley[6] referred to tubes that have variable cross-sections at different sections of the tube as "anharmonic" resonators. Denardo and Alkov[7] referred to them as having "nonuniformity," and Ilinskii *et al.*[8] referred to them as "dissonant" resonators. In this study, the resonator-cavity systems will be considered as Helmholtz-like resonators, having the same qualities emphasized by the descriptions of the various authors.

With Helmholtz-like systems, the pressure and velocity are necessarily not negligible to allow for positive feedback and appreciable pressure at the closed end of the cavity for energy extraction. This system is similar to a Helmholtz resonator in that a resonator, closed at the outer end and open at the end joining the cavity, is used for the tube and the cavity is used for the volume. The system cannot be a true Helmholtz resonator because the mass of gas within the tube, i.e., the resonator region, must be compressible to generate thermoacoustic work. This system has a new standing wave pattern and frequency, different from the simple quarter-wave resonator. It must take into account the compressibility within the neck region. Applying the modification made to the Helmholtz resonator developed to explain modes of glass bottles, as presented by Crawford,[4] the pressure profile is characterized and the resonant frequency is calculated.

The process of matching boundary conditions is similar to that presented by Panton and Miller,[9] Rott and Zouzoulas,[10] Gaitan and Atchley,[6] and Denardo and Alkov.[7] Here, results are presented for the Helmholtz-like systems with a similar approach to Crawford's[4] results for the closed bottle.

III. MATHEMATICAL APPROACH

Crawford[4] extended the concept of a Helmholtz resonator to approximate acoustic modes within bottles by considering that the flow in both the neck and cavity are compressible. Rather than a mass-spring system, it becomes a series of springs in the cavity and in the tube. This approach does not take into consideration the radial motion of the air but only the axial motion from the tube to the cavity. Crawford[4] approximated a bottle as two cylinders of different radii joined together and calculated the resonance frequency based on the dimensions of the cylinders and the boundary conditions where the two cylinders meet and at the ends of the two cylinders.

If the cavity used for a thermoacoustic engine is also a cylinder as with the resonator, the method used by Crawford[4] is effective for modeling the pressure profile within the cavity, as well as the frequency. Figure 1 shows the two regions considered here.

Following Crawford,[4] the general equation for the displacement, ψ, of air within the system, ψ, as a function of time, t, and x, is a wave equation where the general solution is

$$\psi_+ = \cos(\omega t + \alpha)\{A \cos[k(L-x)] + B \sin[k(L-x)]\} \tag{3}$$

within the resonator and

$$\psi_- = \cos(\omega t + \alpha)\{A_0 \cos[k(x+L_0)] + B_0 \sin[k(x+L_0)]\} \tag{4}$$

within the cavity where α is a phase constant, k is the wave vector defined as $2\pi f/c$, and A, B, A_0, and B_0 are amplitudes.

The boundary conditions for the closed end of both the cavity and the resonator are that ψ_+ and ψ_-, must be zero at the closed ends, $x=-L_0$ and $x=L$. The boundary conditions at $x=0$ are conservation of mass given by

$$\pi R_0^2 \psi_- = \pi R^2 \psi_+, \tag{5}$$

where πR_0^2 and πR^2 are the cross-sectional areas of the cavity and resonator regions, respectively, and continuity of pressure,

$$\frac{\partial \psi_-}{\partial x} = \frac{\partial \psi_+}{\partial x}. \tag{6}$$

These provide the condition on k (and thus the resonant frequency),

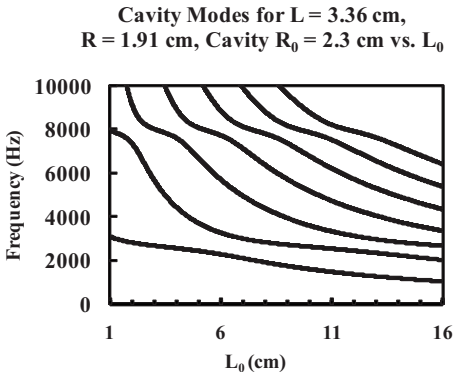


FIG. 2. Solutions of the transcendental equation as a function of the cavity length. The fundamental mode is the lowest curve, with the higher modes following, in ascending order, upward up to the seventh.

$$\cot(kL)\tan(kL_0) = -\frac{R^2}{R_0^2}. \tag{7}$$

The pressures, p_+ and p_-, within the resonator are 90° out of phase with and proportional to the displacement, giving

$$p_+ = C\cos[k(L-x)] \tag{8}$$

and

$$p_- = C_0\cos[k(x+L_0)], \tag{9}$$

where C and C_0 are proportional to B and B_0 by the same factor.

The condition that determines the resonant frequency is a transcendental equation. This has transformed the simple Helmholtz equation [Eq. (3)] to a more complex system, but not as complicated as those given by Rott and Zouzoulas[10] or Gaitan and Atchley[6] because this geometry is simpler. Numeric techniques can be used to determine the solutions. The solutions are the allowed values for k.

Figure 2 shows a typical pattern for the first seven modes of a Helmholtz-like resonator, with the resonant frequency plotted as a function of the cavity length. The ranges and step sizes for k and L_0 must be adjusted properly to find the correct solution for each corresponding mode. Separating the modes is achieved after sweeping through from longer values of the cavity length to shorter values and for each cavity length starting at low values for k and sweeping through to higher values for k. From Fig. 2, it is obvious that the higher modes are not harmonics of the fundamental.

IV. EXPERIMENTS

A. Theoretical and experimental comparison of the Helmholtz-like resonator

The Helmholtz-like method discussed above is used to analyze multiple resonator-cavity configurations with various values of the parameters. Six different cavity lengths ranging from 1.18 to 10 cm were used, all having a radius of 2.3 cm. There are three different resonator diameters compared in this study: 1.91, 1.27, and 1.91 cm. They all are approximately 3.3 cm long, giving aspect ratios of length to diameter of 3.5, 2.6, and 1.7, respectively. These are all significantly smaller than the engines discussed by Swift[11] and Migliori and Swift.[12] The source of sound for the resonators is thermoacoustic driven oscillations, produced internally at the stack. Heat is supplied to each engine via a heating element wrapped around the resonator just above the hot-heat exchanger. The metal cavity acts as a heat sink to the ambient temperature for the cold-heat exchanger.

The heat exchangers and stack material are the same for all systems tested. The stack is a random fibrous material (steel wool) with a volume filling factor of about 4%, giving a hydraulic radius comparable to the thermal penetration depth, and the heat exchangers are 40×40 wires/in. (0.254 mm in diameter) or 60×60 wires/in. (0.19 mm in diameter) copper mesh screens. The stacks are approximately located at the midpoint of the quarter-wave resonator. Air at atmospheric pressure is the working fluid. One additional 1.27-cm-diameter resonator with a length of 2.9 cm is also tested. The stack material and heat exchanger geometry for all devices were the same. The heat was supplied to the devices via the heating element receiving 8–12 W of electrical power. (Most used exactly 9 W, and less or more power was required in a few cases that shall be mentioned.)

B. Helmholtz-like resonator and half-wave resonator comparison

A 1.91-cm-diameter half-wave resonator by itself is compared with the Helmholtz-like system with a 1.91-cm-diameter resonator and a 10 cm cavity. The total length of the half-wave resonator, L_{hw}, is 10.7 cm to operate at the same fundamental frequency as the Helmholtz-like resonator. Both have the same type of heat exchangers and stack material and size. The masses of the top part of each engine above the hot-heat exchanger are equal. Both received an input power of 11.15 W through a heating element wrapped around the resonator above the heat exchanger.

Using Eq. (1), the ratio of the expected Q values for the two resonators is 2.25, neglecting the surface area of the heat exchangers, the stack, and the end of the resonator. This ratio is compared with measured values for the Q of the two resonators. A model 7060 Exact Generator is used to send 100 μs pulses to a piezotweeter, MCM model 53-800, speaker placed near the end of the resonator. A microphone placed within the resonator records the oscillations. The stack and heat exchangers are present for these measurements. An oscilloscope, triggered by the output from the generator, displays the signal from the microphone. By counting the number of sound oscillations, the Q of the resonators can then be approximated. Finally, the acoustic pressures, measured with a Honeywell XCA401DN sensor, are compared.

V. RESULTS

A. Helmholtz-like resonators

Table I displays the theoretical and measured frequencies produced for 6 of the 16 different resonator-cavity combinations tested. The 1.91-cm-diameter resonator with the 1.18 cm cavity only attained oscillations after supplying a

TABLE I. Measured frequencies for five Helmholtz-like resonators compared with corresponding theoretical values (* indicates the second mode).

L (cm)	L_0 (cm)	R_0 (cm)	f_{theory} (kHz)	f_{measured} (kHz)
2.87	1.18	0.635	3.17	3.16
3.29	5.67	0.476	2.50	2.48
3.36	5.67	0.635	2.40	2.37
3.26	5.67	0.953	2.32	2.29
3.29	7.25	0.476	2.24, 2.75*	2.24, 2.75*
3.26	10.0	0.953	1.59, 2.62*	1.61, 2.61*

power greater than 9 W. This is probably because there was not enough positive feedback provided by the short cavity to initiate oscillations at a lower power. With the 10 cm cavity, all systems were measured to be operating in the second mode, with the 1.91-cm-diameter resonator beginning in the first mode and sometimes switching to the second mode as the temperature rose sufficient to excite the second mode. An input power of 8 W was used to keep this system in the first mode. Likewise, the 0.952-cm-diameter resonator was seen to achieve both first and second modes with the 7.25 cm cavity. Kwon[13] also observed a crossover from the first to the second mode and even from the second to the third mode as the cavity length was increased.

The pressure profile within the cavity and resonator is measured from the end of the cavity up to the stack along the axis of the system. It is measured using the pressure sensor. The probe of the sensor is a 1/16-in.-outer diameter and 0.031-in.-inner diameter ceramic tube attached to a nylon tube that runs into the sensor. The pressure is measured every 0.5 cm along the axis of the cylinders from the end of the cavity up to the stack position. Because of the very small size of probe, the pressure field was not affected. There are two o-rings where the ceramic tube enters the base of the cavity to reduce leakage at the interface. Since the presence of any leak around the tube reduces the pressure, a Swagelok connection was attached at the interface. However, there was a tiny void just beyond the end of the cavity where the swedge-lock was joined to the mount. This void distorted the system substantially, and the attachment with o-rings was used instead, yielding more accurate results. The data measured with the pressure sensor is plotted together with the theoretical results. The theoretical function for the pressure within the cavity is taken to be continuous within the system, but it is not assumed to be smooth, following the approach of Sec. III. Recall also that the longitudinal motion only of the gas is considered and not the radial motion. Hence, it is expected that the measured values will deviate from the theory near the interface of the resonator and cavity.

The theoretical equations must be normalized to the measured pressure at some point within the system. Since there is a discrepancy near the interface between the resonator and the cavity, a logical place for normalization is at $x=-L_0$, the end of the cavity. The frequency used to calculate the pressure profile is the measured value of the operating Helmholtz-like resonator prime mover, except for some of the resonators with the 10 cm cavity, which have dimensions

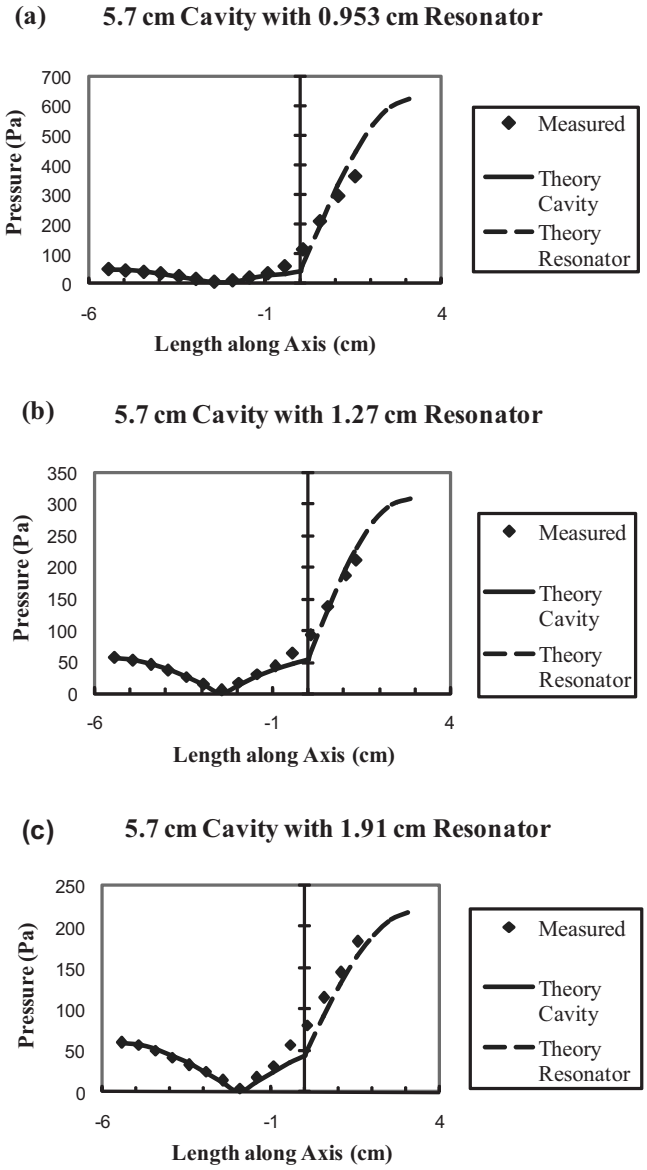


FIG. 3. The theoretical and experimental pressure profiles of the 5.7-cm-long cavity with the following resonators: (a) 0.953 cm diameter, (b) 1.27 cm diameter, and (c) 1.91 cm diameter. The diamond symbols indicate the measured values taken with the pressure sensor. The solid lines indicate the theoretical pressure within the cavity regions, and the dashed lines indicate the theoretical values. The uncertainty is within the marker for the values.

near $L_0=nL$, n odd, which causes C_0 to be skewed using measured values of the frequency. In those cases, the theoretical value rather than the measured value of the frequency, which keeps C_0 reasonable, was used.

Figure 3 displays the positive value of the pressure profile within the system for the 5.7 cm cavity, with the resonators having diameters of 0.953, 1.27, and 1.91 cm. There is one visible node in each case.

Figure 4 shows the first and second modes achieved with the 1.91-cm-diameter resonator and 10-cm-long cavity combination. There are two nodes present in the second mode, as to be expected. The acoustic pressure achieved at the end of the cavity in the first mode is substantially greater than that achieved by the second mode.

The acoustic pressures achieved at the bottom of the 5.67 cm cavity (at $x=-L_0$) are 81, 56, and 122 Pa for the

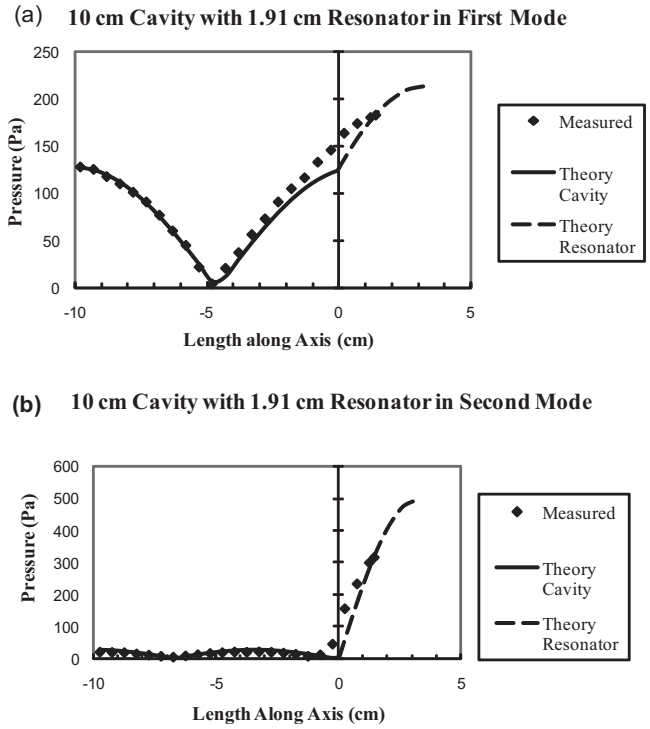


FIG. 4. Pressure profile of the 1.91-cm-diameter resonator with the 10-cm-long cavity in the (a) first and (b) second modes. The diamond symbols indicate the measured values taken with the pressure sensor. The solid lines indicate the theoretical pressure within the cavity regions, and the dashed lines indicate the theoretical values. The uncertainty is within the marker for the values.

resonators having diameters of 0.953, 1.27, and 1.91 cm, respectively. The acoustic pressures for the 7.25 cm cavity are 182, 153, and 215 Pa for the three resonators, respectively. Each engine received an input power of 9 W. The 1.91-cm-diameter resonator achieved the highest pressure of all engines of 285 Pa when attached to the 10 cm cavity when it remained in the first mode. The pressures in the second mode for the 0.953-, 1.27-, and 1.91-cm-diameter resonators were 55, 56, and 65 Pa, respectively. The uncertainty on these measurements is 4 Pa.

B. Half-wave resonator comparison

The measured Q values for the Helmholtz-like resonator and the half-wavelength resonator are 58.1 and 25.1, respectively, giving a ratio of 2.5. The acoustic pressures measured at the peak performance for the Helmholtz-like resonator prime mover and the half-wave resonator are 524 and 220 Pa, respectively. The ratio of pressures is 2.4. The acoustic pressure ratios achieved are closely proportional to the ratio of Q values measured.

VI. CONCLUSIONS

The 0.953-cm-diameter resonator works without a cavity, but the SWR is too low for the 1.27- or the 1.91-cm-diameter resonators to work without some reflections. With a cavity, the 1.91-cm-diameter resonator engine produces the greatest pressure compared with the other two for all cavity lengths. Thus cavities, providing positive feedback, can allow for larger-diameter engines, producing more sound according to Eq. (3) despite poor SWRs. It is unclear why the 1.91-cm-diameter engine with the 10-cm-long cavity would sometimes spontaneously jump from the first to the second mode.

For two combinations of the resonators and cavities, the first and second modes were seen. This makes sense because with larger cavities the Q should also be larger, and as the temperature of the devices increases, there is sufficient energy to excite the second mode. However, for the 0.953- and the 1.27-cm-diameter resonators with the 10 cm cavity, the second mode only was seen. This could be due to thermoviscous dissipation. Q competition between modes can be caused by changes in the drive mechanism.[14]

The method presented accurately describes both the frequency and pressure profiles of Helmholtz-like resonators. While the pressure within the resonator region is roughly equal to the theoretical value, the pressure within the cavity region is an excellent match. The average frequencies measured match the prediction to within 2% for all 16 combinations tested. This approach shall allow for further study to optimize geometric parameters of resonator-cavity systems.

It is shown that a Helmholtz-like resonator achieves a higher acoustic pressure, by a factor of 2.4, than a half-wave resonator, operating at the same frequency. The ratio of pressures observed is close to the ratio of the Q values of the two systems. Thus, Helmholtz-like resonators show promise as resonators for use in thermoacoustic standing wave engines.

Future work could include the effect of stack placement on the acoustic energy produced. It would also be interesting to study how the stack placement affects which mode is excited and the transitions between the modes. Since the presence of the stack and heat exchangers complicated the problem, it would be useful to do a modeling of the system.

ACKNOWLEDGMENTS

This research was supported by the U.S. Army Space Missile Defense Command. The authors would also like to acknowledge Jack Pitts for the machining of parts for the devices in this study.

[1]G. W. Swift, "Thermoacoustic engines and refrigerators," Phys. Today **48**, 22–28 (1995).
[2]T. Hofler, "Thermoacoustic refrigerator design and performance," Ph.D. dissertation, Physics Department, University of California at San Diego, 1986.
[3]Lord Rayleigh (J. W. Strutt), *The Theory of Sound*, 2nd ed. (Dover, New York, 1945), Vol. **2**, Sec. 322.
[4]F. S. Crawford, "Lowest modes of a bottle," Am. J. Phys. **56**, 702–712 (1988).
[5]L. E. Kinsler, A. R. Frey, A. B. Coppens, and J. V. Sanders, *Fundamentals of Acoustics* (Wiley, New York, 1982), 3rd ed., pp. 16 (Eq. 1.37) and 227 (Eq. 10.8).
[6]D. F. Gaitan and A. A. Atchley, "Finite amplitude standing waves in harmonic and anharmonic tubes," J. Acoust. Soc. Am. **93**, 2489–2495 (1993).
[7]B. Denardo and S. Alkov, "Acoustic resonators with variable nonuniformity," Am. J. Phys. **62**, 315–321 (1994).
[8]Y. A. Ilinskii, B. Lipkens, T. S. Lucas, T. W. Van Doren, and E. A. Zabolotskaya, "Nonlinear standing waves in an acoustical resonator," J. Acoust. Soc. Am. **104**, 2664–2674 (1998).
[9]R. L. Panton and J. M. Miller, "Resonant frequencies of cylindrical Helmholtz resonators," J. Acoust. Soc. Am. **57**, 1533–1535 (1975).

[10]N. Rott and G. Zouzoulas, "Thermally driven acoustic oscillations, Part IV: Tubes with variable cross-section," Z. Angew. Math. Phys. **27**, 197–224 (1976).
[11]G. W. Swift, "Analysis and performance of a large thermoacoustic engine," J. Acoust. Soc. Am. **92**, 1551–1563 (1992).
[12]A. Migliori and G. W. Swift, "Liquid-sodium thermoacoustic engine," Appl. Phys. Lett. **53**, 355–357 (1988).
[13]Y. S. Kwon, "Study of thermoacoustic engines operating between 2 kHz and 25 kHz," Ph.D. dissertation, Department of Physics, University of Utah, 2006.
[14]N. H. Fletcher, "Nonlinear interaction in organ flue pipes," J. Acoust. Soc. Am. **56**, 645–652 (1974).

Density imaging using inverse scattering

Roberto J. Lavarello[a] and Michael L. Oelze
Bioacoustics Research Laboratory, Department of Electrical and Computer Engineering, University of Illinois at Urbana-Champaign, 405 North Matthews, Urbana, Illinois 61801

(Received 20 February 2008; revised 6 October 2008; accepted 11 November 2008)

Inverse scattering is considered one of the most robust and accurate ultrasonic tomography methods. Most inverse scattering formulations neglect density changes in order to reconstruct sound speed and acoustic attenuation. Some studies available in literature suggest that density distributions can also be recovered using inverse scattering formulations. Two classes of algorithms have been identified. (1) The separation of sound speed and density contributions from reconstructions using constant density inverse scattering algorithms at multiple frequencies. (2) The inversion of the full wave equation including density changes. In this work, the performance of a representative algorithm for each class has been studied for the reconstruction of circular cylinders: the dual frequency distorted Born iterative method (DF-DBIM) and the *T*-matrix formulation. Root mean square error values lower than 30% were obtained with both algorithms when reconstructing cylinders up to eight wavelengths in diameter with moderate density changes. However, in order to provide accurate reconstructions the DF-DBIM and *T*-matrix method required very high signal-to-noise ratios and significantly large bandwidths, respectively. These limitations are discussed in the context of practical experimental implementations.
 [DOI: 10.1121/1.3050249]

PACS number(s): 43.35.Wa, 43.60.Pt [TDM] Pages: 793–802

I. INTRODUCTION

Ultrasonic computerized tomography (UCT) is an imaging modality used to reconstruct quantitative images of acoustic properties and has been studied since the 1970s. Initial attempts were performed using ray propagation algorithms to form images of attenuation α (Ref. 1) and speed of sound c (Ref. 2). These algorithms can only correct for refraction and their spatial resolution is limited by diffraction effects. Diffraction tomography was developed in the 1980s as a means to compensate for diffraction effects by linearizing the wave equation using either the Born or Rytov approximations, but this approach provides erroneous solutions even for low contrast between the acoustic properties of the background and the scattering object.[3]

Traditionally, inverse scattering algorithms in the frequency domain based on Newton-type approaches such as the distorted Born iterative method (DBIM)[4,5] attempt to solve for the full wave equation assuming no changes in density ρ. Under this assumption, these algorithms allow the reconstruction of c and α. Newton-type inverse scattering algorithms are not limited by diffraction effects, and convergence for large contrast objects can be obtained by properly using multiple frequency data.[6]

However, experimental evidence is available in literature suggesting that relative ρ changes in tissues may be comparable in magnitude to relative c changes.[7,8] The effects of variable density in the reconstruction of sound speed were observed to result in overshoots of sound speed estimates at the edges of objects where density underwent abrupt changes.[9] Sharper changes in density were found to lead to larger artifacts in reconstructed images of sound speed.

Currently, UCT has been proposed for the early detection and diagnosis of breast cancer. Clinical trials indicate that α reconstructions may be more important than c reconstructions for differentiating benign from malignant lesions.[10] Determining density distributions may provide additional information or contrast for cancer detection.

Quantitative ultrasound techniques based on the backscatter (QUBS) may also benefit from the determination of density distributions.[11] QUBS consists of estimating properties of tissue microstructure based on backscattered pressure measurements and scattering models. Under weak scattering assumptions, the backscattered power spectrum can be related to the three-dimensional spatial autocorrelation function of the acoustic impedance, $Z=\rho c$, of the underlying tissue microstructure.[12] Therefore, using variable density UCT in conjunction with c reconstructions at high frequencies could in theory be useful for QUBS by providing three-dimensional impedance maps of tissues.

The number of UCT studies that consider variable density is limited. Variable density UCT was introduced in the context of single scattering formulations using bistatic scanning configurations with infinite bandwidth transducers.[13,14] Density reconstructions were later explored using diffraction tomography methods,[15] with some researchers developing similar approaches with both algebraic[16] and Fourier-based[17] algorithms. However, the fact that these works are based on linearized scattering theory limits their applicability.

Some inverse scattering methods designed to recover density information have also been developed. Two classes of algorithms have been identified: (1) the separation of sound speed and density contributions from reconstructions

[a]Electronic mail: lavarell@illinois.edu

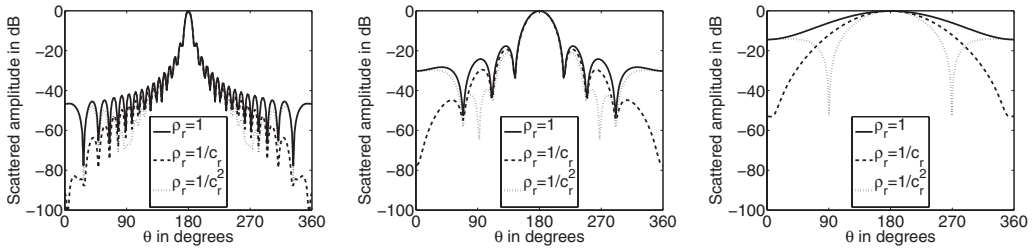


FIG. 1. Effect of ρ_r on the scattered pressure patterns of circular cylinders with $\Delta c=2\%$. All plots show the normalized amplitude of the scattered pressure in decibels. The radii of the cylinders are 4λ (left), λ (center), and λ/4 (right), respectively.

using constant density inverse scattering algorithms at multiple frequencies[18,19] and (2) the inversion of the full wave equation including density changes.[20–22] All of these works claim that UCT can also be used to obtain quantitative images of density distributions.

The goal of the present work is to analyze through simulations the performance of the two classes of variable density inverse scattering algorithms when reconstructing circular cylinders. The effects of scatterer size, density and speed of sound contrast values, and noise are considered. The root mean square error (RMSE) of the reconstructed profiles is used as a quality metric when assessing the accuracy of both approaches. As a result of this work, the fundamental limitations of variable density inverse scattering methods will be better understood and presented in a more comprehensive manner.

The remainder of this paper is organized as follows. Section II reviews the effects of changes in density in the pressure field scattered by a circular cylinder. The performances of the first and the second class of algorithms are analyzed in Sec. III [dual frequency DBIM (DF-DBIM) approach] and Sec. IV (*T*-matrix approach), respectively. Finally, Sec. V discusses the results obtained in this work.

II. EFFECTS OF DENSITY IN THE PRESSURE FIELD SCATTERED BY A CIRCULAR CYLINDER

The analytic solution for the scattering of a cylindrical wave by a circular cylinder is well documented in literature,[23] which allows the generation of exact data for all methods tested in the present work. Consider the case of a cylinder of radius a, density ρ, compressibility κ, speed of sound c, wave number k, and acoustic impedance Z embedded in a homogeneous background. Throughout the present work, for a given acoustical property X, the relative X_r and contrast ΔX values are defined as $X_r=X/X_0$ and $\Delta X=X_r-1$, respectively, where X_0 is the value of the property in the background. The pressure scattered by the cylinder when a line source is located at $x=R$ can be written as

$$p^{sc}(\boldsymbol{r}) = \sum_{m=0}^{\infty} A_m R_m(\kappa,\rho) H_m^{(1)}(k_0 R) H_m^{(1)}(k_0 r)\cos m\theta, \qquad (1)$$

where r and θ are the cylindrical coordinates of the observation point, $A_0=1$ and $A_m=2$, $m>0$, and $H_m^{(1)}(\cdot)$ is the mth order Hankel function of the first kind. The scattering coefficient $R_m(\cdot)$ can be calculated as

$$R_m(\kappa,\rho) = \frac{\frac{1}{Z_r} J_m(k_0 a) J_m'(ka) - J_m(ka) J_m'(k_0 a)}{J_m(ka) H_m'^{(1)}(k_0 a) - \frac{1}{Z_r} J_m'(ka) H_m^{(1)}(k_0 a)}, \qquad (2)$$

where $J_m(\cdot)$ is the mth order Bessel function and the $'$ symbol represents derivative with respect to the total argument. In the Rayleigh limit ($\lambda \gg a$) the scattered pressure in the far field can be approximated as

$$p^{sc}(\boldsymbol{r}) \xrightarrow[\lambda \gg a]{} \frac{k_0 a^2}{2} \frac{e^{ik_0(R+r)}}{\sqrt{Rr}} \left\{ [\kappa_r - 1] - 2\left[\frac{\rho_r - 1}{\rho_r + 1}\right] \cos\theta \right\}. \qquad (3)$$

The first term in the brackets in Eq. (3) represents monopole scattering with dependence on κ and the second term represents dipole scattering with dependence on ρ.

The effects of density variations on the scattering patterns of circular cylinders with three different radii are shown in Fig. 1. The cylinders had radii λ/4, λ, and 4λ with a fixed speed of sound contrast $\Delta c=2\%$. Three cases were evaluated per cylinder size: $\rho_r=1$ (no changes in density), $\rho_r=1/c_r$ (equal changes in compressibility and density), and $\rho_r=1/c_r^2$ (no changes in compressibility). The pressure fields for each case are shown in Fig. 1. The RMSEs between the $\rho_r=1$ and $\rho_r=1/c_r$ cases were 1.86%, 7.77%, and 31.46% for $a=4\lambda$, λ, and λ/4, respectively. Similarly, the RMSEs between the $\rho_r=1$ and $\rho_r=1/c_r^2$ cases were 3.72%, 15.58%, and 62.93%, respectively. These results illustrate the fact that unless $a \ll \lambda$, the scattered field will be fairly insensitive in the mean square sense to changes in density for low contrast scatterers when $\Delta\rho$ is not much larger than Δc.

III. VARIABLE DENSITY AND THE DBIM

The first class of algorithms for density reconstruction is based on the separation of c and ρ contributions from reconstructed profiles obtained using constant density inversion algorithms at two frequencies. This approach is analogous to the one presented in Refs. 15 and 16 for the diffraction tomography case and was proposed in Ref. 18 using the alternating variable algorithm.[24] Preliminary but limited results using Newton-type methods can be found in Ref. 19. The DBIM,[4] one of the most well-studied constant density inverse scattering methods, will be used for the remainder of this section.

A. The distorted Born iterative method

The details of DBIM are presented here for completeness. The wave propagation in an inhomogeneous medium is described by[25]

$$\rho(\boldsymbol{r})\,\nabla\cdot[\rho^{-1}(\boldsymbol{r})\,\nabla p(\boldsymbol{r})]+k^2(\boldsymbol{r})p(\boldsymbol{r})=-\phi^{\mathrm{inc}}(\boldsymbol{r}), \tag{4}$$

where $p(\boldsymbol{r})$ is the acoustical pressure and $\phi^{\mathrm{inc}}(\boldsymbol{r})$ is the acoustic source. By applying the change of variables $p(\boldsymbol{r})=f(\boldsymbol{r})\rho^{1/2}(\boldsymbol{r})$,[26,27] Eq. (4) can be rewritten as

$$\nabla^2 f(\boldsymbol{r})+(k^2(\boldsymbol{r})-\rho^{1/2}(\boldsymbol{r})\nabla^2\rho^{-1/2}(\boldsymbol{r}))f(\boldsymbol{r})=-\Phi^{\mathrm{inc}}(\boldsymbol{r}). \tag{5}$$

Equation (5) can be expressed in the integral form

$$p(\boldsymbol{r})=e_s(\boldsymbol{r})+\int_{\Omega}d\boldsymbol{r}'\,\mathcal{O}(\boldsymbol{r}')p(\boldsymbol{r}')G_0(\boldsymbol{r},\boldsymbol{r}'), \tag{6}$$

where $e_s(\boldsymbol{r})$ is the incident field caused by a source located at $\boldsymbol{r}_s$, $s=0,1,\ldots,N_s$, and $G_0(\boldsymbol{r},\boldsymbol{r}')=(i/4)H_0^{(1)}(k_0|\boldsymbol{r}-\boldsymbol{r}'|)$ is the Green's function in cylindrical coordinates. The object function $\mathcal{O}$ is given by

$$\mathcal{O}(\boldsymbol{r})=(k^2(\boldsymbol{r})-k_0^2)-\rho^{1/2}(\boldsymbol{r})\nabla^2\rho^{-1/2}(\boldsymbol{r}). \tag{7}$$

Equation (6) can be discretized using the method of moments (MoMs) and written in matrix form, both for the pressure field inside the computational domain $\bar{p}$ and the scattered field outside the computational domain $\bar{p}^{\mathrm{sc}}$, as

$$\bar{p}=(\bar{I}-\bar{C}\cdot\mathcal{D}(\bar{\mathcal{O}}))^{-1}\cdot\bar{p}^{\mathrm{inc}}, \tag{8}$$

$$\bar{p}^{\mathrm{sc}}=\bar{D}\cdot\mathcal{D}(\bar{\mathcal{O}})\cdot\bar{p}, \tag{9}$$

where $\bar{D}$ is a matrix with the Green's coefficients from each pixel to the receivers, $\bar{C}$ is a matrix with the Green's coefficients among all the pixels, and $\mathcal{D}$ is an operator that transforms a vector into a diagonal matrix.

In order to reconstruct the object function from the scattered field data, an iterative algorithm is used. A trial $\bar{\mathcal{O}}_{(0)}$ is chosen for which the corresponding scattered field is calculated. Next, the object function is updated as $\bar{\mathcal{O}}_{(n+1)}=\bar{\mathcal{O}}_{(n)}+\Delta\bar{\mathcal{O}}_{(n)}$, where $\Delta\bar{\mathcal{O}}_{(n)}$ is given by the regularized optimization problem

$$\Delta\mathcal{O}_{(n)}=\arg\min_{\Delta\mathcal{O}}\|\Delta\bar{p}^{\mathrm{sc}}-\bar{F}_{(n)}\cdot\Delta\mathcal{O}\|_2^2+\gamma\|\Delta\mathcal{O}\|_2^2, \tag{10}$$

where $\Delta\bar{p}^{\mathrm{sc}}$ contains the difference between the predicted and measured scattered fields and γ is the regularization parameter. The Frechet derivative matrix $\bar{F}_{(n)}$ is composed of N_s stacked matrices $\bar{F}_s$ of the form[5]

$$\bar{F}_s=\bar{D}\cdot\{\bar{I}-\mathcal{D}(\bar{\mathcal{O}})\cdot\bar{C}\}^{-1}\cdot\mathcal{D}(\bar{p}_s). \tag{11}$$

The iterative process is repeated until the residual error (RRE), given by $\mathrm{RRE}=\|\Delta\bar{p}^{\mathrm{sc}}\|_2/\|\bar{p}^{\mathrm{sc}}\|_2$, falls within a desired termination tolerance. The regularization parameter was chosen using an extension of the approach presented in Ref. 9,

$$\gamma=0.5\sigma_0^2\max\{10^{\log_2\mathrm{RRE}},10^{-4}\}, \tag{12}$$

where σ_0^2 is the square of the dominant singular value of $\bar{F}_{(n)}$ calculated using the Rayleigh quotient iteration.

The DBIM diverges when the magnitude of the excess phase $\Delta\phi$ accumulated by the acoustic wave when traveling through the scatterer approaches π.[24] For a homogeneous circular cylinder, $\Delta\phi$ can be estimated as

$$\Delta\phi=2k_0a(c_r^{-1}-1). \tag{13}$$

This adimensional quantity will be used for the remainder of this work to report the c_r values for different cylinders.

B. The DF-DBIM approach

From Eq. (5), a linear combination of the reconstructions $\mathcal{O}_i$ at frequencies ω_i, $i=1,2,\ldots,N_f$, allows the separation of c and ρ contributions. Specifically,

$$\mathcal{F}_\rho(\boldsymbol{r})=\frac{(\Sigma_{i=1}^{N_f}\omega_i^2)(\Sigma_{i=1}^{N_f}\omega_i^2\mathcal{O}_i(\boldsymbol{r}))-(\Sigma_{i=1}^{N_f}\omega_i^4)(\Sigma_{i=1}^{N_f}\mathcal{O}_i(\boldsymbol{r}))}{N_f\Sigma_{i=1}^{N_f}\omega_i^4-(\Sigma_{i=1}^{N_f}\omega_i^2)^2}, \tag{14}$$

where $\mathcal{F}_\rho=\rho^{1/2}(\boldsymbol{r})\nabla^2\rho^{-1/2}(\boldsymbol{r})$. The simplest approach, the DF-DBIM, is to use $N_f=2$ because for a fixed spatial resolution (dictated by the maximum frequency f_0 used) only one parameter (the lowest frequency $f_{\min}$ used) has to be chosen. To obtain ρ profiles using DF-DBIM, the differential equation

$$\nabla^2 u(\boldsymbol{r})-\mathcal{F}_\rho(\boldsymbol{r})u(\boldsymbol{r})=\mathcal{F}_\rho(\boldsymbol{r}),\quad \boldsymbol{r}\in\Omega,$$

$$u(\boldsymbol{r})=0,\quad \boldsymbol{r}\notin\Omega \tag{15}$$

has to be solved, where $u(\boldsymbol{r})=(\rho_r^{-1/2}(\boldsymbol{r})-1)$. Equation (15) was solved by converting it to a matrix equation, with ∇^2 implemented using a finite difference template.

The effect of some parameters in the quality of ρ reconstructions is shown in Fig. 2. The minimum frequency $f_{\min}$ was varied between $0.9f_0$ and $0.1f_0$, the DBIM termination tolerance was set to 0.1%, and cylinders with radii of λ_0, $2\lambda_0$, and $4\lambda_0$ were reconstructed, where $\lambda_0=c_0/f_0$. Both the dependence on the value of ρ_r compared to c_r (fixed $\Delta\phi=0.9\pi$ and ρ_r values of $1/c_r$, $1/c_r^2$, and $1/c_r^4$) and $\Delta\phi$ (fixed $\rho_r=1/c_r$ and $\Delta\phi$ values of -0.9π, 0.45π, and -0.45π) were studied. For illustration, ρ profiles corresponding to $\Delta\phi=0.9\pi$ and $\rho_r=1/c_r$ are shown in Fig. 3. Larger density changes required the use of lower $f_{\min}$ values for optimum accuracy at the cost of reduced spatial resolution. In general, larger cylinder radii resulted in more unstable reconstructions when f_0 and $f_{\min}$ were relatively close. Therefore, the optimum $f_{\min}$ value depends on the actual imaging target, but results suggest that reliable results can only be obtained when $f_{\min}$ is small compared to f_0.

The effect of the DBIM termination tolerance was also studied. Figure 4 shows the RMSE curves when reconstructing an $a=2\lambda_0$, $\rho_r=1/c_r$, and $\Delta\phi=0.9\pi$ cylinder using tolerances of 0.1%, 1%, and 2%. The RMSE curves behave smoothly when the DBIM tolerance is low (0.1%) but degrade significantly as the tolerance increases unless $f_{\min}\ll f_0$. This behavior has a direct impact in practical imaging

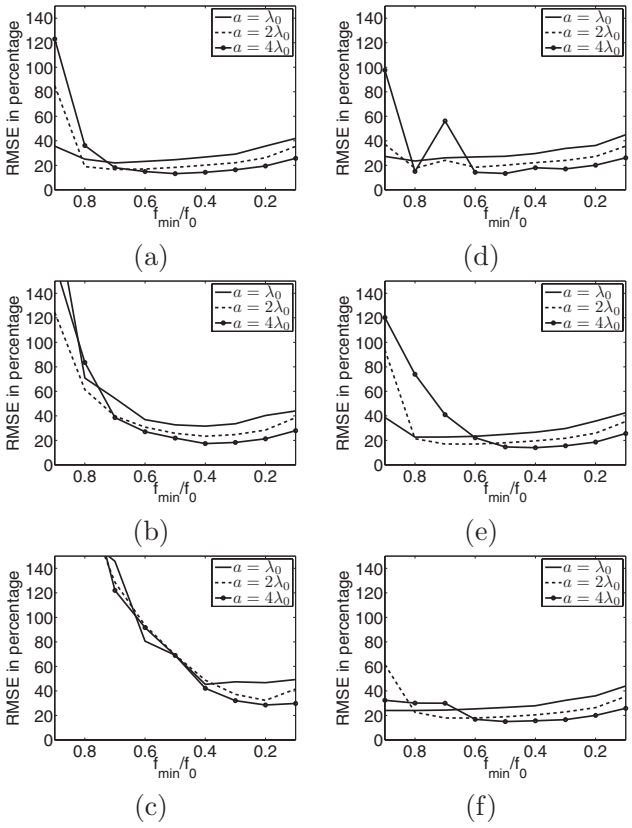


FIG. 2. RMSEs in density reconstructions using the DF-DBIM approach. The corresponding properties of the cylinders are (a) $\rho_r=1/c_r$, $\Delta\phi=0.9\pi$, (b) $\rho_r=1/c_r^2$, $\Delta\phi=0.9\pi$, (c) $\rho_r=1/c_r^4$, $\Delta\phi=0.9\pi$, (d) $\rho_r=1/c_r$, $\Delta\phi=-0.9\pi$, (e) $\rho_r=1/c_r$, $\Delta\phi=0.45\pi$, and (f) $\rho_r=1/c_r$, $\Delta\phi=-0.45\pi$. The DBIM termination tolerance was set to 0.1%.

scenarios because DBIM has to be truncated when the RRE falls below the noise floor to avoid divergence.[4]

It should be emphasized that the RMSE curves are not shown with the intention of identifying an optimum $f_{\min}$ value. In fact, the results from Fig. 2 suggest that several choices of $f_{\min}$ provide solutions with very similar RMSE values. All cases explored, including imaging target and termination tolerance variations, point to using low auxiliary frequencies (less than $0.5f_0$) which limits the spatial resolution of the algorithm. Regardless, the large sensitivity to termination tolerance makes the results provided by DF-DBIM unreliable and therefore alternatives to stabilizing this algorithm need to be studied.

C. DF-DBIM and total variation regularization

Regularization can be used to stabilize the separation of speed of sound and density profiles. One common approach is the generalized Tikhonov regularization, which consists of solving the optimization problem[28]

$$\hat{u}=\arg\min_{\bar{u}}\|\bar{F}_\rho-\bar{G}\cdot\bar{u}\|_2^2+\gamma\sum_{i=1}^{N}(|(\bar{L}\cdot\bar{u})_i|^2+\beta)^{k/2}, \quad (16)$$

where $\bar{u}$ is a vector representation of $u(\boldsymbol{r})$, $\bar{F}_\rho$ is a vector with the values of $\mathcal{F}_\rho(\boldsymbol{r})$, $\bar{G}=(\mathcal{L}-\mathcal{D}(\bar{F}_\rho))$, $\mathcal{L}$ is the Laplacian matrix, and β is a small positive constant ($\beta=10^{-10}$) introduced to avoid gradient singularities. The solution $\hat{u}$ of Eq. (16) is given by

$$\hat{u}=[(\bar{G}^H\cdot\bar{G}+\gamma\bar{L}^H\cdot\mathcal{D}(\bar{W}_\beta(\hat{u}))\cdot\bar{L})^{-1}\bar{G}^H]\bar{F}_\rho, \quad (17)$$

where $\bar{W}_\beta(\hat{u})_i=(k/2)((|(\bar{L}\cdot\hat{u})_i|^2+\beta)^{1-k/2})$. In particular, total variation (TV) regularization ($k=1$ and L equal to the gradient operator)[29] was explored. For each value of $f_{\min}$ several γ values were used and the optimum reconstruction was selected based on RMSE minimization. It should be emphasized that in real applications the selection of γ becomes an important issue because the ideal object function is not available.

It was found through simulations that even though TV was able to improve on the reconstructions of single cylinders with termination tolerances below 2%, either slightly larger tolerances or more complicated scatterer geometries resulted in eventual divergence due to the inherent sensitivity of DF-DBIM to the termination tolerance. This is illustrated in Fig. 5 for the reconstruction of a concentric cylinder phantom, for which the analytic scattering solution is available.[9] The radii and Δc of the inner and outer cylinders were λ_0 and -10% and $2\lambda_0$ and -5%, respectively. In both cylinders, $\rho_r=1/c_r$. The termination tolerance was set to 5%, higher than in previous examples.

Even though TV regularization improved upon the nonregularized inversion as reported in Fig. 5, the performance was still unsatisfactory as evidenced by the large RMSEs (larger than 40% for all cases). A sample reconstruction using DF-DBIM is shown in Fig. 5(b) using $f_{\min}=0.4f_0$. When regularization is not used, spurious slopes appeared in ideally homogeneous regions, which is the cause of the large RMSE. TV regularization helps in reducing these artifacts and improving the RMSE, but the reconstruction error is still fairly large. When using very low frequencies ($f_{\min}=0.2f_0$), the effect of TV regularization was negligible and the optimum reconstruction was similar to the one obtained with no regularization. Therefore, TV is an aid but not a complete solution to the DF-DBIM robustness problem, and further algorithmic developments are required to reduce the sensitivity of DF-DBIM prior to applying TV regularization.

IV. *T*-MATRIX APPROACH FOR DENSITY IMAGING

The second class of algorithms consists of directly inverting the full wave equation including changes in ρ. Implementations include the use of MoM formulations,[20] *T*-matrix approaches,[21] and contrast source inversion methods.[22,30] The work in Ref. 30 only deals with the case of constant κ. Of the other approaches, the *T*-matrix method in Ref. 21 was chosen as representative for this class of algorithms because it is the only work, to the authors' knowledge, that went as far as providing experimental reconstructions.

A. Theoretical background

The details of the *T*-matrix algorithm are presented in Ref. 21. The computational domain is divided in N homoge-

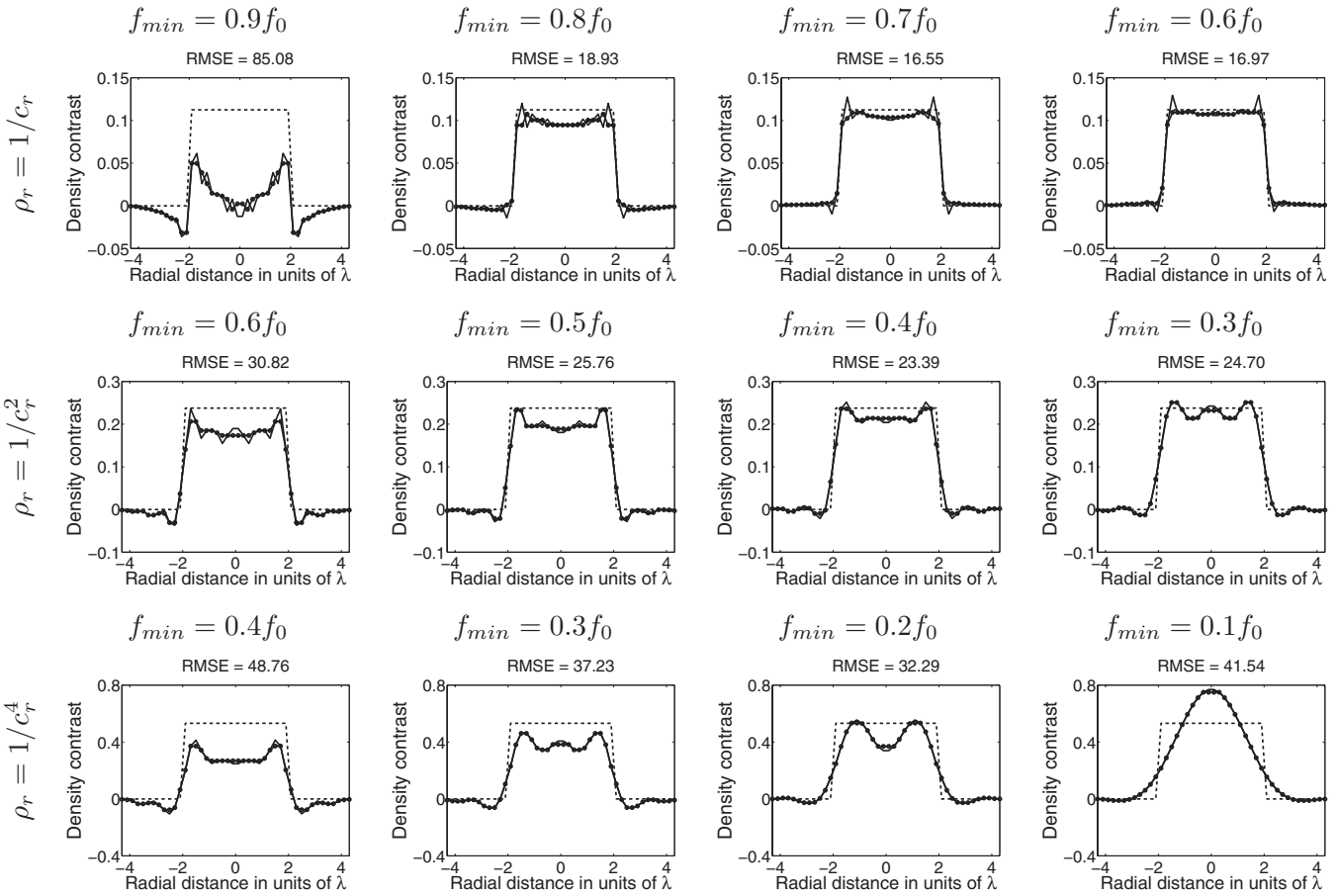


FIG. 3. Reconstruction of ρ profiles of cylinders with $\Delta\phi=0.9\pi$ and radius $2\lambda_0$ using DF-DBIM. The reconstructed (solid), ideal (dashed), and median filtered (dotted) profiles are shown. The DBIM termination tolerance was set to 0.1%.

neous subscatterers distributed on a rectangular grid of pixel size h. The total acoustic field produced at some point $\boldsymbol{r}_p$ in space is given by

$$p(\boldsymbol{r}_p) = \psi(\boldsymbol{r}_p - \boldsymbol{r}_s) \cdot \bar{f}_s + \sum_{m=1}^{N} \psi(\boldsymbol{r}_p - \boldsymbol{r}_m) \cdot \bar{a}_m, \tag{18}$$

where $\boldsymbol{r}_s$ is the location of the source, $\boldsymbol{r}_m$ is the location of the mth subscatterer, $\psi(\boldsymbol{r})$ is a vector of cylindrical harmonics, and $\bar{f}_s$ and $\bar{a}_m$ are vectors containing the amplitudes of the cylindrical harmonic fields generated by the source and the mth subscatterer, respectively.

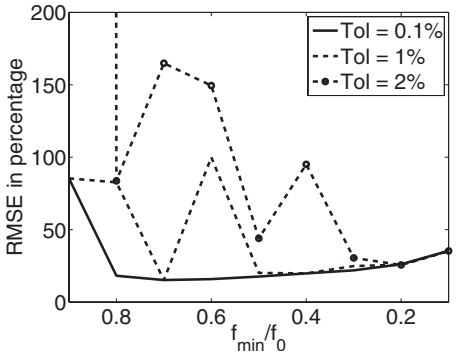


FIG. 4. Effect of the DBIM termination tolerance by reconstructing a cylinder with radius $2\lambda_0$. $\rho_r=1/c_r$, and $\Delta\phi=0.9\pi$ using termination tolerances of 0.1%, 1%, and 2%.

The equation above can be rewritten using the jth subscatterer as the origin for all the cylindrical harmonics using the addition theorem of Bessel functions[31] as

$$p(\boldsymbol{r}_p) = \psi(\boldsymbol{r}_{pj}) \cdot \bar{a}_j + \hat{\psi}(\boldsymbol{r}_{pj}) \cdot \left(\sum_{m \neq j} \bar{\alpha}_{jm} \cdot \bar{a}_m + \bar{e}_{js} \right),$$

$$[\alpha_{jm}]_{kl} = H_{k-l}^{(1)}(k_0|\boldsymbol{r}_{mj}|) e^{-i(k-l)\theta_{mj}}, \tag{19}$$

where $[\hat{\psi}(\boldsymbol{r})]_k = J_k(k_0 r)e^{il\theta}$ and, for line sources, $[\bar{e}_{js}]_k = H_k^{(1)} \times (k_0|\boldsymbol{r}_{sj}|)e^{-ik\theta_{sj}}$. If $h \ll \lambda$, the harmonics $l=0,1,-1$ are sufficient to characterize the scattering process. The vector of equivalent induced sources $\bar{a}_s$ when the transmitter is at the position $\boldsymbol{r}_s$ is approximated as

$$\{\bar{I} - \mathcal{D}(\bar{R}) \cdot \bar{A}\} \cdot \bar{a}_s = \mathcal{D}(\bar{R}) \cdot \bar{e}_s, \tag{20}$$

where $\bar{A}$ is a matrix containing the $[\alpha_{jm}]_{kl}$ coefficients, $\mathcal{D}(\bar{R})$ is a diagonal matrix with the reflection coefficients given by Eq. (2) using a pixel radius $a=h/\sqrt{\pi}$ for the harmonics $k=0,1,-1$, and $\bar{e}_s$ is a vector whose elements are given $\bar{e}_{js}$. If the total pressure $\bar{e}_{ts}$ at the scatterer is defined such that $\bar{a}_s = \mathcal{D}(\bar{R}) \cdot \bar{e}_{ts}$, then from Eq. (20),

$$\bar{e}_{ts} = [\bar{I} - \bar{A} \cdot \mathcal{D}(\bar{R})]^{-1} \cdot \bar{e}_s. \tag{21}$$

The T-matrix formulation can be inverted using the same iterative process used in the DBIM. The object function

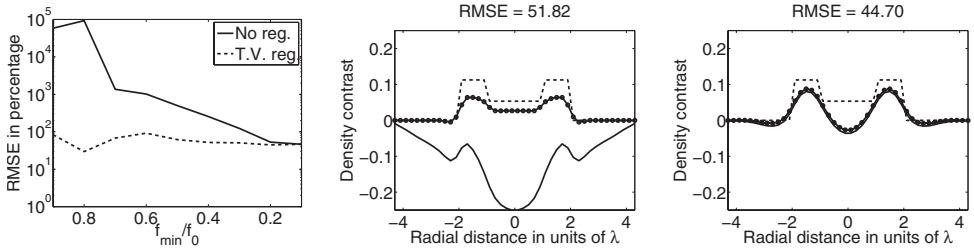


FIG. 5. Reconstructions of $\Delta\rho$ profiles of a concentric cylinder phantom using DF-DBIM and TV. The DBIM termination tolerance was set to 5%. The RMSE curves both without (solid) and with (dashed) TVs are shown in (a). Reconstructions using $f_{\min}=0.4f_0$ (b) and $f_{\min}=0.2f_0$ (c) are also shown. In (b) and (c), the ideal (dashed) and reconstructed profiles both with (dotted) and without (solid) TV regularizations are shown.

vector is here defined as $\mathcal{O}=[\{\bar{R}\}_{k=0};\{\bar{R}\}_{k=1}]$ because $R_1(\kappa,\rho)=R_{-1}(\kappa,\rho)$. By analogy with Eq. (11), the Frechet derivative matrix blocks $\bar{F}_s$ are given by

$$\bar{F}_s = \bar{\psi}\cdot\{\bar{I}-\mathcal{D}(\bar{R})\cdot\bar{A}\}^{-1}\cdot\mathcal{M}(\bar{e}_{ts}), \tag{22}$$

$$\mathcal{M}(\bar{e}_{ts}) = \begin{bmatrix} \mathcal{D}(\{\bar{e}_{ts}\}_{k=0}) & 0 \\ 0 & \mathcal{D}(\{\bar{e}_{ts}\}_{k=1}) \\ 0 & \mathcal{D}(\{\bar{e}_{ts}\}_{k=-1}) \end{bmatrix}. \tag{23}$$

B. Convergence of the *T*-matrix algorithm using single frequency data

The T-matrix approach was used to reconstruct ρ and κ profiles of homogeneous circular cylinders with radius 2λ and $\Delta\phi=0.9\pi$. Four cases were considered: $\rho_r=1$, $\rho_r=1/c_r$, $\rho_r=1/c_r^2$, and $\rho_r=1/c_r^4$. The results are shown in Fig. 6. It can be observed that the mean reconstructed density value ρ_m did not significantly change among all cases despite the fact that the true values changed over a large range. This is due to the fact that the scatterer was not small compared to the wavelength and therefore the scattered field was mainly dominated by the changes in c, which was held constant for all simulations. For all simulations c was the only parameter reconstructed with good numerical accuracy.

Given that the mean reconstructed compressibility κ_m and density ρ_m converge in the mean to the same values, it is of interest to analyze the characteristics of the expected reconstructed profiles for κ and ρ. If the c contrast dominates the scattering, c_r will be properly reconstructed and all candidate solutions should satisfy $\kappa_r\rho_r=1/c_r^2$. For homogeneous objects, the candidate object function can be represented by the 2×1 vector $\hat{\mathcal{Q}}(\kappa_r,\rho_r)=[R_0(\kappa_r,\rho_r),R_1(\kappa_r,\rho_r)]$. Given that the algorithm solves for the object function in the least squares sense, the quantities $\hat{\kappa}$ and $\hat{\rho}$ defined as

$$\hat{\kappa},\hat{\rho} = \arg\min_{\kappa_r,\rho_r}\{\|\hat{\mathcal{O}}\|^2\} \quad \text{subject to } \kappa_r\rho_r=\frac{1}{c_r^2} \tag{24}$$

are an approximation of the minimum norm solution of the inversion problem, and therefore it is expected that κ_m and ρ_m should correlate with these quantities when using an all-zero initial guess.

Several simulations were performed in order to study the behavior of the pairs (κ_m,ρ_m) and $(\hat{\kappa},\hat{\rho})$ using cylinders of radii $2\lambda_0$ and $4\lambda_0$ as imaging targets. The speed of sound

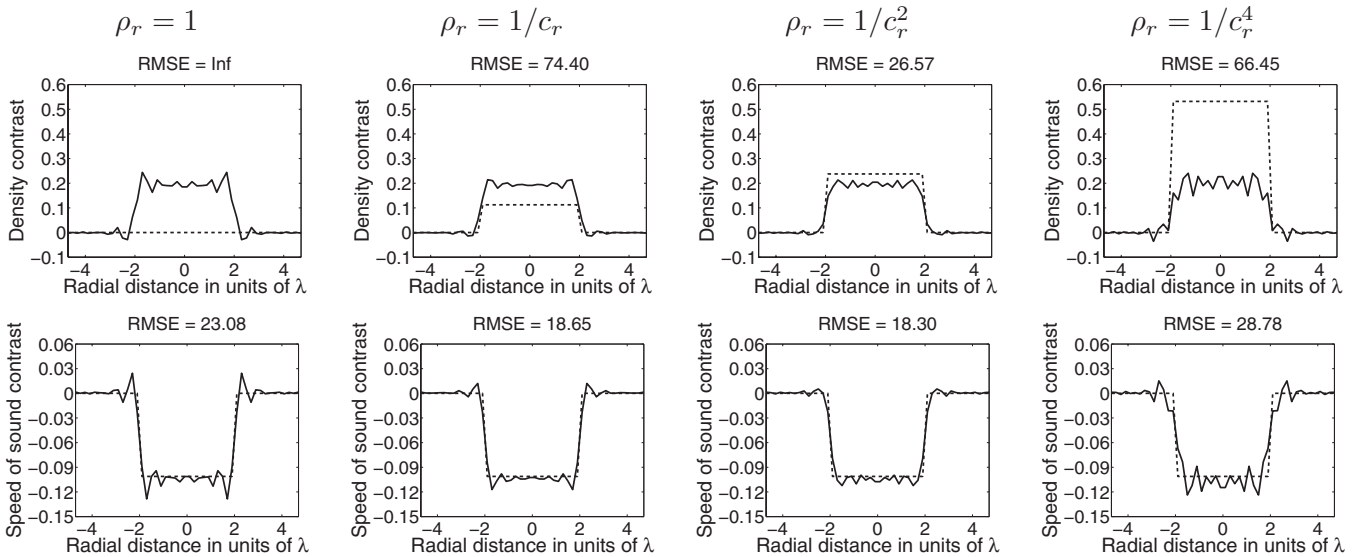


FIG. 6. Reconstructions of 2λ radius cylinders with $\Delta\phi=0.9$ using the single frequency T-matrix approach. First column: $\rho_r=1$. Second column: $\rho_r=1/c_r$. Third column: $\rho_r=1/c_r^2$. Fourth column: $\rho_r=1/c_r^4$. Both the reconstructed (solid) and ideal (dashed) profiles are shown. The T-matrix termination tolerance was set to 2%.

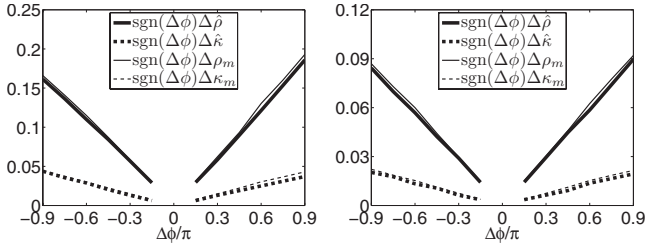


FIG. 7. Mean reconstructed excess density ρ_m and compressibility κ_m values and the corresponding $\hat{\kappa}$ and $\hat{\rho}$ values when reconstructing $2\lambda_0$ (left) and $4\lambda_0$ (right) radii cylinders with different speed of sound values.

contrast for each cylinder size was changed so that the excess phase $\Delta\phi$ ranged between -0.9π and 0.9π. The point $\Delta\phi=0$ was excluded because it would imply $c_r=1$, and therefore density changes only govern the scattering process. The ρ_m and κ_m values were calculated from tomographic reconstructions using scattered data corresponding to $\rho_r=1/c_r$ ($2\lambda_0$ radii cylinders) and $\rho_r=1/c_r^2$ ($4\lambda_0$ radii cylinders). The $\hat{\kappa}$ and $\hat{\rho}$ values for each case were estimated using an exhaustive search in the interval $\kappa_r \in [0.8, 1.2]$. The results are shown in Fig. 7. The mean error between the reconstructed mean density contrast ρ_m and the corresponding minimum norm solution $\hat{\rho}$ was only 4.1% for the $2\lambda_0$ cylinders and 3.9% for the $4\lambda_0$ cylinders, which indicates a high correlation between the two quantities.

C. Convergence of the *T*-matrix algorithm using multiple frequency data

As observed in Sec. II, the dependence of the scattered field on density changes becomes more significant for low frequencies. Therefore, the use of multiple frequencies may improve the convergence of the *T*-matrix approach, which has been briefly explored in Ref. 21. The multiple frequency data were processed using the frequency hopping approach,[6] i.e., the sequential use of progressively larger frequencies to refine the inverse scattering reconstructions. The frequency hopping profiles were obtained using all frequencies $f=2^{-m}f_0$, $m=M-1, M-2, \ldots, 0$, where M and f_0 are the number of frequencies and the maximum frequency used in the reconstructions, respectively.

First, a set of simulations was performed to reconstruct cylinders of radii λ_0, $2\lambda_0$, and $4\lambda_0$. For all simulations, $\Delta\phi=0.9\pi$ and $\kappa_r=\rho_r$. The reconstructions were obtained using minimum frequencies $f_{\min}$ ranging between $f_0/2$ and $f_0/64$. The reconstructed density profiles are shown in Fig. 8. The RMSE in the ρ reconstructions initially decreases with decreasing $f_{\min}$. However, at some point the improvement stops and the RMSE starts to increase slowly. For the analyzed cases, the best ρ reconstructions in terms of the RMSE occurred when $f_{\min}=(\lambda_0/8a)f_0$ was used.

The trend observed in the cases analyzed in Fig. 8 indicates that the ability of the *T*-matrix approach using fre-

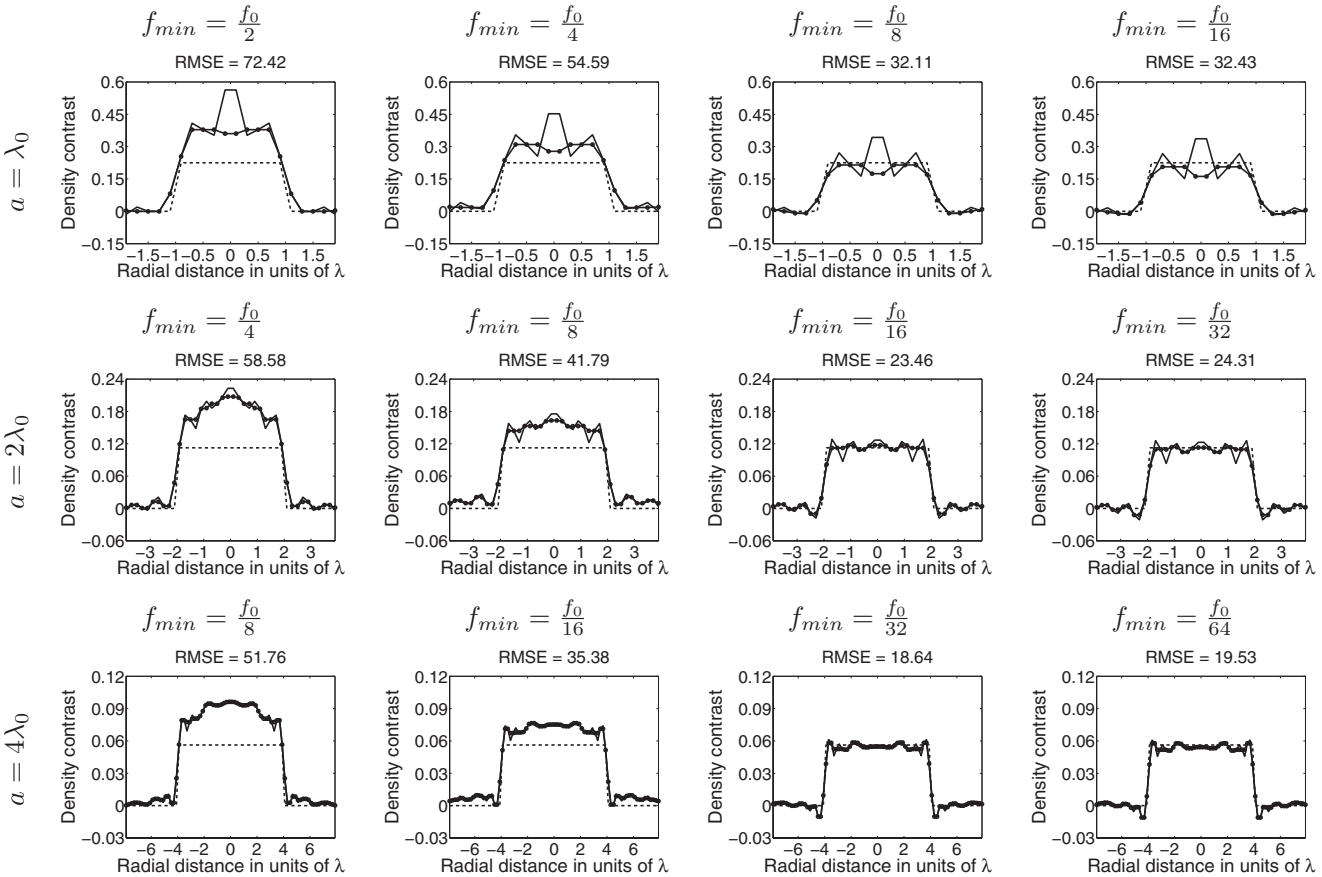


FIG. 8. Frequency hopping reconstructions using the *T*-matrix approach for cylinders with a fixed excess phase $\Delta\phi=0.9\pi$ and $\rho_r=1/c_r$. The reconstructed (solid), ideal (dashed), and median filtered (dotted) density profiles are shown. The reported RMSE values correspond to the nonfiltered reconstructions. The *T*-matrix termination tolerance was set to 2%.

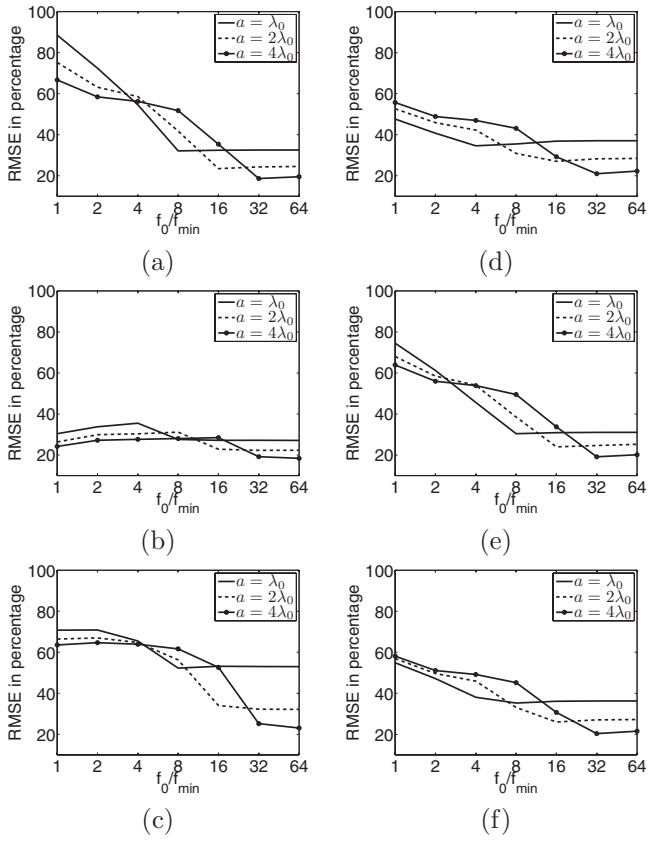


FIG. 9. RMSEs in density reconstructions using frequency hopping and the *T*-matrix approach. The corresponding properties of the cylinders are (a) $\rho_r=1/c_r$, $\Delta\phi=0.9\pi$, (b) $\rho_r=1/c_r^2$, $\Delta\phi=0.9\pi$, (c) $\rho_r=1/c_r^4$, $\Delta\phi=0.9\pi$, (d) $\rho_r=1/c_r$, $\Delta\phi=-0.9\pi$, (e) $\rho_r=1/c_r$, $\Delta\phi=0.45\pi$, and (f) $\rho_r=1/c_r$, $\Delta\phi=-0.45\pi$. The *T*-matrix termination tolerance was set to 2% for all simulations.

quency hopping works only if the cylinder radius is small compared to the maximum λ used. In order to verify that such a trend indeed exists, the behavior of the RMSE in the reconstructed $\Delta\rho$ using the same simulation sets as in Sec. III B was studied. The results are presented in Fig. 9. For all cases, the scatterer size was the main factor that affected the reconstruction quality.

V. DISCUSSION

The results from Secs. III and IV suggest that both the DF-DBIM and *T*-matrix algorithms were able to generate density images with RMSE values lower than 30% when reconstructing cylinders up to eight wavelengths in diameter with moderate density changes, provided that certain requirements related to the imaging apparatus [i.e., signal-to-noise ratio (SNR) and bandwidth] were met. These requirements and their implications for the experimental implementations of these algorithms are discussed below.

A. DF-DBIM approach

The DF-DBIM approach is severely affected by inaccuracies of reconstructions performed at any one frequency. Even in the absence of noise, the denominator of Eq. (14) is very small when f_0 and $f_{\min}$ are too close to each other and therefore numerical errors are amplified. On the other hand, when $f_0 \gg f_{\min}$, density reconstructions will suffer from spatial resolution degradation because of the limited spectral support of the far-field measurements.[32]

Simulation results suggest that the performance of DF-DBIM depends on properties of the imaging target that are not known *a priori*, i.e., ρ_r. However, the main drawback of the technique is its sensitivity to the termination tolerance of the single frequency DBIM reconstructions. Even for low termination tolerances (1%–2%), the RMSE curves become highly irregular, as shown in Fig. 2. This further complicates the selection of an appropriate $f_{\min}$, especially for real applications in which the SNR dictates the minimum tolerance that can be used in the reconstructions. The limited results presented here suggest that reconstructions using very low $f_{\min}$ values are less sensitive to changes in the termination tolerance.

TV regularization was also briefly explored as a means to improve the performance of the DF-DBIM approach. TV may help improve the reconstruction error, but even with the use of TV, DF-DBIM was plagued by its large sensitivity to the DBIM termination tolerance. Further algorithmic developments may be tested to ameliorate the performance of this approach given its potential to obtain density information using relatively small bandwidths (less than an order of magnitude) which may readily be obtainable with current transducer technology.

B. *T*-matrix approach

The *T*-matrix approach fails to converge when the scatterer size a is of moderate size compared to λ because the scattering pattern is dominated in the mean square sense by c variations rather than the actual values of κ and ρ, as illustrated in Fig. 1. As a becomes smaller than λ, scattering theory predicts that the scattered pressure pattern becomes more sensitive to the actual values of κ and ρ. This explains why the use of frequency hopping with minimum frequencies such that $ka \approx 1$ resulted in accurate ρ reconstructions. However, using even lower frequencies did not help improve the accuracy of the inversions. This is expected from the results of Sec. II because when $ka \ll 1$, the scattered pressure pattern reaches the asymptotic Rayleigh regime given by Eq. (3) and no further information is obtained.

For a homogeneous cylinder of radius a, divergence in c can be avoided by the use of frequencies low enough such that[24]

$$k_0 a < \frac{\pi}{2}\frac{c_r}{|c_r - 1|}. \tag{25}$$

Density imaging using the *T*-matrix approach diverges due to a different mechanism, namely, the weak dependence of the scattering pattern on ρ for large ka values. The convergence condition can be approximated as $ka < 1$. Therefore, for all practical biomedical imaging applications the condition for *T*-matrix convergence in κ and ρ is more restrictive than the one for DBIM convergence in c. Convergence in the reconstruction in c_r when using the *T*-matrix approach is obtained regardless of the divergence of κ and ρ, as long as the condition in Eq. (25) is satisfied. These results have a profound

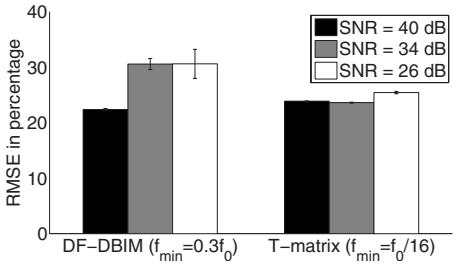


FIG. 10. RMSE in the reconstruction of circular cylinders with $a=2\lambda_0$ and $\rho_r=1/c_r$. The SNRs were set to 40 dB (black bars), 34 dB (gray bars), and 26 dB (white bars), respectively.

impact on the feasibility of the use of the *T*-matrix approach. Even for imaging targets as small as four wavelengths in size ($a=2\lambda_0$) a frequency jump larger than an order of magnitude is needed in order to obtain good numerical accuracy. This requirement becomes more restrictive with increasing target size and severely limits experimental implementations of the algorithm.

Quadratic regularization, which was used in the present work to stabilize the *T*-matrix inversions, is commonly chosen because of simplicity but not because of optimality. Some studies[22,30] suggest that improved results in terms of spatial resolution in variable density inverse scattering can be obtained by using TV regularization. This approach has also been used in constant density inverse scattering problems.[33] However, it remains to be studied if the *T*-matrix bandwidth requirement can be reduced when using this approach.

C. Effects of noise

Additive noise is another practical consideration that has an effect on the quality of the reconstructions. In order to provide a preliminary evaluation of the effects of noise, a cylinder with radius $2\lambda_0$ and $\rho_r=1/c_r$ was reconstructed using both the DF-DBIM and *T*-matrix approaches. The minimum frequencies $f_{\min}$ were chosen using values that provided good reconstructions in Secs. III and IV of the present work: $f_{\min}=0.3f_0$ and $f_{\min}=f_0/16$ for the DF-DBIM and *T*-matrix approaches, respectively. The SNRs were set to 40 dB (1% noise), 34 dB (2% noise), and 26 dB (5% noise). All the reconstructions were stopped when the RRE dropped below the noise level. Twenty-four simulations were conducted for each setting with different zero-mean, Gaussian noise realizations. The mean and standard deviation of the reconstruction RMSEs are shown in Fig. 10.

The small variation in both the mean and variance of the RMSEs as a function of the noise level for the *T*-matrix approach suggests that the technique is reasonably robust to noise. The mean RMSEs for the DF-DBIM case are very close to the ones obtained in Sec. III B, which suggests that the RMSEs are mainly biased by the termination tolerance. The variance induced by the noise itself was not too large compared to the mean RMSE values, although it was larger than the one for the *T*-matrix reconstructions especially for smaller SNR values. Based on diffraction tomography studies[15] it could be expected that DF-DBIM is a noise sensitive technique. However, in the present case f_0 is large compared to $f_{\min}$ which keeps the noise from being significantly amplified. In addition, the DBIM profiles were obtained using a regularized algorithm that efficiently reduces reconstruction fluctuations due to noise in the measurements.

VI. CONCLUSIONS

The present work presents several contributions to the field of inverse scattering. The most important contribution is the joint performance analysis of two inverse scattering algorithms (DF-DBIM and *T*-matrix) designed to reconstruct density profiles. The performances of both approaches were studied under different conditions considering the effects of the cylinder size relative to the acoustic wavelength, density contrast, and speed of sound contrast.

Second, this work suggests that the performance of the DF-DBIM algorithm is severely affected by the termination tolerance of the single frequency DBIM reconstructions. The use of TV regularization improved the performance of DF-DBIM but was not enough to guarantee convergence to a proper solution for moderate termination tolerances. Further improvements may be studied in order to increase the robustness of this algorithm.

Third, this work indicated that the single frequency *T*-matrix algorithm only provides accurate reconstructions of speed of sound, as long as speed of sound dominates scattering over density contributions. This condition is met for targets of moderate to large sizes ($ka>1$) even when density and speed of sound contrasts are comparable. Even further, the profiles to which density and compressibility reconstructions converge in the single frequency case were successfully characterized.

Finally, the convergence criteria for the multiple frequency *T*-matrix algorithm were found through simulations to be mainly affected by the imaging target size. The minimum frequency required to obtain accurate density reconstructions must satisfy $ka<1$. This constraint is more severe than the one for speed of sound reconstructions and was related to the insensitivity of scattering patterns to density variations for large *ka* values. This algorithm may require an excessively large bandwidth to reconstruct the density profile of large targets. Further studies are required to investigate if variations in the algorithm allow a reduction in the bandwidth required for convergence.

In summary, neither algorithm appears amenable for direct experimental implementation. Their fundamental limitations, presented here in a cohesive manner, will serve as reference points for further algorithmic improvements required for practical experimental implementation of density imaging on ultrasound tomographic systems.

ACKNOWLEDGMENTS

The authors would like to thank Dr. S. Bond for discussions on the DF-DBIM algorithm. This work was supported in part by a grant from the 3M Corporation.

[1]J. Greenleaf, S. Johnson, S. Lee, G. Herman, and E. Wood, "Algebraic reconstruction of spatial distributions of acoustic absorption within tissue from their two-dimensional acoustic projections," Acoust. Hologr. **5**, 591–603 (1974).

[2]J. Greenleaf, S. Johnson, W. Samayoa, and F. Duck, "Algebraic reconstruction of spatial distributions of acoustic velocities in tissue from their time-of-flight profiles," Acoust. Hologr. **6**, 71–90 (1975).
[3]B. Robinson and J. Greenleaf, "The scattering of ultrasound by cylinders: Implications for diffraction tomography," J. Acoust. Soc. Am. **80**, 40–49 (1986).
[4]W. C. Chew and Y. M. Wang, "Reconstruction of two-dimensional permittivity distribution using the distorted Born iterative method," IEEE Trans. Med. Imaging **9**, 218–225 (1990).
[5]D. Borup, S. Johnson, W. Kim, and M. Berggren, "Nonperturbative diffraction tomography via Gauss-Newton iteration applied to the scattering integral equation," Ultrason. Imaging **14**, 69–85 (1992).
[6]W. C. Chew and J. H. Lin, "A frequency-hopping approach for microwave imaging of large inhomogeneous bodies," IEEE Microw. Guid. Wave Lett. **5**, 440–441 (1995).
[7]S. A. Goss, R. L. Johnston, and F. Dunn, "Comprehensive compilation of empirical ultrasonic properties of mammalian tissues," J. Acoust. Soc. Am. **64**, 423–457 (1978).
[8]S. A. Goss, R. L. Johnston, and F. Dunn, "Compilation of empirical ultrasonic properties of mammalian tissues II," J. Acoust. Soc. Am. **68**, 93–108 (1980).
[9]R. J. Lavarello and M. L. Oelze, "A study on the reconstruction of moderate contrast targets using the distorted Born iterative method," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **55**, 112–124 (2008).
[10]S. A. Johnson, T. Abbott, R. Bell, M. Berggren, D. Borup, D. Robinson, J. Wiskin, S. Olsen, and B. Hanover, "Noninvasive breast tissue characterization using ultrasound speed and attenuation," Acoust. Imaging **28**, 147–154 (2007).
[11]M. L. Oelze, W. D. O'Brien, Jr., J. P. Blue, and J. F. Zachary, "Differentiation and characterization of rat mammary fibroadenomas and 4T1 mouse carcinomas using quantitative ultrasound imaging," IEEE Trans. Med. Imaging **23**, 764–771 (2004).
[12]J. Mamou, M. L. Oelze, W. D. O'Brien, Jr., and J. F. Zachary, "Identifying ultrasonic scattering sites from three-dimensional impedance maps," J. Acoust. Soc. Am. **117**, 413–423 (2005).
[13]S. J. Norton, "Generation of separate density and compressibility images in tissue," Ultrason. Imaging **5**, 240–252 (1983).
[14]S. Mensah and J. P. Lefebvre, "Enhanced compressibility tomography," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **44**, 1245–1252 (1997).
[15]A. J. Devaney, "Variable density acoustic tomography," J. Acoust. Soc. Am. **78**, 120–130 (1985).
[16]M. Moghaddam and W. Chew, "Variable density linear acoustic inverse problem," Ultrason. Imaging **15**, 255–266 (1993).
[17]M. A. Anastasio, D. Shi, and T. Deffieux, "Image reconstruction in variable density acoustic tomography," Proc. SPIE **5750**, 326–331 (2005).
[18]M. J. Berggren, S. A. Johnson, B. L. Carruth, W. W. Kim, F. Stenger, and P. K. Kuhn, "Ultrasound inverse scattering solutions from transmission and/or reflection data," Proc. SPIE **671**, 114–121 (1986).
[19]S. Kwon and M. Jeong, "Ultrasound inverse scattering determination of speed of sound, density and absorption," Proc.-IEEE Ultrason. Symp. **2**, 1631–1634 (1998).
[20]S. A. Johnson and F. Stenger, "Ultrasound tomography by Galerkin or moment methods," in *Lecture Notes in Medical Informatics*, edited by O. Nalcioglu and Z. Cho (Springer-Verlag, New York, NY, 1984), Vol. 23, pp. 254–275.
[21]J. Lin and W. Chew, "Ultrasonic imaging by local shape function method with CGFFT," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **43**, 956–969 (1996).
[22]K. W. A. van Dongen and W. M. D. Wright, "A full vectorial contrast source inversion scheme for three-dimensional acoustic imaging of both compressibility and density profiles," J. Acoust. Soc. Am. **121**, 1538–1549 (2007).
[23]T. Cavicchi and W. D. O'Brien, Jr., "Acoustic scattering of an incident cylindrical wave by an infinite circular cylinder," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **35**, 78–80 (1988).
[24]T. Cavicchi, S. Johnson, and W. D. O'Brien, Jr., "Application of the sinc basis moment method to the reconstruction of infinite circular cylinders," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **35**, 22–33 (1988).
[25]P. M. Morse and K. U. Ingard, *Theoretical Acoustics* (McGraw-Hill, Princeton, NJ, 1968).
[26]S. A. Johnson, F. Stenger, C. Wilcox, J. Ball, and M. J. Berggren, "Wave equations and inverse solutions for soft tissue," Acoust. Imaging **11**, 409–424 (1982).
[27]S. Pourjavid and O. J. Tretiak, "Numerical solution of the direct scattering problem through the transformed acoustical wave equation," J. Acoust. Soc. Am. **91**, 639–645 (1992).
[28]W. C. Karl and M. Cetin, "Feature-enhanced synthetic aperture radar image formation based on nonquadratic regularization," IEEE Trans. Image Process. **10**, 623–631 (2001).
[29]R. Akar and C. R. Vogel, "Analysis of bounded variation penalty methods for ill-posed problems," Inverse Probl. **10**, 1217–1229 (1994).
[30]G. Pelekanos, A. Abubakar, and P. M. van den Berg, "Contrast source inversion methods in elastodynamics," J. Acoust. Soc. Am. **114**, 2825–2834 (2003).
[31]W. C. Chew, *Waves and Fields in Inhomogeneous Media* (IEEE, Piscataway, NJ, 1995).
[32]T. J. Cui, W. C. Chew, X. X. Yin, and W. Hong, "Study of resolution and super resolution in electromagnetic imaging for half-space problems," IEEE Trans. Antennas Propag. **52**, 1398–1411 (2004).
[33]X. Zhang, S. Broschat, and P. Flynn, "A comparison of material classification techniques for ultrasound inverse imaging," J. Acoust. Soc. Am. **111**, 457–467 (2002).

Coupled vibration analysis of the thin-walled cylindrical piezoelectric ceramic transducers

Boris Aronov[a)]
BTech Acoustic LLC, Acoustics Research Laboratory, Advanced Technology and Manufacturing Center, and Department of Electrical and Computer Engineering, University of Massachusetts Dartmouth, 151 Martine Street, Fall River, Massachusetts 02723

(Received 21 June 2008; revised 1 December 2008; accepted 4 December 2008)

Analysis of electromechanical transducers employing axially symmetric vibrations of piezoelectric ceramic thin-walled tubes of arbitrary aspect ratio is presented based on application of the energy method [B. S. Aronov, J. Acoust. Soc. Am. **117**, 210–220 (2005)]. The suggestion made by Giebe and Blechshmidt [Ann. Physik, Ser. 5 **18**, 417–485 (1933)] regarding representation of a tube vibration as a coupled vibration of two partial systems is used to choose the assumed modes of the piezoceramic tube vibration, and the energy of the flexural deformations accompanying the extensional vibration of a tube is also taken into consideration. The Lagrange type equations describing the transducer vibrations are derived, and the equivalent electromechanical circuit is introduced that conveniently represents the solution. The resonance frequencies, effective coupling coefficients, and velocity distributions for differently poled piezoceramic tubes as functions of their height-to-diameter aspect ratio are calculated. The validity of the equivalent circuit is extended to the case that a piezoceramic tube is mechanically and acoustically loaded on the ends and acoustically loaded on its outer surface. The results of calculations are in good agreement with the results of experimental investigations.
 [DOI: 10.1121/1.3056560]

PACS number(s): 43.38.Ar, 43.38.Fx, 43.40.At, 43.40.Ey [AJZ] Pages: 803–818

I. INTRODUCTION

The cylindrical piezoelectric ceramic transducers are widely used for underwater applications. Calculation of their parameters is well known in the case that the transducers are built from the thin-walled short rings, for which the assumption of one-dimensional nature of vibrations in the circumferential direction is valid. With increasing of the height-to-diameter aspect ratio of the comprising piezoelements the one-dimensional approximation fails, and vibrations of the cylindrical piezoelements have to be considered as two-dimensional coupled vibrations in the circumferential and axial directions.

The first treatment of vibration of piezoelectric ceramic thickness poled shells in the two-dimensional approximation was carried out by Haskins and Walsh[1] in the framework of the "membrane" theory of shells following the treatment of vibration of the passive thin isotropic elastic tubes described by Love.[2] Afterwards a number of papers were published, related to different aspects of the problem of coupled vibrations in piezoceramic tubes.[3–9] Only a few of these references[4,9] considered calculation of properties of underwater electroacoustic transducers, made from the tubes, for which the analysis of the coupled vibrations was applied.

The common feature of all of the references is that they treat the problem by means of solving the partial differential equations of motion of the tubes as piezoelectric elastic medium under certain electrical and mechanical boundary conditions. In the case that the tubes are treated as electroacoustic transducers, the boundary conditions on their radiating surface involve a reaction of acoustical loading, and the problem becomes a rather complicated coupled piezoelectric-elastic and coupled elastoacoustic problem. The main differences between the cited references consist in the mathematical treatment of the problem and the specific assumptions made regarding the thickness of the tubes, namely, whether the tubes are considered as thin-walled (the "membrane theory" approximation)[3–5] or as finite thickness.[6–8] Because of complexity of the problem, solutions in all of the referenced cases are obtained by numerical computation, and the most of the papers are concerned with improving of the mathematical models rather than with revealing the trends in the electromechanical and electroacoustic characteristics of transducer behavior as functions of aspect ratio of the comprising piezoceramic tubes. The dependence of parameters of the thickness poled air-backed cylindrical transducers from aspect ratios were considered by means of experimental studies and discussed in two recent papers.[10,11]

However, a reasonably accurate theoretical prediction of performance of the transducers, employing vibrations of variously poled tubes mechanically and acoustically loaded as a function of their aspect ratio, has not been previously presented.

The objective of this paper is to establish an analytical solution to the coupled electromechanical and mechanoacoustical problem that extends the physical insight and provides a foundation for calculating the cylindrical electroacoustic transducer parameters for various polarization states and aspect ratios of the comprising piezoceramic tubes. The solution is based on employing the energy method[12] and in-

a)Electronic mail: baronov@comcast.net

volves application of equivalent electromechanical circuits for modeling the transducers. The method permits separation of the radiation problem from the treatment of a transducer as an electromechanical system and thus simplifies the overall procedure for obtaining and interpreting a solution. In the application of the energy method it is crucial to appropriately choose the assumed modes of vibration for the mechanical system of a transducer. In the current paper vibration of a thin-walled tube is considered as the dynamical interaction of two coupled partial mechanical systems, namely, of a thin ring undergoing radial vibrations and of a thin longitudinally vibrating bar, which was first proposed by Giebe and Blechshmidt.[13] This is an alternative to the partial differential equations of motion treatment of the problem, which was introduced by Love.[2] The Giebe and Blechshmidt's solution[13] was derived in the form of Lagrange's equations and gave exactly the same results for the resonance frequencies of the finite length elastic tubes as obtained by Love's approach.[2] A common prediction of both of these analyses was the existence of a so called "dead zone," i.e., some frequency range, in which no resonance vibrations of a tube may occur. This physically improbable result is rooted in the membrane theory approximation, in which case the thickness of a tube is assumed to be small to the extent that the energy of flexural deformations can be neglected in comparison with the energy of extensional deformations at any aspect ratio. It was shown by Junger and Rosato[14] that such an assumption is especially wrong for the range of aspect ratios around the point of strongest coupling. In order to correct this shortcoming, the energy related to flexural vibration of the bar as one of the partial systems is taken into account in the current paper.

The structure of the paper is as follows. In Sec. II assumptions are made on the modes of deformation in the thin-walled elastic tubes of arbitrary aspect ratio, following the idea to represent the vibration of a tube as the coupled vibrations of two partial systems. In distinction from the membrane theory approach the flexural deformations accompanying the extensional vibrations of a tube are included. In Secs. III and IV the energy method[12] is applied to derive Lagrange's equations describing vibrations of thickness poled piezoelectric tubes, and an electromechanical circuit equivalent to the equations of motion is introduced for the most practical range of aspect ratios pertaining to the first region of strong coupling. The results of calculating the resonance frequencies, modes of vibration, and effective coupling coefficients of the tubes and their experimental verification are presented. In Sec. V the analysis is extended for the tubes that are polarized in circumferential and axial directions. Section VI deals with including the acoustic and mechanical effects on a transducer in the equivalent electromechanical circuit. Examples of the transmit and receive responses of air-backed transducers built from the tubes having different aspect ratios are presented.

It has to be noted that a terminology vagueness exists regarding a naming of the thin-walled cylindrical piezoelements that are considered in the course of treating their two-dimensional vibration. So far as the passive elastic cylindrical elements are concerned, it is common to refer to elements as "rings" at aspect ratios $h/2a \ll 1$ and as the "tubes" at $h/2a \gg 1$. But it is not clear when a ring transitions into a tube. Probably by this reason the piezoelectric elements considered in terms of two-dimensional (coupled) vibration are interchangeably referred to as cylindrical shells,[1] cylindrical tubes,[3,4] piezoelectric shells of finite length,[5] piezoceramic cylinders of finite size,[7] and ceramic cylinders.[8] At the same time, a convention exists among suppliers of piezoceramic parts to specify the thin-walled cylindrical parts with electrodes applied to the side surfaces as tubes regardless of their height-to-diameter ratio. Following this convention, we will name the objects of this treatment as the piezoceramic tubes (or just tubes, when it is clear from context that they are made from piezoelectric ceramics).

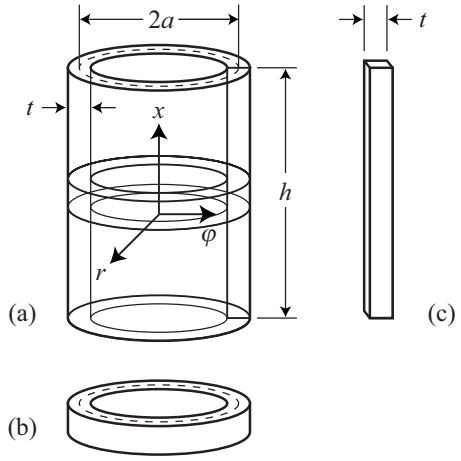


FIG. 1. A thin-walled elastic tube (a) and its partial subsystems: the radially vibrating short ring (b) and the longitudinally vibrating thin bar (c).

II. ASSUMPTIONS ON THE DISTRIBUTION OF DEFORMATIONS IN THE THIN-WALLED ELASTIC TUBES

Consider extensional axially symmetric vibrations of a thin-walled isotropic elastic tube shown in Fig. 1(a). In this treatment we will represent the vibration in the tube as the coupled vibrations of the two partial one-dimensional elastic systems, namely, of the radial vibrating ring [Fig. 1(b)] and of the longitudinally vibrating thin bar [Fig. 1(c)].

The common assumption for the thickness, t, of the "thin-walled" tube is that $t \ll 2a$. In other words this means that the resonance frequency of vibration of the ring in the direction of thickness, f_t ($f_t \doteq c/2t$, where $c=\sqrt{Y/\rho}$ is the sound speed in a thin bar, Y is Young's modulus, and ρ is the density of the tube material), and the resonance frequency of the radial vibration of the ring, f_r ($f_r=c/2\pi a$), are very far apart. Thus, these vibrations can be considered as independent, and at frequencies close to the radial resonance of the ring the thickness is very small compared with the extensional wavelength. Therefore the stresses in radial direction, T_r, are practically constant, and being zero on the ring surfaces they remain negligible inside of its volume, i.e., $T_r=0$, which allows the problem to be treated as two dimensional.

Consider the extreme cases for the height, h, of a tube, in which $h/2a \ll 1$ and $h/2a \to \infty$. In the case that $h/2a \ll 1$, the tube reduces to ring Fig. 1(b), the first resonance fre-

quency of the axial vibrations, $f_{h1} \doteq c/2h$, is much higher than the resonance frequency of the radial vibration, f_r, and by the reasons discussed for the thickness of a ring it can be assumed that the axial stress $T_x=0$ in the volume of the ring. Thus, the problem becomes one dimensional with a well-known solution. In the case that $h/2a \to \infty$, the tube becomes very long, the resonance frequency of the axial vibrations becomes much lower than for the radial vibration, and the vibrations in the vicinity of the radial resonance frequency may be considered as one dimensional with the mechanical conditions $T_r=0$ and $S_x=0$, where S_x is the strain in the axial direction. The latter condition is valid, strictly speaking, for an infinitely long tube because of the symmetry considerations, but it may be assumed to be applicable to a tube long in comparison with the extensional wavelength in a frequency range of interest. For this case the resonance frequency of the radial vibration, $f_{r\infty}$, is known to be $f_{r\infty} = \sqrt{Y/\rho(1-\sigma^2)}$, where σ is Poisson's ratio.

Both of these one-dimensional approximations are valid practically to a broad extent of aspect ratios so far as the separation between the resonance frequencies of vibration in the axial and radial directions is sufficiently large. But it remains to be estimated what is large enough in terms of an acceptable accuracy of calculations based on these approximations. From a general theory of coupled vibrations it follows that the strongest coupling between the partial systems should take place under the condition that the resonance frequencies of the partial systems are equal. In our case this condition takes place at first in the vicinity of the aspect ratio $h/2a=\pi/2$, at which point $f_r=f_{h1}$, and then repeatedly at the aspect ratios related to the harmonics of the axial resonance frequency, f_{hi}. Thus, it can be expected that the one-dimensional ring approximation may be valid for the tube with the aspect ratios $h/2a \ll \pi/2$. It is not clear what the lowest acceptable value of the aspect ratio is for the one-dimensional long tube approximation to be valid. This has to be determined.

When considering vibrations of a tube as being two dimensional, an assumption regarding the distribution of displacements over the tube surface can be made based on representation of the vibrations, as a result of interaction of the partial system vibrations. At first consider deformations of a ring [Fig. 1(b)]. Denote the radial displacement of the ring surface as ξ_0, then the strain S_φ in the circumferential direction may be presented as

$$S_\varphi = \frac{2\pi(a+\xi_a) - 2\pi a}{2\pi a} = \frac{\xi_0}{a}, \tag{1}$$

and the strain S_x in the axial direction will be determined as $S_x=-\sigma S_\varphi=-\sigma\xi_0/a$. Correspondingly, the displacement in the axial direction ξ_{xr} generated in the tube by the radial displacement will be

$$\xi_{xr} = -\xi_0 \frac{\sigma x}{a}. \tag{2}$$

Consider now deformations of a thin bar [Fig. 1(c)]. Displacement in the axial direction, ξ_{xx}, may be represented as an expansion in the series

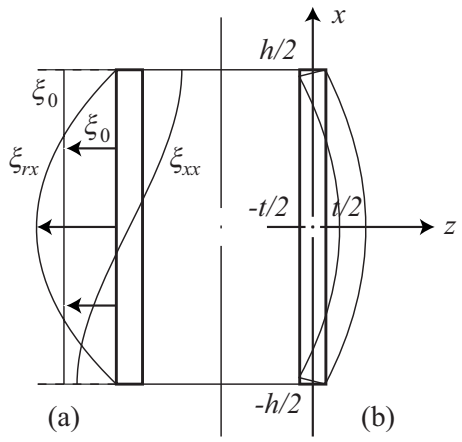


FIG. 2. Distribution of displacements in a tube: (a) in the extensional vibrations and (b) in the flexural vibrations.

$$\xi_{xx} = \sum_{i=1}^{2L-1} \xi_{xi} \sin(i\pi x/h), \tag{3}$$

where L is a number of modes taken into account for an approximation, which may be considered as acceptable. Expression for the strain in the axial direction is $S_x=\partial\xi_{xx}/\partial x$, and the strain in the lateral direction, S_φ, will be found as $S_\varphi=-\sigma S_x=-\sigma\partial\xi_{xx}/\partial x$. The deformation corresponding to this strain, which is produced in the circumferential direction of the ring, should cause a displacement of the ring surface that is proportional to deformation S_φ according to formula (1). Thus, the deformation of the bar in axial direction generates the radial displacement of the tube surface, ξ_{rx}, which can be defined as

$$\xi_{rx} \sim (\partial\xi_{xx}/\partial x) \tag{4}$$

and can be expressed with reference to Eq. (3) in the general form of

$$\xi_{rx} = \sum_{i=1}^{2L-1} \xi_{ri} \cos(i\pi x/h). \tag{5}$$

Summarizing the assumptions made above, the displacement distribution in an axial symmetrically vibrating thin-walled tube can be represented as follows:

$$\xi_x = \xi_{xx} + \xi_{xr} = (-\sigma x/a)\xi_0 + \sum_{i=1}^{2L-1} \xi_{xi} \sin(i\pi x/h), \tag{6}$$

$$\xi_r = \xi_0 + \xi_{rx} = \xi_0 + \sum_{i=1}^{2L-1} \xi_{ri} \cos(i\pi x/h). \tag{7}$$

Distribution of the displacements in a tube is shown qualitatively in Fig. 2, as a superposition of displacements generated by vibration of the partial systems (only the fundamental mode of a bar vibration is illustrated for simplicity). So far as the displacement distribution is defined by Eqs. (6) and (7), the axial symmetric strain distribution in the body of the tube can be represented in the cylindrical coordinates x and φ as follows:

$$S_x = \frac{\partial \xi_x}{\partial x} - z\frac{\partial^2 \xi_r}{\partial x^2} = -\frac{\sigma}{a}\xi_0 + \sum_{i=1}^{2L-1} \xi_{xi}\frac{i\pi}{h}\cos(i\pi x/h)$$
$$- z\sum_{i=1}^{2L-1} \xi_{ri}\left(\frac{i\pi}{h}\right)^2 \cos(i\pi x/h), \tag{8}$$

$$S_\varphi = \frac{\xi_r}{a} = \frac{\xi_0}{a} + \frac{1}{a}\sum_{i=1}^{2L-1} \xi_{ri}\cos(i\pi x/h). \tag{9}$$

The term $(-z\partial^2\xi_r/\partial x^2)$ in Eq. (8) accounts for the strains due to the flexural deformations of the wall of the tube [the coordinate axis z goes in the radial direction and has its origin on the mean circumferential surface of the tube, as it is shown in Fig. 2(b)]. This term is a matter of principle in this treatment. It takes into consideration the energy of the flexural deformations of the wall of a tube having a finite thickness, and makes the proposed approach to the problem different from that, which was used in the framework of the membrane theory.

The stresses in the tube will be found as follows[15] (remember that $T_r=0$):

$$T_x = \frac{Y}{1-\sigma^2}(S_x + \sigma S_\varphi), \tag{10}$$

$$T_\varphi = \frac{Y}{1-\sigma^2}(\sigma S_x + S_\varphi). \tag{11}$$

By substituting the strains S_x and S_φ, defined by Eqs. (8) and (9), into Eq. (10) it is easy to make certain that the boundary conditions on the free ends of a tube, $T_x(\pm h/2)=0$, are met. With distribution of strain in the tube known, the equations of motion of an isotropic passive tube can be derived as Lagrange's equations.

A note has to be made regarding a number of terms in series (3) to be taken into account for practical calculations in order to achieve an acceptable level of accuracy for the results. This depends on the range of aspect ratios under consideration. For the range below and around the first region of strong coupling, namely, $0<h/2a<3$, to the first approximation it should be sufficient to retain only the first term of the series, which corresponds to the fundamental mode of the axial vibration. For the higher range of aspect ratios, but below and around the second region of a strong coupling, that is, for $0<h/2a<6$, the next term has to be included, which corresponds to the third harmonic of the axial vibration.

One more note has to be made regarding representation of the flexural term in Eq. (8). The flexural term is represented based on the elementary theory of bending. This can be considered as sufficiently accurate until the half length of flexure-to-thickness ratio for a bar as a partial system is much larger than unity, i.e., $h/it \gg 1$, where i is a number of half waves of flexure on the length of the bar. Otherwise, corrections accounted for the effects of rotary inertia and shearing deformations on the kinetic and potential energies related to the flexural vibrations must be taken into consideration.[15] It is shown in Ref. 16 that with these corrections for rectangular bars a good agreement between calculated and measured resonance frequencies takes place at $h/it>5$, even this condition can not be met at small aspect ratios, especially for a high order flexural vibration. Thus, for example, for the tubes having the mean diameter $2a=35.8$ mm and thickness $t=3.2$ mm, which were used for experimental investigations reported in Ref. 10, this condition meets at aspect ratios $h/2a>0.5$ and $h/2a>1.5$, when $i=1$ and $i=3$, respectively. For this reason the frequency branch in Fig. 3, which corresponds to the mode of vibration at $i=3$, starts from the aspect ratio $h/2a=1.6$.

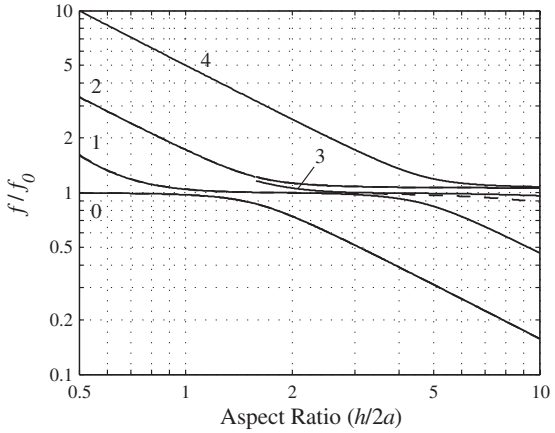


FIG. 3. The resonance frequencies of the thin-walled tube (PZT-4, $2a=35$ mm, $t=3.2$ mm) normalized by the resonance frequency of a short ring, $f_0=30$ kHz, as a function of aspect ratio $h/2a$. The dotted line shows the changes in the case that only the first approximation is used. Branch 3 starts from the aspect ratio $h/2a=1.6$, because at smaller aspect ratios calculations for the corresponding mode of vibration become inaccurate.

III. EQUATIONS OF FREE VIBRATION OF ISOTROPIC PASSIVE ELASTIC TUBES

The equations of free vibration of isotropic elastic tubes can be obtained as Lagrange's equations. The second approximation will be considered, in which case $L=2$ in Eqs. (8) and (9). The new notations will be introduced for the generalized coordinates as follows: $\xi_{r1}\rightarrow\xi_1$, $\xi_{r3}\rightarrow\xi_3$, $\xi_{x1}\rightarrow\xi_2$, and $\xi_{x3}\rightarrow\xi_4$. Thus, the five generalized coordinates, ξ_i, where $i=0,1,\ldots,4$, will be used to describe a solution to the problem. To obtain Lagrange's equations the kinetic and potential energies of a tube have to be considered.

Under the assumption that for flexural deformations the elementary theory of bending is applicable they will be determined as follows.

The kinetic energy of a tube is

$$W_{\text{kin}} = \frac{1}{2}\int_{\tilde{V}} \rho(\dot{\xi}_r^2 + \dot{\xi}_x^2)d\tilde{V} = \frac{1}{2}2\pi at\rho\int_{+h/2}^{h/2} (\dot{\xi}_r^2 - \dot{\xi}_x^2)dx. \tag{12}$$

Here $\dot{\xi}$ is the time derivative of the displacement or the velocity, and $\tilde{V}$ is the volume of the tube. The following expression for the kinetic energy will be obtained after substituting the displacements ξ_r and ξ_x from Eqs. (6) and (7) and after integrating over the height:

$$W_{\text{kin}} = \frac{1}{2}\left(\sum_{i=0}^{4} M_{ii}\dot{\xi}_i^2 + 2\sum_{i=1}^{4} M_{0i}\dot{\xi}_0\dot{\xi}_l\right), \tag{13}$$

where

$$M_{00} = M\left(1 + \frac{\sigma^2 h^2}{12a^2}\right), \quad M_{ii} = \frac{M}{2} \quad (i = 1, \ldots, 4);$$

$$M_{01} = \frac{2}{\pi}M, \quad M_{02} = -\frac{2\sigma h}{\pi^2 a}M,$$

$$M_{03} = -\frac{2}{3\pi}M, \quad M_{04} = \frac{2\sigma h}{9a\pi^2}M. \tag{14}$$

Here $M = 2\pi aht\rho$ is the mass of a tube.

The potential energy of a tube is

$$W_{\text{pot}} = \frac{1}{2}\int_{\tilde{V}} (S_x T_x + S_\varphi T_\varphi) d\tilde{V}, \tag{15}$$

Substituting expressions (8) and (9) for the strain and expressions (10) and (11) for the stress under integral (15), and integrating over the volume of the tube and a little manipulation yield

$$W_{\text{pot}} = \frac{1}{2}\left[\sum_{i=0}^{4} K_{ii}\xi_i^2 + 2(K_{01}\xi_0\xi_1 + K_{12}\xi_1\xi_2 + K_{03}\xi_0\xi_3 + K_{34}\xi_3\xi_4)\right], \tag{16}$$

where it is denoted as

$$K_{00} = \frac{2\pi thY}{a},$$

$$K_{ii} = \frac{\pi htY}{a(1-\sigma^2)}\left[1 + \frac{(i\pi)^4}{48}\left(\frac{t}{h}\right)^2\left(\frac{2a}{h}\right)^2\right] \quad \text{at } i = 1 \text{ and } 3,$$

$$K_{22} = \frac{\pi^3 atY}{h(1-\sigma^2)},$$

$$K_{44} = \frac{9\pi^3 atY}{h(1-\sigma^2)}, \quad K_{01} = \frac{4thY}{a}, \quad K_{03} = \frac{4thY}{3a}, \tag{17}$$

$$K_{12} = \frac{\pi^2 t\sigma Y}{1-\sigma^2}, \quad K_{34} = \frac{3\pi^2 t\sigma Y}{1-\sigma^2}.$$

As it was noted in Sec. II, for small aspect ratios, corrections to the flexure related parameters of the masses and rigidities must be introduced. Following Ref. 16 for the masses and rigidities at $i = 1, 3$ with the corrections accounted for the rotary inertia and shearing deformations will be obtained as

$$M_{ii} = \frac{M}{2}\left[1 + \frac{(i\pi)^2}{12}\left(\frac{t}{h}\right)^2\right], \tag{14'}$$

$$K_{ii} = \frac{\pi thY}{a(1-\sigma^2)}\left\{1 + \frac{(i\pi)^4}{48}\left(\frac{t}{h}\right)^2\left(\frac{2a}{h}\right)^2 \times\left[1 - \frac{(i\pi)^2}{10}\frac{Y}{(1-\sigma^2)\mu}\left(\frac{t}{h}\right)^2\right]\right\}, \tag{17'}$$

where μ is the shear modulus. The factors in brackets represent corrections, with which calculations are accurate in the range of aspect ratios $h/2a > 0.5$ at $i = 1$ and $h/2a > 1.5$ at $i = 3$.

Lagrange's equations of free vibration of a tube,

$$\frac{d}{dt}\left(\frac{\partial W_{\text{kin}}}{\partial \dot{\xi}_i}\right) + \frac{\partial W_{\text{pot}}}{\partial \dot{\xi}_i} = 0 \quad (i = 0, 1, \ldots, 4) \tag{18}$$

after substituting expressions (13) and (16) for the kinetic and potential energies, differentiating, and converting to the complex quantities, yield

$$(j\omega M_{\text{eqv}} + K_{\text{eqv}i}/j\omega)U_i = 0 \quad (i = 0, 1, \ldots, 4). \tag{19}$$

Here U_i is the complex amplitude of velocity $\dot{\xi}_i$, and $M_{\text{eqv}i}$ and $K_{\text{eqv}i}$ are the equivalent parameters defined as follows:

$$M_{\text{eqv0}} = M_{00} + \sum_{i=1}^{4} M_{0i}(U_i/U_0),$$

$$M_{\text{eqv1}} = M_{11} + M_{01}(U_0/U_1),$$

$$M_{\text{eqv2}} = M_{22} + M_{02}(U_0/U_2),$$

$$M_{\text{eqv3}} = M_{33} + M_{03}(U_0/U_3),$$

$$M_{\text{eqv4}} = M_{44} + M_{04}(U_0/U_4). \tag{20}$$

$$K_{\text{eqv0}} = K_{00} + K_{01}(U_1/U_0) + K_{03}(U_3/U_0),$$

$$K_{\text{eqv2}} = K_{22} + K_{12}(U_1/U_2),$$

$$K_{\text{eqv4}} = K_{44} + K_{34}(U_3/U_4),$$

$$K_{\text{eqv1}} = K_{11} + K_{01}(U_0/U_1) + K_{12}(U_2/U_1),$$

$$K_{\text{eqv3}} = K_{33} + K_{03}(U_0/U_3) + K_{34}(U_4/U_3). \tag{21}$$

After substituting the parameters thus determined into Eq. (19) these equations finally may be represented as

$$Z_{ii}U_i + \sum_{l\neq i} z_{il}U_i = 0 \quad (i, l = 0, 1, \ldots, 4), \tag{22}$$

where

$$z_{il} = z_{li}, \quad z_{13} = z_{23} = z_{14} = z_{24} = 0,$$

$$Z_{ii} = (K_{ii}/j\omega)(1 - \omega^2/\omega_{ii}^2),$$

$$z_{01} = (K_{01}/j\omega)(1 - \omega^2/\omega_{01}^2), \quad z_{02} = j\omega M_{02},$$

$$z_{03} = (K_{03}/j\omega)(1 - \omega^2/\omega_{03}^2),$$

$$z_{04} = j\omega M_{04}, \quad z_{12} = K_{12}/j\omega, \quad z_{34} = K_{34}/j\omega. \tag{23}$$

When representing impedance (23), relations (21) and (22) are taken into consideration and it is denoted as

$$\omega_{ii}^2 = K_{ii}/M_{ii}, \quad \omega_{01}^2 = K_{01}/M_{01}, \quad \omega_{03}^2 = K_{03}/M_{03}. \tag{24}$$

The frequencies ω_{ii} can be interpreted in the following way:

$$\omega_{00}^2 = K_{00}/M_{00} = \frac{Y}{a^2\rho[1+\sigma^2h^2/(12a^2)]} = \omega_{\text{ring}}^2, \tag{25}$$

$$\omega_{11}^2 = K_{11}/M_{11} = \frac{Y}{a^2\rho(1-\sigma^2)} + \frac{\pi^4t^2Y}{12h^4\rho(1-\sigma^2)} = \omega_{\text{tube}}^2 + \omega_{fh1}^2, \tag{26}$$

$$\omega_{22}^2 = K_{22}/M_{22} = \frac{\pi^2Y}{h^2\rho(1-\sigma^2)} = \omega_{h1}^2, \tag{27}$$

$$\omega_{33}^2 = K_{33}/M_{33} = \omega_{\text{tube}}^2 + \omega_{fh3}^2 = \omega_{\text{tube}}^2 + 9\omega_{fh1}^2,$$

$$\omega_{44}^2 = K_{44}/M_{44} = \omega_{h3}^2 = 9\omega_{h1}^2. \tag{28}$$

[The expressions for ω_{11} and ω_{33} are given for the aspect ratios, at which elementary theory of bending is applicable, i.e., when the corrections in Eqs. (14′) and (17′) can be neglected.] In Eqs. (25)–(28) the following notations are introduced: ω_{ring} for the resonance frequency of the radial vibration of a thin ring of the height h small compared with its radius, ω_{h1} and ω_{h3} for the fundamental resonance frequency and the third harmonics of vibration of infinite thin strip in direction of its width h, ω_{tube} for the resonance frequency of the radial vibration of a thin-walled tube of infinite height, and ω_{fh1} and ω_{fh3} for the first and third resonance frequencies of the flexural vibrations in the "width" mode of an infinite strip of thickness t simply supported on the edges. The expressions for the frequencies in Eqs. (25)–(27) may be found in Ref. 2. Equation (26) can be transformed to the form

$$\omega_{11}^2 = \omega_{\text{tube}}^2\left[1 + \frac{\pi^4}{48}(t/h)^2(2a/h)^2\right], \tag{29}$$

where from it follows that frequency ω_{11} depends on both the aspect ratio and the thickness-to-height ratio of a tube.

The set of Eq. (22) provides the solution to the problem of free coupled vibrations in a passive isotropic elastic tube. In particular, the results of calculating the resonance frequencies as the functions of aspect ratio $h/2a$ from Eq. (22) are represented in Fig. 3. They are normalized to the resonance frequency of a short ring, f_0=30 kHz. The calculations were carried out for PZT-4 ceramic tubes with the outer diameter D_o=38.2 mm and thickness t=3.2 mm, which correspond to the dimensions of tubes used in the experimental studies.[10,11] The parameters of PZT-4 were used as presented in Sec. IV. In the Fig. 3 the changes are shown by the dotted line that have to be made, if only the first approximation is taken into consideration [i.e., the generalized velocities U_3 and U_4, and all the related impedances in Eq. (22) are set to zero]. It can be concluded that at least up to the aspect ratios $h/2a \doteq 3$ the second approximation does not contribute noticeably to values of frequencies calculated using just the first approximation. Therefore, further analysis will be restricted by the first approximation given that the range of aspect ratios $h/2a < 3$ is the most interesting for practical applications and the restriction simplifies analysis for this range without loss of accuracy. Besides, the exactly same technique can be used to analyze the solution at larger values of the aspect ratios in case this is needed. The part of the plot in Fig. 3 related to the first approximation only is displayed in Fig. 4 up to values $h/2a$=3 in a larger scale together with results of calculations made for different thicknesses of the wall of the tube and with experimental data included. The results of experimental investigation we refer to throughout this paper are reported in Refs. 10 and 11, where an experimental setup and measurement procedures are described in detail.

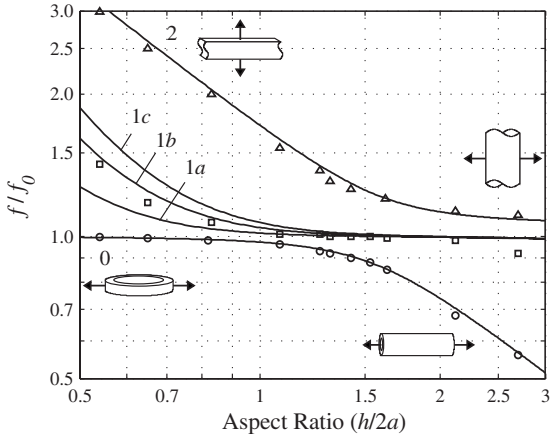


FIG. 4. The normalized resonance frequencies of the tube (PZT-4, $2a$ =35 mm) calculated to the first approximation at different wall thicknesses t: 2.0 mm (1a), 3.2 mm (1b), and 4.0 mm (1c). Experimental data from Ref. 10 are shown by circles for branch 0, by squares for branch 1, and by diamonds for branch 2.

The notable difference between the frequency plots displayed in Fig. 3 and those presented in Ref. 1 is in existence of the flexure related branches 1 and 3, which could not be predicted by the membrane theory and which cross the so called dead zone. It is of note that for these branches the normalized numerical values are valid for the particular thickness t=3.2 mm only, whereas the extensional vibration related frequency branches do not depend on the thickness. The results of calculations made for the tubes with different thicknesses presented in Fig. 4 show significant dependence of the "intermediate" frequency branch 1 from the thickness at aspect ratios up to values in the vicinity of the point of the strongest coupling. This is indicative of a flexural nature of the corresponding modes of vibration in this range of aspect ratios. Above the point of the strongest coupling this frequency branch becomes independent of thickness as the extensional branches do. This is in accord with the results of observations made in Ref. 10.

IV. EQUATIONS OF MOTION OF A PIEZOCERAMIC TUBE AS AN ELECTROMECHANICAL TRANSDUCER (THICKNESS POLARIZATION)

Consider now the vibration problem in the case that the tubes are made from piezoelectric ceramics and constitute an electromechanical system. In this section a piezoceramic

tube will be considered as an electromechanical transducer without any external load applied (vibrating in air). Following the procedure of application of the energy method to deriving the equations of motion of a piezoelectric body,[12] the energies characterizing its status under assumed distribution of deformation in the body have to be calculated. Namely, the mechanical energy of deformation at electric field $E_3=0$, W_m^E; the electrical energy W_{el}^S, which would be stored in the transducer, if the acting generalized coordinates were clamped (under the assumption that $S_1=S_2=0$ in the case of the thickness poled thin-walled tubes); the electromechanical energy W_{em}, as a part of the electrical energy applied, which is converted into energy of deformation of the transducer's mechanical system with rigidity, determined under assumption that the electrical field is set to zero ($E_3=0$). The two latter energy terms are related to the total electrical energy supplied to a transducer, W_{el}, by Eq. (38) in Ref. 17, as follows:

$$W_{\text{el}} = W_{\text{el}}^S + W_{\text{em}}. \tag{30}$$

When calculating all of the energies needed, the piezoelectric equations of state have to be used together with the assumed distributions of deformation in the tubes. The results of calculation depend on how the piezoceramic tubes are oriented relative to a crystallographic coordinate system. The most widespread are the thickness poled tubes especially among those having relatively large aspect ratio. Therefore a detailed analysis will be made for this particular case. The changes that have to be made, if the poling axis is directed circumferentially or axially, will be considered later.

Following the common notations for the crystallographic axes with poling axis denoted as 3, in the case of thickness polarization we have $S_x=S_1$, $S_\varphi=S_2$, and $T_r=T_3$. Remembering that $T_r=T_3=0$, the piezoelectric equations simplify to the following form:[18]

$$S_1 = s_{11}^E T_1 + s_{12}^E T_2 + d_{31}E_3, \tag{31}$$

$$S_2 = s_{12}^E T_1 + s_{11}^E T_2 + d_{31}E_3, \tag{32}$$

$$D_3 = d_{31}(T_1 + T_2) + \varepsilon_{33}^T E_3. \tag{33}$$

Substituting the stresses T_1 and T_2 from Eqs. (31) and (32) into Eq. (33) yields

$$D_3 = D_e^{S_{1,2}} + D_{\text{em}}(S_1, S_2), \tag{34}$$

where the components of charge density are introduced as follows:

$$D_e^{S_{1,3}} = \varepsilon_{33}^{S_{1,3}} E_3, \tag{35}$$

where $\varepsilon_{33}^{S_{1,2}}=\varepsilon_{33}^T(1-k_p^2)$ and $k_p^2=2d_{31}^2/[\varepsilon_{33}^T s_{11}^E(1+s_{12}^E/s_{11}^E)]$ are the dielectric constant and the planar coupling coefficient square of a piezoceramic material in the clamped tube (subscripts 1 and 2 indicate that deformations S_1 and S_2 are set to zero);

$$D_{\text{em}}(S_1, S_2) = \frac{d_{31}}{s_{11}^E + s_{12}^E}(S_1 + S_2) \tag{36}$$

is the charge density induced by deformations.

The stresses in a tube at $E_3=0$ will be found from Eqs. (31) and (32) as

$$T_1^E = \frac{Y_1^E}{1-(\sigma_1^E)^2}(S_1 + \sigma_1^E S_2), \tag{37}$$

$$T_2^E = \frac{Y_1^E}{1-(\sigma_1^E)^2}(\sigma_1^E S_1 + S_2). \tag{38}$$

In Eqs. (37) and (38) it is denoted as

$$Y_1^E = 1/s_{11}^E, \quad \sigma_1^E = -s_{12}^E/s_{11}^E. \tag{39}$$

Equations (37) and (38) are identical with Eqs. (10) and (11), if to replace Y_1^E and σ_1^E for the previous Y and σ. Therefore all the results obtained for the isotropic elastic tubes in terms of their mechanical behavior will be completely applicable to the thin-walled thickness poled piezoceramic tubes and vice versa. Thus, the expressions for all of the equivalent parameters, impedances, and resonance frequencies introduced in Sec. III are valid for piezoceramic tubes upon substituting Y_1^E and σ_1^E for Y and σ. In order to distinguish the mechanical quantities in the case that a piezoelectric material is involved, the notations W_m^E instead of W_{pot} and K_{il}^E instead of K_{il} will be used. The expressions for the rigidities related to the first approximation (at i, $l=0,1,2$) are presented in Table I.

After multiplying both parts of Eq. (35) by E_3, integrating over the tube volume, and comparing the corresponding terms of resulting equation and Eq. (30), the following expressions for energies W_{el}^S and W_{em} will be obtained.

For the electrical energy of a clamped tube,

$$W_{\text{el}}^{S_{1,2}} = \frac{1}{2}\int_{\tilde{V}} D_e^{S_{1,2}} E_3 d\tilde{V} = \frac{1}{2}\int_{\tilde{V}} \varepsilon_{33}^{S_{1,2}} E_3^2 d\tilde{V} = \frac{1}{2} C_e^{S_{1,2}} v^2. \tag{40}$$

Here the electric capacitance of a clamped tube is denoted as

$$C_e^{S_{1,2}} = 2\pi a h \varepsilon_{33}^{S_{1,2}}/t. \tag{41}$$

For the electromechanical energy,

$$W_{\text{em}} = \frac{1}{2}\int_{\tilde{V}} D_{\text{em}}(S_1, S_2) E_3 d\tilde{V} = \frac{1}{2}\frac{d_{31}}{s_{11}^E + s_{12}^E}\int_{\tilde{V}} (S_1 + S_2) E_3 d\tilde{V}. \tag{42}$$

Upon substituting $E_3=v/t$, where v is an instantaneous value of the voltage applied to the electrodes, and expressions (8) and (9) for the deformations $S_1=S_x$ and $S_2=S_\varphi$, and integrating over the volume of a tube, the electromechanical energy can be expressed as

$$W_{\text{em}} = \frac{v}{2}\sum_{i=0}^{2} n_i \xi_i, \tag{43}$$

where it is denoted as

$$n_0 = \frac{2\pi h d_{31}}{s_{11}^E}, \quad n_1 = \frac{4 h d_{31}}{s_{11}^E(1-\sigma_1^E)}, \quad n_2 = \frac{4\pi a d_{31}}{s_{11}^E(1-\sigma_1^E)}. \tag{44}$$

TABLE I. Rigidity parameters for tubes with different polarizations.

	K_{il}^E				
Polarization	K_{00}^E, $\times 2\pi th/a$	K_{11}^E, $\times \pi th/a$	K_{22}^E, $\times \pi^3 at/h$	K_{01}^E, $\times 4th/a$	K_{12}^E, $\times \pi^2 t$
Thickness (radial)	Y_1^E	$\dfrac{Y_1^E\left[1+\dfrac{\pi^4}{48}\left(\dfrac{t}{h}\right)^2\left(\dfrac{2a}{h}\right)^2\right]}{1-(\sigma_1^E)^2}$	$\dfrac{Y_1^E}{1-(\sigma_3^E)^2}$	Y_1^E	$\dfrac{\sigma_1^E Y_1^E}{1-(\sigma_1^E)^2}$
Circumferential	Y_3^E	$\dfrac{Y_3^E\left[1+\dfrac{\pi^4}{48}\left(\dfrac{t}{h}\right)^2\left(\dfrac{2a}{h}\right)^2\dfrac{Y_1^E}{Y_3^E}\right]}{1-\sigma_3^E\sigma_{13}^E}$	$\dfrac{Y_1^E}{1-\sigma_3^E\sigma_{13}^E}$	Y_3^E	$\dfrac{\sigma_3^E Y_1^E}{1-\sigma_3^E\sigma_{13}^E}$
Axial	Y_1^E	$\dfrac{Y_1^E\left[1+\dfrac{\pi^4}{48}\left(\dfrac{t}{h}\right)^2\left(\dfrac{2a}{h}\right)^2\dfrac{Y_3^E}{Y_1^E}\right]}{1-\sigma_3^E\sigma_{13}^E}$	$\dfrac{Y_3^E}{1-\sigma_3^E\sigma_{13}^E}$	Y_1^E	$\dfrac{\sigma_3^E Y_1^E}{1-\sigma_3^E\sigma_{13}^E}$

A set of equations of motion of a piezoceramic tube may be represented in the general form of Eqs. (39) and (41) from Ref. 12, namely,

$$(j\omega M_{\mathrm{eqv}i} + K_{mi}^E/j\omega + r_{\mathrm{nl},i} + Z_{\mathrm{ac}i})U_i = Vn_i \quad (i=0,1,2), \tag{45}$$

$$I = \left(j\omega C_e^{S_{1,2}} + \frac{1}{R_{\mathrm{eL}}}\right)V + \sum_{i=1}^{2} U_i n_i. \tag{46}$$

(Note that the capital letters V and U stand for the complex magnitudes of the voltage v and velocity $u=\dot{\xi}$.) All the constants in Eqs. (45) and (46) are already obtained for the particular case of the thickness poled tube except for the radiation impedance, $Z_{\mathrm{ac}i}$, and for the dissipative terms represented as resistances of the electrical and mechanical losses R_{eL} and r_{mL}. In this section a transducer is considered as electromechanical without any load; therefore, the acoustical radiation impedances, $Z_{\mathrm{ac}i}$, in Eq. (45) have to be set to zero. The resistance of electrical losses is usually represented as

$$R_{\mathrm{eL}} = 1/\omega C_e^{S_{1,2}} \tan\delta_e, \tag{47}$$

where $\tan\delta_e$ is the dielectric loss tangent of a ceramic material.

A special note has to be made as to resistances of the mechanical losses, $r_{\mathrm{mL}i}$. Using the common assumption that the dissipative forces in elastic bodies are proportional to the velocity, the power of dissipation can be represented in the complex form as $\bar{\dot{W}}_{\mathrm{mL}} = r_{\mathrm{mL}}|U|^2$. From the definition of the mechanical quality factor, Q_m, as a ratio of the reactive elastic power to the power of dissipation, it follows that $\bar{\dot{W}}_{\mathrm{mL}} = |\bar{\dot{W}}_m^E|/Q_m$. The reactive elastic power may be obtained from Eq. (16), and finally we arrive at

$$\bar{\dot{W}}_{\mathrm{mL}} = \frac{1}{Q_m\omega}\left[\sum_0^2 K_{ii}^E|U_i|^2 + K_{01}^E(U_0U_1^* + U_1U_0^*) + K_{12}^E(U_1U_2^* + U_2U_1^*)\right]. \tag{48}$$

From Eq. (48) it follows that each rigidity related term has to be accompanied by a resistance of losses and they have to be correlated as $r_{\mathrm{mL}\ il} = K_{m\ il}^E/Q_m\omega$. Therefore the equivalent resistance of the mechanical losses corresponding to a generalized velocity U_i in Eq. (45) is

$$r_{\mathrm{ml},i} = K_{mi}^E/Q_m\omega, \tag{49}$$

where K_{mi}^E are given by expression (21) (for the passive tubes they were denoted as $K_{\mathrm{eqv}i}$).

The set of Eqs. (45) and (46) solves the problem of calculating a piezoceramic tube as an electromechanical transducer. Equation (45) is the equation of forced vibrations of a tube, which becomes Eq. (22) of free vibrations under the conditions that applied voltage is zero ($V=0$) and the resistances of the mechanical losses are neglected (the formal difference also is that notations for the rigidities and constants Y and σ are changed, as it was noted). The resonance frequencies of a piezoceramic tube may be found from Eq. (22), if to substitute all the impedances Z_{il} by Z_{il}^E. The results of calculating the resonance frequencies presented in Figs. 3 and 4 were obtained under exactly this condition. Thus, they are valid for the thickness poled PZT-4 tubes.

Returning to the set of Eqs. (45) and (46), we note that these equations may be considered as Kirchhoff's equations for the circuit that is shown in Fig. 5(a) for the particular case of an unloaded transducer. This circuit is equivalent to the set of Eqs. (45) and (46) in terms of calculating the electromechanical parameters of a transducer and velocity distribution over its surface. The mechanical branches of the equivalent circuit are coupled. The coupling is introduced through the mutual impedances z_{01}, z_{02}, and z_{12} [see Eq. (23)] as follows:

$$Z_{01} = z_{01}(U_1/U_0), \quad Z_{10} = z_{01}(U_0/U_1),$$

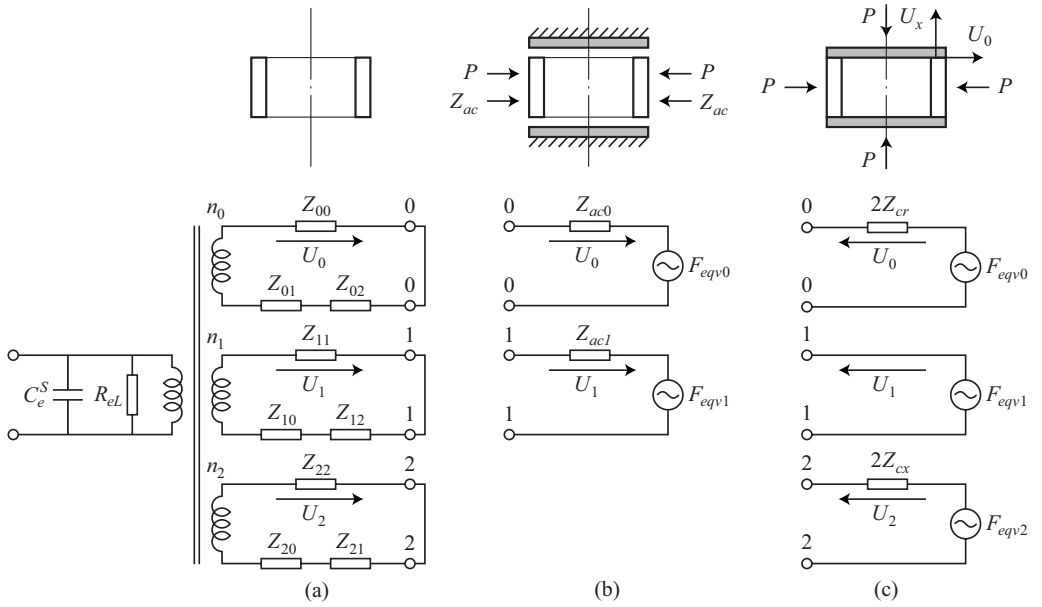


FIG. 5. The equivalent electromechanical circuit of a transducer made from a piezoceramic tube undergoing the two-dimensional vibration for different transducer modifications: (a) with free ends and without an acoustical load applied to the side surface, (b) with the ends shielded and acoustical load applied to the side surface, and (c) with the caps attached to the ends and sound pressure acting all over the surface (the case typical for hydrophones). In cases (b) and (c) the one-port circuits have to be connected to the corresponding terminals of the circuit (a).

$$Z_{02} = z_{02}(U_2/U_0), \quad Z_{20} = z_{02}(U_0/U_2),$$

$$Z_{12} = z_{12}(U_2/U_1), \quad Z_{21} = z_{12}(U_1/U_2). \tag{50}$$

After the generalized velocities are calculated and the admittance of a transducer between input terminals of the equivalent circuit in Fig. 5(a) is obtained, the calculation of a cylindrical thickness poled transducer as an unloaded electromechanical device may be considered completed. The calculated and measured values of modulus of admittance $|Y|$ are shown in Fig. 6 for the tube that we consider in the examples throughout the paper, at aspect ratio $h/2a=1.1$.

The distribution of radial velocity over surface, $U_r(x)$, and the magnitude of vibration of the ends of a tube, $U_x(\pm h/2)$, are of a great interest for a better understanding of the mechanism of coupled vibration in the tubes and for calculating or predicting the acoustic field related parameters of a transducer. It follows from Eqs. (6) and (7) rewritten in the complex form that

$$U_r(x) = U_0 + U_1 \cos(\pi x/h), \tag{51}$$

$$U_x(h/2) = -(\sigma_1^E h/2a)U_0 + U_2. \tag{52}$$

The mode shape of the surface vibration defined as the distribution of radial velocity normalized to its value at $x=0$ is

$$\theta_r(x) = \frac{1}{U_0 + U_1}\left(U_0 + U_1 \cos\frac{\pi x}{h}\right). \tag{53}$$

The calculated mode shapes corresponding to the resonance frequencies of a tube at $h/2a=1.1$ and the results of their experimental verification are shown in Fig. 7. Evidently the mode shape related to branch 1 is typical of the flexural vibration of a bar with free ends at its lowest resonance frequency. The calculated and experimentally measured dependences of the ratio of magnitudes of velocity in the radial direction at $x=0$ to velocity of the ends in the axial direction, $U_r(x=0)/U_x(x=\pm h/2)$, as a function of aspect ratio are shown in Fig. 8. The dependences clearly show how the radial component of vibration, being predominant at lower frequency branch 0 below the point of the strongest coupling, becomes predominant at the upper branch 2 above this point. A very important feature resulting from calculations and confirmed experimentally is that U_r and U_x are in antiphase, when related to branch 0, and in phase, when related to branch 2 (the displacements leading to expansion are conventionally considered as positive). This fact, which was also pointed out in Ref. 6, explains peculiarities in behavior of the effective coupling coefficients, pertaining to the frequency branches.

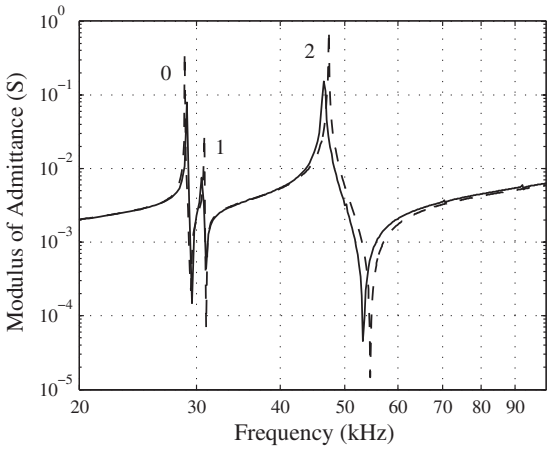


FIG. 6. Admittances for $h/2a=1.1$ measured (solid line) and calculated (dashed line).

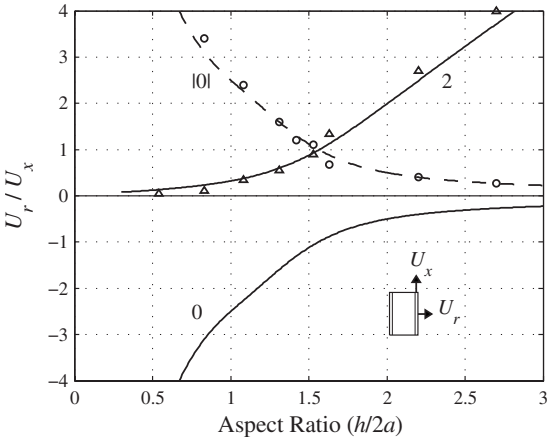


FIG. 7. The resonance mode shapes of a tube ($2a=35$ mm, $t=3.2$ mm) at $h/2a=1.1$: $f_0=28.8$ kHz at branch 0 and $f_1=30.8$ kHz at branch 1. Calculated mode shapes are shown by lines, whereas experimental data from Ref. 10 are shown by circles and squares.

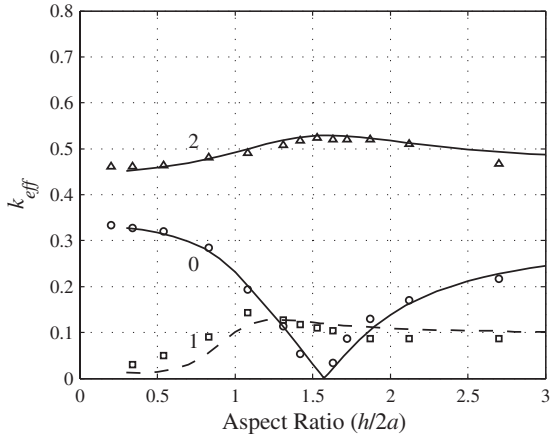


FIG. 9. Effective coupling coefficients as a function of aspect ratio along the frequency branches 0, 1, and 2 (the plots are labeled, respectively). The numerical values for k_{eff1} depend on the wall thickness and are valid for $t=3.2$ mm. Experimental data from Ref. 10 are shown as markers.

The effective coupling coefficients corresponding to the resonance mode shapes of a tube can be calculated based on analysis of the input admittance, namely, by the formula $k_{\text{eff}}^2=1-(f_r/f_{ar})^2$, where f_r and f_{ar} are the resonance and antiresonance frequencies. For example, the coupling coefficients at $h/2a=1.1$ can be obtained from the data presented in Fig. 6. It is of note that, while being the most common, this method is valid for the piezoelements having only one mechanical degree of freedom, and the accuracy of its application to determining the effective coupling coefficients of real systems depends on how far their resonances are from each other. An alternative way of calculating the effective coupling coefficients that does not have this disadvantage is by taking into consideration the energies associated with the resonance modes of vibration by formula $k_{\text{eff}}^2/(1-k_{\text{eff}}^2)=W_{\text{em}}^2/(W_{\text{el}}^S W_m^E)$, which can be obtained following Ref. 19 [Eq. (27)]. All the energies involved in the formula are determined above.

The results of calculating the effective coupling coefficients for the thickness poled tubes of different aspect ratios are shown in Fig. 9 together with the experimental data.[10]

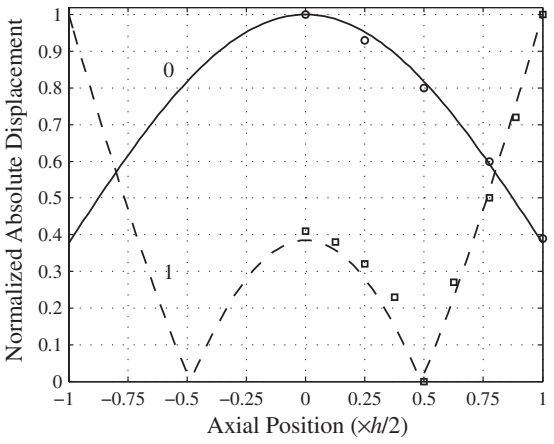


FIG. 8. The ratio of the magnitudes of vibration in the radial and axial directions $[U_r(x=0)/U_x(\pm h/2)]$ vs $h/2a$ along branches 0 and 2. The modulus of $[U_r/U_x]_0$ is shown by a dashed line.

For the extensional branches 0 and 2 the results do not depend on the wall thickness (so far as the thin-walled assumption remains applicable). For the flexure related branch 1 the results are numerically valid for the particular tube thickness, for which the calculations are made, at aspect ratios below the strongest coupling point. Above this point they become thickness independent because the nature of vibration gradually changes to the extensional, as it was pointed out. Given the facts that the thin-walled piezoceramic tube is electromechanically isotropic and that deformations in the circumferential and axial directions are in antiphase along branch 0 and in phase along branch 2, it could be expected that electromechanical effects being subtracted at branch 0 and adding up at branch 2 will result in k_{eff0} dropping to zero and k_{eff2} raising to maximum at the point of strongest coupling, where the magnitudes of deformation in the circumferential and axial directions become equal. It is of note that although at this point (at $h/2a=1.57$) k_{eff2} has maximum, the electromechanical energy at this point is equally distributed between vibration in the radial and axial directions, whereas the operational direction is radial. At larger aspect ratios k_{eff2} decreases slightly, but most of the electromechanical energy goes for radial vibration, which makes this range of aspect ratios advantageous for use of piezoceramic tubes as single cylindrical projectors.

V. PIEZOCERAMIC TUBES POLED IN THE AXIAL AND CIRCUMFERENTIAL DIRECTIONS

Consider now the changes that have to be made in calculating the transducer parameters in the case that the piezoceramic tubes are poled in the circumferential or in the axial directions. Orientations of the axis of the crystallographic coordinate system in these cases correlate with the geometrical axis shown in Fig. 1 as follows: for the circumferential polarization $x\to 1$, $r\to 2$, and $\varphi\to 3$; for the axial polarization $x\to 3$, $r\to 2$, and $\varphi\to 1$.

In the case of the circumferential (tangential) polarization we will assume that electrodes are embedded into the body of a tube, as it is shown in Fig. 10(a), and the electric field is uniform and may be calculated as $E_3=V/\delta_c$, where δ_c

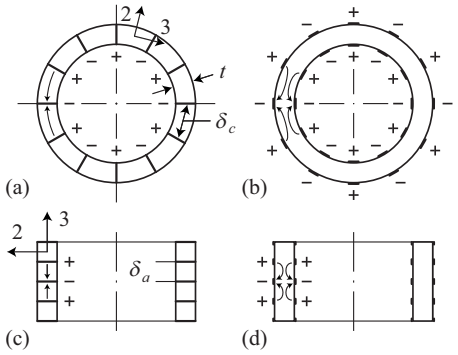


FIG. 10. The variants of electrode configuration: [(a) and (b)] for the circumferential polarization and [(c) and (d)] for the axial polarization and the corresponding crystallographic coordinate system orientation.

is the separation between electrodes (the segments are supposed to be electrically connected in parallel). However, in reality, the stripped electrodes are usually used to tangentially polarize the thin-walled tubes, as shown in Fig. 10(b). In this case the electric field, strictly speaking, cannot be considered as uniform and it is hard to make quantitatively accurate calculations, but qualitatively the results should be similar to those for the segmented design.

In the case of the axial polarization, which can be of interest in the application of transducers operating in the axial direction, the piezoelement design can be imagined, as illustrated in Figs. 10(c) and 10(d), namely, segmented tubular stack cemented from a number of end-electroded rings [Fig. 10(c)] and a tube tangentially poled in axial direction by using stripped electrodes in the shape of the rings [Fig. 10(d)]. We will consider the segmented tube design, in which case the electric field may be assumed to be uniform with magnitude $E_3=V/\delta_a$, where δ_a is the separation between electrodes.

The mechanical boundary condition $T_2=0$ is the same for circumferential and axial polarizations. Therefore for both cases the following piezoelectric equations are valid:

$$S_1 = s_{11}^E T_1 + s_{13}^E T_3 + d_{31} E_3, \tag{54}$$

$$S_3 = s_{13}^E T_1 + s_{33}^E T_3 + d_{33} E_3, \tag{55}$$

$$D_3 = d_{31} T_1 + d_{33} T_3 + \varepsilon_{33}^T E_3. \tag{56}$$

The expressions for stresses T_1 and T_3 will be found from Eqs. (54) and (55) as follows:

$$T_1 = \frac{Y_1^E}{1-\sigma_3^E\sigma_{13}^E}(S_1 + \sigma_3^E S_3) - \frac{Y_1^E}{1-\sigma_3^E\sigma_{13}^E}(d_{31} + \sigma_3^E d_{33})E_3, \tag{57}$$

$$T_3 = \frac{Y_3^E}{1-\sigma_3^E\sigma_{13}^E}(\sigma_{13}^E S_1 + S_3) - \frac{Y_3^E}{1-\sigma_3^E\sigma_{13}^E}(d_{33} + \sigma_{13}^E d_{31})E_3. \tag{58}$$

In Eqs. (57) and (58) it is denoted as

$$Y_3^E = 1/s_{33}^E, \quad \sigma_3^E = -s_{13}^E/s_{33}^E, \quad \sigma_{13}^E = -s_{13}^E/s_{11}^E. \tag{59}$$

The mechanical energy of a tube at $E_3=0$ following Eqs. (15) and (16) can be represented as

$$W_m^E = \frac{1}{2}\int_{\tilde{V}} (S_1 T_1^E + S_3 T_3^E) d\tilde{V} = \frac{1}{2}\left[\sum_{i=0}^{2} K_{ii}^E \xi_i^2 + 2(K_{02}^E \xi_0 \xi_2 + K_{12}^E \xi_1 \xi_2)\right]. \tag{60}$$

Substituting T_1 and T_3 into Eq. (56) produces the following expressions for the components of the charge density analogous to those in Eq. (34): $D_e^{S_{1,3}} = \varepsilon_{33}^{S_{1,3}} E_3$, where

$$\varepsilon_{33}^{S_{1,3}} = \varepsilon_{33}^T (1 - k_{S_{1,3}}^2) \quad \text{and}$$

$$k_{S_{1,3}}^2 = (k_{31}^2 + k_{33}^2 - 2|k_{31}|k_{33}\sqrt{\sigma_3^E \sigma_{13}^E})/(1-\sigma_3^E\sigma_{13}^E) \tag{61}$$

are the dielectric constant and coupling coefficient squared of ceramic material under the mechanical boundary conditions $S_1=S_3=0$, $T_2=0$;

$$D_{\text{em}}(S_1, S_3) = \frac{d_{31} Y_1^E}{1-\sigma_3^E\sigma_{13}^E}(S_1 + \sigma_3^E S_3) + \frac{d_{33} Y_3^E}{1-\sigma_3^E\sigma_{13}^E}(\sigma_{13}^E S_1 + S_3). \tag{62}$$

Analogous to Eqs. (40) and (42) the energies W_{el}^S and W_{em} can be represented as

$$W_{\text{el}}^{S_{1,3}} = \frac{1}{2}\int_{\tilde{V}} D_e^{S_{1,3}} E_3 d\tilde{V} = \frac{1}{2}\int_{\tilde{V}} \varepsilon_{33}^{S_{1,3}} E_3^2 d\tilde{V} = \frac{1}{2} C_e^{S_{1,3}} v^2, \tag{63}$$

$$W_{\text{em}}(S_1, S_3) = \frac{1}{2}\int_{\tilde{V}} D_{\text{em}}(S_1, S_3) E_3 d\tilde{V} = \frac{v}{2}\sum_{i=0}^{2} n_i \xi_i. \tag{64}$$

A difference in the results of calculating the energies involved in the electromechanical conversion for the cases of circumferential and axial polarizations arises from the different relative orientation of the geometrical and crystallographic axes. Namely, in the case of the circumferential polarization $S_1=S_x$, $S_3=S_\varphi$ and in the case of the axial polarization $S_1=S_\varphi$, $S_3=S_x$. In both cases the strains S_x and S_φ are given by Eqs. (8) and (9), respectively. Hence, the calculations of the energies and of the corresponding equivalent parameters for different modes of polarization must be fulfilled separately.

A. Circumferential polarization

After substituting the strains S_x and S_φ, given by Eqs. (8) and (9), for S_1 and S_3 in Eq. (57) at $E_3=0$, it will be found from the condition $T_x(\pm h/2)=T_1^E(\pm h/2)=0$ that in this case σ in Eqs. (8) and (16) should be replaced with σ_3^E. Upon substituting the stresses T_1^E and T_3^E from Eqs. (57) and (58) and the strains S_x and S_φ instead of S_1 and S_3 under the integral in Eq. (60), and after using the same procedure, as was used in the case of the thickness polarization, the expressions for the rigidities K_{il}^E will be obtained as presented in

TABLE II. Electromechanical transformation coefficients for tubes with different polarizations.

Polarization	n		
	n_0, $\times 2\pi h$	n_1, $\times 4h$	n_2, $\times 4\pi a$
Thickness (radial)	$\frac{d_{31}}{s_{11}^E}$	$\frac{d_{31}}{s_{11}^E+s_{12}^E}$	$\frac{d_{31}}{s_{11}^E+s_{12}^E}$
Circumferential, $\times t/\delta_c$	$\frac{d_{33}}{s_{33}^E}$	$\frac{d_{33}}{s_{33}^E}\left[1-\frac{d_{31}s_{13}^E}{d_{33}s_{11}^E}\right]$	$\frac{d_{31}}{s_{11}^E}\left[1-\frac{d_{33}s_{13}^E}{d_{31}s_{33}^E}\right]$
Axial, $\times t/\delta_a$	$\frac{d_{31}}{s_{11}^E}$	$\frac{d_{31}}{s_{11}^E}\left[1+\frac{d_{33}s_{13}^E}{d_{31}s_{33}^E}\right]$	$\frac{d_{33}}{s_{33}^E}\left[1-\frac{d_{31}s_{13}^E}{d_{33}s_{11}^E}\right]$

Table I. The capacitance of a tube at $S_1=S_3=0$ and $T_2=0$ and under the condition that all the segments are connected in parallel will be found from Eq. (63) as

$$C_e^{S_{1,3}} = 2\pi aht\varepsilon_{33}^{S_{1,3}}/\delta_c^2, \tag{65}$$

where $\varepsilon_{33}^{S_{1,3}}$ is given by formula (61). Expressions for the electromechanical transformation coefficients will be obtained after substituting under the integral in Eq. (64) relation for $D_{\text{em}}(S_1,S_3)$ from Eq. (62), the strains $S_1=S_x$ and $S_3=S_\varphi$ and the electric field $E_3=V/\delta_c$. The resulting expressions are presented in Table II.

B. Axial polarization

In the case of the axial polarization from the boundary condition $T_x(\pm h/2)=T_3^E(\pm h/2)=0$ it follows that σ should be replaced with σ_{13}^E in Eqs. (8) and (14). All the other parameters, such as the rigidities, electromechanical transformation coefficients, and blocked capacitance, may be determined in exactly the same way as it is done for the circumferential polarization. The main difference is that in this case S_3 should be replaced with S_x and S_1 by S_φ. Also peculiarity exists in calculating the equivalent rigidities. It is shown in Ref. 17 that, if the strains change in the direction of electric field (and in the case of the axial polarization both S_x and S_φ change along the electric field), an additional mechanical energy related term ΔW must be considered, resulting in an increase in the rigidity of a mechanical system. From Ref. 17 it follows that, if $\delta_a=h$, this effect can be accounted for by introducing the rigidity $\Delta K=0.2K_{22}^E k_{S_{1,3}}^2/(1-k_{S_{1,3}}^2)$ in addition to K_{22}^E. For PZT-4 $k_{S_{1,3}}^2\approx 0.4$ and $\Delta K\approx 0.13K_{22}^E$. However, when $h/\delta_a>4$ (and this is usually the case for a range of aspect ratios, for which the effects of the coupled vibrations are significant), the additional rigidities drop and become negligible. Therefore, for simplicity, the corresponding corrections will not be included. The resulting expressions for the equivalent rigidities are presented in Table I.

The expressions for the electromechanical transformation coefficients, obtained after substituting the electric field $E_3=V/\delta_a$ and the strains S_x and S_φ instead of S_3 and S_1, respectively, into Eqs. (62) and (64), are presented in Table II.

The capacitance of a tube at $S_1=S_3=0$ and $T_2=0$ and under the condition that all the segments are connected in parallel is

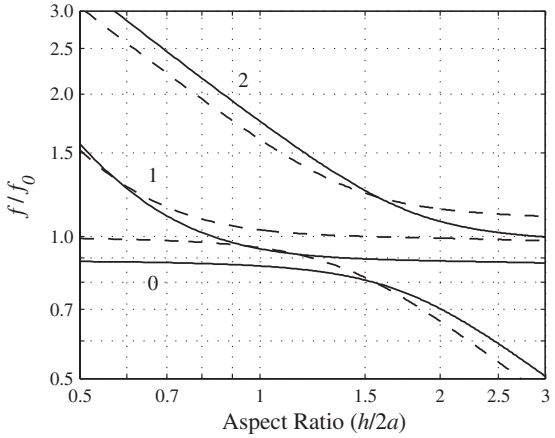


FIG. 11. The resonance frequencies of the axially (dashed lines) and circumferentially (solid lines) poled tubes normalized by $f_0=30$ kHz. Labeling 0, 1, and 2 of the curves corresponds with numbering of the frequency branches.

$$C_e^{S_{1,3}} = 2\pi a\varepsilon_{33}^{S_{1,3}}t/\delta_a^2. \tag{66}$$

Once the equivalent electromechanical parameters of the circumferentially and axially poled tubes are determined, all the necessary characteristics can be calculated in the same way as it was done in the case of the thickness polarization. In particular, the results of calculating the resonance frequencies, radial-to-axial velocities' ratio, and effective coupling coefficients for the tubes, having the same dimensions as in the case of the thickness polarization, are depicted in Figs. 11–13. The plots for the resonance frequencies are normalized to the same values as in the case of the thickness polarization, i.e., to the resonance frequency of a short thickness poled tube, $f_0=30$ kHz. The numerical values of the normalized resonance frequencies versus aspect ratio are presented in Table III.

The results show clear distinctions due to different modes of polarization and consequent elastic and piezoelectric anisotropies of the tubes. For example, it can be seen that in terms of a qualitative behavior the plots for k_{eff} related to branches 0 and 2 changed places compared with the case of the thickness polarization. This effect could be expected by the following reason. In the cases of the axial and circumfer-

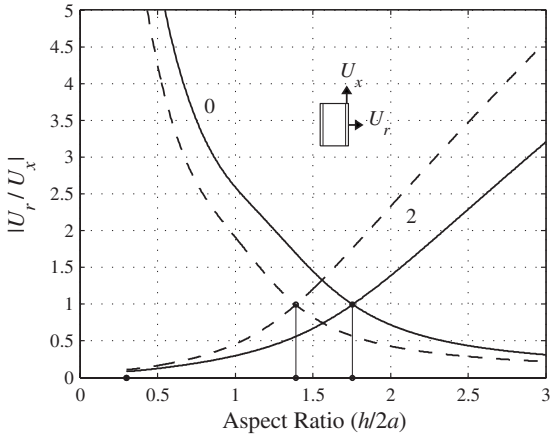


FIG. 12. Magnitude of radial-to-axial velocity ratios, $[U_r/U_x]_i$, for the axially (dashed curves) and circumferentially (solid curves) poled tubes.

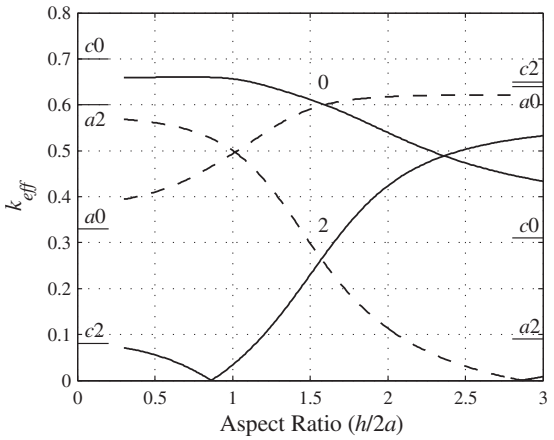


FIG. 13. Effective coupling coefficients for the axially (dashed curves) and circumferentially (solid curves) poled tubes. The horizontal markers denote the asymptotic limits of the curves as $h/2a \to 0$ and $h/2a \to \infty$.

ential polarizations the piezoelectric moduli effective in the axial and circumferential directions (d_{31} and d_{33}) have different signs, whereas in the case of the thickness polarization they (both being d_{31}) are of the same sign. Therefore the electromechanical effects are subtracted along branch 2 and adding up along branch 0 on the contrary to the case of the thickness polarization.

The results of calculating the effective coupling coefficients comply with their expected values for the extreme aspect ratios at $h/2a \to 0$ and $h/2a \to \infty$, in which cases the one-dimensional approximations for the corresponding piezoelements are valid. Expressions for the coupling coefficients of piezoceramic material for the mechanical boundary conditions corresponding to the extreme cases of $h/2a \to 0$ [a low pulsating tube and an infinitely long strip vibrating in the direction of its width with $\cos(\pi x/h)$ distribution] and $h/2a \to \infty$ [a long pulsating tube and a long tube of small diameter axially vibrating with $\cos(\pi x/h)$ distribution] may be found in Ref. 18 and they are summarized in Table IV. These values for PZT-4 ceramic are shown in Fig. 13 as the extreme lines to which the calculated plots have to approach asymptotically (in the case that cosine distribution of deformation exists the coupling coefficients of material are multiplied by the corresponding factors in order to get values for the effective coupling coefficients).

VI. ON THE ACOUSTIC RADIATION RELATED ASPECTS OF THE PROBLEM

All the calculations made so far were concerned with electromechanical parameters of unloaded (vibrating in air) transducers, whereas it is impossible to complete treatment of a transducer for practical applications without considering their acoustical and/or mechanical loading and other external actions typical for the particular applications. Although the energy method used[12,20] allows acoustical loads and the external actions to be formally included into equations of motion [see Eq. (45)], a detailed analysis of the particular electroacoustical transducers and corresponding acoustic loads is beyond the scope of this paper. However, in order to complete the transducer analysis, at least in the methodical sense, it is necessary to demonstrate how the external loads and actions can be included in the overall procedure of calculating electroacoustic parameters of a transducer in the framework of the developed approach.

One of the merits of the energy method[12] is that it allows the radiation problem for a transducer to be considered separately from its treatment as an electromechanical system. This may be recognized as especially advantageous, considering that the radiation problems itself can be rather complicated (even more complicated than the electromechanical part of a transducer treatment) and that they can vary depending on the transducer application, whereas the electromechanical part of solution may remain unchanged. A way to combine the results of solving the radiation problems, which we will assume to be known, with the electromechanical part of an unloaded transducer treatment that is presented by the equivalent circuit in Fig. 5(a), will be illustrated with examples of cylindrical air-backed underwater transducer designs, shown schematically as icons in Figs. 5(b) and 5(c).

The variant of transducer design in Fig. 5(b) with caps mechanically isolated from the ends of a piezoceramic tube is typical for the projectors and broadband receivers. The ends of the tube are free to vibrate and are isolated from the acoustic field by the caps that are considered to be absolutely rigid. The design variant in Fig. 5(c) with the caps attached to the ends of a piezoceramic tube and exposed to acoustic

TABLE III. Calculated resonance frequencies normalized to $f=30$ kHz.

Polarization	Thickness (radial)		Circumferential		Axial	
$h/2a$	Branch 0	Branch 2	Branch 0	Branch 2	Branch 0	Branch 2
0.50	1.00	3.35	0.89	3.43	0.99	3.06
0.75	0.99	2.25	0.88	2.30	0.98	2.08
1.00	0.97	1.72	0.87	1.76	0.95	1.61
1.25	0.94	1.42	0.84	1.44	0.89	1.36
1.50	0.89	1.25	0.81	1.25	0.82	1.24
1.75	0.82	1.17	0.76	1.14	0.74	1.18
2.00	0.74	1.13	0.70	1.08	0.66	1.14
2.25	0.67	1.11	0.65	1.04	0.60	1.13
2.50	0.61	1.09	0.59	1.02	0.54	1.12
2.75	0.56	1.08	0.55	1.01	0.50	1.11
3.00	0.51	1.08	0.51	1.00	0.46	1.10

TABLE IV. Effective coupling coefficients for the extreme mechanical systems.

	Branch 0		Branch 2	
Polarization	$(h/2a)\to 0$ $T_1=T_2=0,\ S_3\neq 0$	$(h/2a)\to\infty$ $T_2=T_3=0,\ S_1\neq 0$	$(h/2a)\to 0$ $T_2=0,\ S_3=0,\ S_1\neq 0$	$(h/2a)\to\infty$ $T_2=0,\ S_3=0,\ S_1\neq 0$
Circumferential				
k_m [a]	k_{33}	k_{31}	$\lvert k''_{31}\rvert$ [b]	$\lvert k'_{33}\rvert$
k_{eff}, PZT-4	0.70	0.30	0.08	0.65
Axial				
k_m [a]	k_{31}	k_{33}	$\lvert k'_{33}\rvert$ [b]	$\lvert k''_{31}\rvert$
k_{eff}, PZT-4	0.33	0.64	0.60	0.09
Thickness (radial)				
k_m [a]	k_{31}	k'_{31}	k'_{31}	k_{31}
k_{eff}, PZT-4	0.33	0.30	0.45	0.50

k_m is the coupling coefficient for piezoelectric ceramic material under different boundary conditions. Following Ref. 18, $k'_{31}=(\lvert k_{31}\rvert/\sqrt{1-k_{31}^2})\sqrt{1+\sigma_1^E/1-\sigma_1^E}$, $k''_{31}=-\lvert k_{31}\rvert+k_{33}\sqrt{\sigma_3^E\sigma_{13}^E}/\sqrt{1-k_{33}^2}$, and $k'_{33}=k_{33}-\lvert k_{31}\rvert\sqrt{\sigma_3^E\sigma_{13}^E}/\sqrt{(1-\sigma_3^E\sigma_{13}^E)(1-k_{31}^2)}$.

field are used predominately for the low frequency cylindrical hydrophones. A similar design can be used for radiation along the axis, but in this case it falls into category of Tonpilz transducers and usually thick-walled axially polarized cylinders are used for this purpose.

The radiation problem to be solved in the variant of Fig. 5(b) is that for a cylinder of finite height vibrating in different modes. A comprehensive analysis of the problem and the vast bibliography on this issue can be found in Ref. 21. Note that in the case that $h/\lambda>(0.6-0.7)$ a simpler model of the cylinder vibrating between two infinite rigid cylindrical baffles can be used. As the radial velocity distribution according to formula (51) is $U_r(x)=U_0+U_1\cos(\pi x/h)$, suppose that radiation problem is solved for uniformly vibrating cylinders and with velocity distribution $\cos(\pi x/h)$, and therefore the following functions can be considered known (subscripts 0 and 1 correspond with uniform and cosine distribution modes of vibration, respectively): the *far field sound pressures* $P_0(r,\omega)=A(r)U_0k_{\text{dif0}}$ and $P_1(r,\omega)=A(r)U_1k_{\text{dif1}}$, where $A(r)$ is a distance depending coefficient and $k_{\text{dif}i}$ are the diffraction coefficients; the *sound pressures on the cylinder surface* $P_{S0}(ka,x)=B_0(ka,x)U_0$ and $P_{S1}(ka,x)=B_1(ka,x)U_1$, where functions $B_i(ka,x)$ are introduced in order to make it obvious for the further analysis that the sound pressures are proportional to velocities.

The acoustical power radiated by a transducer now can be represented as[12]

$$\begin{aligned}\bar{\dot{W}}_{\text{ac}}&=\int_{\Sigma}(P_{S0}+P_{S1})U_r^*d\Sigma\\&=\int_{\Sigma}[B_0(ka,x)U_0+B_1(ka,x)U_1][U_0^*\\&\quad+U_1^*\cos(\pi x/h)]d\Sigma\end{aligned}\tag{67}$$

(here Σ stands for the transducer radiating surface). As the functions $B_i(ka,x)$ are assumed to be known, after integration over the transducer surface in Eq. (67) and simple manipulations the acoustical power can be expressed as $\bar{\dot{W}}_{\text{ac}}=Z_{\text{ac0}}\lvert U_0\rvert^2+Z_{\text{ac1}}\lvert U_1\rvert^2$, where

$$Z_{\text{ac0}}=Z_{\text{ac00}}+z_{\text{ac01}}U_1/U_0,\quad Z_{\text{ac1}}=Z_{\text{ac11}}+z_{\text{ac01}}U_0/U_1\tag{68}$$

are the radiation impedances. Z_{ac0} and Z_{ac1} represent the total acoustical loads related to the generalized velocities U_0 and U_1, Z_{ac00} and Z_{ac11} are the modal impedances for the uniform and cosine modes of vibration, and z_{01} is the mutual impedance between the modes. The impedances Z_{ac0} and Z_{ac1} must be included in Eq. (45) and introduced into the equivalent circuit.

The procedure of applying the energy method to calculate a transducer in the receive mode is described in Ref. 20. It is shown that the same equivalent circuit of Fig. 5 will be valid for calculation, if the equivalent forces

$$F_{\text{eqv0}}=P_fS_tk_{\text{dif0}},\quad F_{\text{eqv1}}=P_fS_tk_{\text{dif1}},\tag{69}$$

which are due to action of acoustic field, are introduced into the contours related to the velocities U_0 and U_1. In expression (69) P_f is the sound pressure in the free acoustic field, S_t is the area of the transducer radiating surface, and k_{dif0} and k_{dif1} are the same diffraction coefficients that are introduced for the radiating mode of operation.

The acoustic loads and equivalent forces determined by expressions (68) and (69) must be introduced into the equivalent circuit between the terminals 0,0 and 1,1 in Fig. 5(a) as the one-port networks, shown in Fig. 5(b) [we will refer to thus obtained circuit as to Figs. 5(a) and 5(b)]. After this is done, the equivalent circuit for electroacoustical transducer operating in the transmit and receive modes can be considered completed.

The far field sound pressure, $P(r,\omega)$, generated by a transducer will be found as

$$P(r,\omega) = P_0(r,\omega) + P_1(r,\omega) = A(r)(U_0 k_{\text{dif0}} + U_1 k_{\text{dif1}}), \tag{70}$$

so far as the velocities U_0 and U_1 are calculated from Eq. (45) or, alternatively, from equivalent circuit in Figs. 5(a) and 5(b).

The problem of calculating parameters of a transducer with symmetrical caps attached to the ends of a piezoceramic tube and exposed to acoustic field, as shown in Fig. 5(c), has both the mechanical and acoustical aspects. The mechanical part of the problem is to formulate the boundary conditions for the joints between the caps and the ends of the tube and to calculate the coupled vibration of the caps and of the tube under these conditions. The acoustical part is to determine the radiation impedances and equivalent forces applied to the surface of the caps and to the side surface of a transducer. The results of a solution to the mechanical part of the problem can be represented by the input impedances on the contour of the caps in the radial and axial directions, Z_{cr} and Z_{cx}. If to consider operation of a transducer in the receive mode at frequencies well below its resonance, which is typical for the low frequency hydrophones, then the acoustical part of the problem simplifies to determining of the diffraction coefficients only. In addition to the equivalent forces that are defined by expression (69), the equivalent forces $F_{\text{eqv.}c}$ applied to the caps have to be taken into account, and $F_{\text{eqv.}c} = P_f S_c k_{\text{dif.}c}$, were S_c is the surface area of a cap and $k_{\text{dif}c}$ is the diffraction coefficient related to the cap radiation.

Information regarding the diffraction coefficients for tubes of the finite height with capped ends can be found in Ref. 22 in addition to Ref. 21. After the input impedances of the caps and diffraction coefficients are determined, the equivalent circuit of a transducer can be completed by connecting the one-port networks, shown in Fig. 5(c), to the corresponding terminals of the mechanical branches in Fig. 5(a). Note that impedances Z_{cr} and Z_{cx} are doubled because of symmetry (associated mechanical energies are doubled), and Z_{cr} is ascribed to the generalized coordinate U_0 only, because according to Eq. (51) $U_r(\pm h/2) = U_0$. Also of note is that the velocity U_2 found from the equivalent circuit is the velocity on the contour of a cap. In the case that the cap cannot be considered as absolutely rigid, distribution of velocity on its surface must be determined. This calculation, as well as determining the input impedances of the caps, are pure mechanical problems, which can be treated separately. However, methodically the equivalent circuit of Fig. 5(a) with the one-port networks of Fig. 5(c) included provides the means to calculate the parameters of capped cylindrical transducers taking into account real elastic properties of the caps.

The results of calculating the transmit (TVR) and receive (FFVS) responses for the particular transducers with aspect ratios 1.32 and 2.2 and their comparison with the results of experimental investigation[11] are presented in Figs. 14 and 15.

The prototype transducers were made from thickness poled PZT-4 tubes with the outer diameter $D_0 = 38$ mm, thickness $t = 3.2$ mm, and heights $h = 46$ and 76 mm.

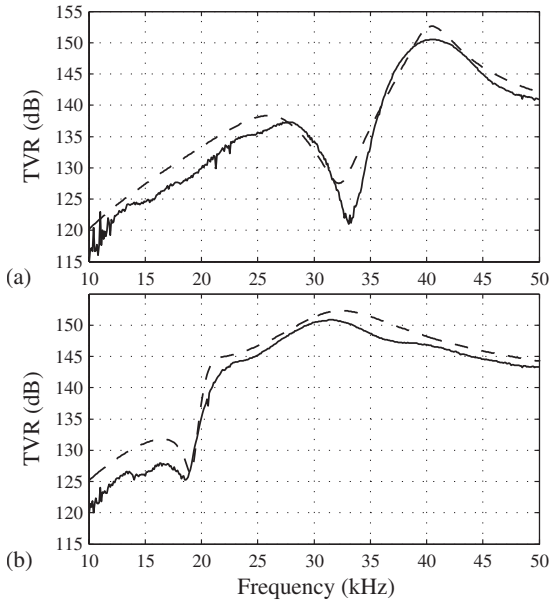


FIG. 14. Transmit frequency responses (TVR) calculated (dashed curve) and measured (solid curve) for a transducer comprised of a thickness poled tube with $h/2a = 1.32$ (a) and $h/2a = 2.2$ (b). Measured frequency responses are taken from Fig. 5 in Ref. 11.

The transducer design falls into the category of designs with free ends, shown schematically in Fig. 5(b) (for more details on the transducers design, see Ref. 11).

Calculations were made using the equivalent circuit [Fig. 5(b)]. All the mechanical $(K^E_{mi}, M_{\text{eqv}i})$, electrical $(C^{S_{1,2}}_{\text{el}}, R_{\text{eL}})$, and electromechanical (n_i) parameters involved in the equivalent circuit were calculated by formulas (14), (17), (41), (44), and (47), respectively. The resistances of mechanical losses, r_{mL}, were measured on the prototype transducers. Their values correspond to the mechanical quality factor $Q_m = 50$.

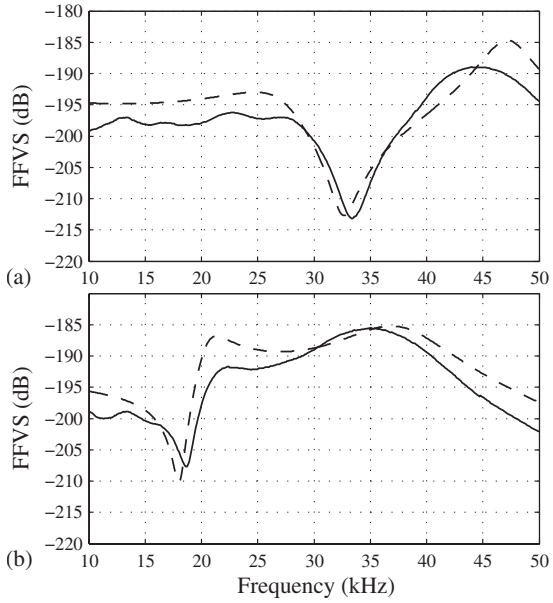


FIG. 15. Receive frequency responses (FFVS) calculated (dashed curve) and measured (solid curve) for a transducer comprised of a thickness poled tube with $h/2a = 1.32$ (a) and $h/2a = 2.2$ (b). Measured frequency responses are taken from Figs. 11(a) and 11(b) in Ref. 11, where they are shown by solid lines.

Given that the height of the transducers is on the order of the wavelength in the frequency ranges around their resonance frequencies, the radiation impedances and diffraction coefficients per unit length were approximated by those per unit length of the infinitely long pulsating cylinder of the same diameter. This resulted in

$$Z_{ac0} = -j\rho c 2\pi a h H_0^{(2)}(ka)/H_0^{(2)\prime}(ka), \quad Z_{ac1} = (2/\pi)Z_{ac0};$$

$$k_{dif0} = -2j/\pi k a H_0^{(2)\prime}(ka), \quad k_{dif1} = (2/\pi)k_{dif0}.$$

It is of note that the calculations confirmed the existence of notches in the frequency responses, for which a qualitative explanation was given in Ref. 11.

VII. CONCLUSION

Analytical treatment of the cylindrical thin-walled piezoelectric ceramic transducers based on employing the energy method[12] and coupled vibration analysis[13] is presented. As it is inherent in the method used, operation of the transducers is described in terms of Lagrange's type equations and of the equivalent electromechanical circuits. It is shown that the approach developed provides clear physical and computationally straightforward means for calculating the electromechanical parameters of piezoceramic tubes in two-dimensional approximation for a broad range of their height-to-diameter aspect ratios and for the different directions of the tube polarization that would be employed. The treatment allows consideration of the operation of the transducers under different mechanical and acoustical loads and other external actions. This is illustrated by presenting the results of calculating the dependences of the resonance frequencies and effective coupling coefficients as functions of aspect ratios for the tubes poled in radial, circumferential, and axial directions. As an example of applying the proposed technique to design the electroacoustical transducers, the results are presented for calculating the transmit and receive frequency responses for the transducers built from the piezoceramic tubes having different aspect ratios.

For the particular case of thickness poled tubes and for transducers of air-backed design built from such tubes the results of calculations are shown to be in good agreement with recently published[10,11] experimental results in a broad range of aspect ratios.

Based on the materials discussed in the paper it can be concluded that the approach presented can be used for designing and optimizing the parameters of the transducers that employ the thin-walled piezoceramic tubes for the practically reasonable range of their aspect ratios.

ACKNOWLEDGMENTS

The author wishes to thank Dr. David A. Brown for his help in editing and revising the paper in the course of preparation for publication. Special thanks go to Corey Bachand for implementing the calculations and for preparing the plots and figures for the paper. Also he is the main force behind manufacturing the transducers and obtaining the experimental data referred to in the paper. This work was supported in part by BTECH Acoustics and by Office of Naval Research 321 MS.

[1] J. F. Haskins and J. L. Walsh, "Vibrations of ferroelectric cylindrical shells with transverse isotropy," J. Acoust. Soc. Am. **29**, 729–734 (1957).
[2] A. E. H. Love, *Mathematical Theory of Elasticity*, 4th ed. (Dover, New York, 1944), p. 546.
[3] G. E. Martin, "Vibrations of longitudinally polarized ferroelectric cylindrical tubes," J. Acoust. Soc. Am. **35**, 510–520 (1963).
[4] H. Wang, "On the tangentially and radially polarized piezoceramic thin cylindrical tube transducers," J. Acoust. Soc. Am. **79**, 164–176 (1986).
[5] D. D. Ebenezer and P. Abraham, "Eigenfunction analysis of radially polled piezoelectric shells of finite length," J. Acoust. Soc. Am. **102**, 1549–1558 (1997).
[6] D. S. Drumheller and A. Kalnins, "Dynamic shell theory for ferroelectric ceramics," J. Acoust. Soc. Am. **47**, 1343–1353 (1970).
[7] V. H. Lazutkin and A. I. Mihailov, "Vibrations of the longitudinally polarized piezoceramic cylinders of finite size," Sov. Phys. Acoust. **22**, 393–399 (1976).
[8] D. D. Ebenezer and P. Abraham, "Piezoelectric thin shell theoretical model and eigenfunction analysis of radially polarized ceramic cylinders," J. Acoust. Soc. Am. **105**, 154–163 (1999).
[9] P. H. Rogers, "Mathematical model for a free-flooded piezoelectric cylinder transducer," J. Acoust. Soc. Am. **80**, 13–18 (1986).
[10] B. S. Aronov, D. A. Brown, and S. Regmi, "Experimental investigation of coupled vibrations in piezoelectric cylindrical shells," J. Acoust. Soc. Am. **120**, 1374–1380 (2006).
[11] B. S. Aronov, D. A. Brown, and C. L. Bachand, "Effects of coupled vibrations on the acoustical performance of underwater cylindrical transducers," J. Acoust. Soc. Am. **122**, 3419–3427 (2007).
[12] B. S. Aronov, "The energy method for analyzing the piezoelectric electroacoustic transducers," J. Acoust. Soc. Am. **117**, 210–220 (2005).
[13] E. Giebe and E. Blechschmidt, "Experimental and theoretical studies of extensional vibrations of rods and tubes," Ann. Phys. **18**, 417–485 (1933).
[14] M. C. Junger and F. G. Rosato, "The propagation of elastic waves in thin-walled cylindrical shells," J. Acoust. Soc. Am. **26**, 709–713 (1954).
[15] S. P. Timoshenko and D. H. Young, *Vibration Problems in Engineering*, 3rd ed. (Van Nostrand, Toronto, 1955).
[16] B. S. Aronov and L. B. Nikitin, "Calculation of the flexural modes of piezoceramic plates," Sov. Phys. Acoust. **27**, 382–387 (1981).
[17] B. S. Aronov, "Energy analysis of a piezoelectric body under nonuniform deformation," J. Acoust. Soc. Am. **113**, 2638–2646 (2003).
[18] D. A. Berlincourt, D. R. Curran, and H. Jaffe, "Piezoelectric and piezomagnetic materials and their functions in transducers," in *Physical Acoustics*, edited by W. P. Mason (Academic, New York, 1964), Vol. **IA**.
[19] B. S. Aronov, "On the optimization of the effective electromechanical coupling coefficients of a piezoelectric body," J. Acoust. Soc. Am. **114**, 792–800 (2003).
[20] B. S. Aronov, "The energy method for analyzing the piezoelectric electroacoustic transducers. II. (With the examples of the flexural plate transducers)," J. Acoust. Soc. Am. **118**, 627–637 (2005).
[21] E. L. Shenderov and V. A. Kozyrev, "Radiation impedance of a cylinder of finite length," Sov. Phys. Acoust. **26**, 230–235 (1980).
[22] W. J. Trott, "Sensitivity of piezoceramic tubes with capped or shielded ends above the omnidirectional frequency range," J. Acoust. Soc. Am. **62**, 565–568 (1977).

Expert diagnostic system for moving-coil loudspeakers using nonlinear modeling

Mingsian R. Bai[a] and Chau-Min Huang
Department of Mechanical Engineering, National Chiao-Tung University, 1001 Ta-Hsueh Road, Hsin-Chu 300, Taiwan

(Received 18 July 2008; revised 5 December 2008; accepted 8 December 2008)

This work aims at the development of an expert diagnostic system for moving-coil loudspeakers. Special emphasis is placed on the defects resulting from loudspeaker nonlinearities. As a loudspeaker operates in the large signal domain, nonlinear distortions may arise and impair sound quality. Analysis of nonlinear responses can shed light on potential design faults of a loudspeaker. By exploiting this fact, this expert diagnostic system enables classification of design faults using a defect database alongside an intelligent fault inference module. Six types of defects are investigated in this paper. A large signal model based on electromechanical analogous circuits is employed for generating the defect database, through which a neural-fuzzy network is utilized for inferring the defect types. Numerical simulations and experimental investigations were undertaken for validating the loudspeaker diagnostic system.
 [DOI: 10.1121/1.3058639]

PACS number(s): 43.38.Dv, 43.80.Qf [AJZ] Pages: 819–830

I. INTRODUCTION

As a loudspeaker operates in the large signal domain, nonlinear distortions may arise and can potentially impair sound quality. Nonlinear causes (Table I) may be attributed to faults resulting from the design or manufacturing process. Traditionally, diagnostics of loudspeakers rely on well trained human experts who are able to evaluate faulty loudspeakers according to the distortions or symptoms. However, human experts are scarce and require long training. Plus, human experts can make false judgments, especially after a long period of work. It is then highly desirable to analyze and diagnose loudspeakers in a systematic and efficient manner that is free of human errors. To address these issues, this paper aims to develop an expert diagnostic system for moving-coil loudspeakers with emphasis on nonlinear causes.

Nonlinear distortions of a loudspeaker may arise as a result of various causes such as nonlinear compliance, nonlinear force factor, nonlinear inductance, Doppler Effect, suspension creep, etc.,[1–3] as defined in an IEC standard PAS 62458. It is often insufficient to account for nonlinearities using linear loudspeaker models.[4] Since sound quality during audio reproduction may suffer due to nonlinear distortions, to be able to assess and classify nonlinear causes would be quite useful in the design phase and even in the manufacturing quality control phase of loudspeakers.[5–8] In the past, metrics including dc-displacement, harmonic distortion, intermodulation, etc., have been suggested to quantify loudspeaker nonlinearities.[9–12] On the other hand, large signal models[13–17] based on electrical equivalent circuits have been suggested for predicting the performance of a loudspeaker in the large signal domain. Remedies for nonlinear faults of loudspeaker have also been reported in literature.[18]

In this paper, we develop an expert system aimed at loudspeaker diagnostics by exploiting nonlinear signatures of loudspeaker responses. An expert system[18] comprises a computer program with feature extraction, knowledge, and inference procedure, which mimics a human expert in analyzing problems and making decisions. Expert system has found many applications in automatic diagnosis, prediction, and control since the advent of the artificial intelligence. The loudspeaker diagnostic system in this paper is developed in two major steps: (1) creation of a cause-symptom database via a nonlinear loudspeaker model based on electrical equivalent circuits, and (2) training the neural-fuzzy network using the database. The loudspeaker diagnostic process begins with measuring numerous nonlinear metrics (symptoms) using a distortion analyzer and then deduce the fault types (causes) automatically via the neural-fuzzy network.

In the proposed system, the expert diagnostic system seeks to infer potential defect types in lieu of human experts, by using a loudspeaker cause-symptom database. The inference engine is constructed on the basis of neural-fuzzy networks. Fuzzy logic[19] was first introduced by Zadeh in 1965. Fuzzy logic mimics human reasoning that allows for some "fuzzy" latitude in making decisions. Neural network[20,21] (NN) is another technology that enables realizing human intelligence in machines. Recently, neural-fuzzy systems [fuzzy neural network (FNN)] (Refs. 21 and 22) that exploit the combined advantages of fuzzy logic and NN have received increasing interest in the AI research. For example, a five-layered FNN (Ref. 22) suggested by Becraft and Isermann[23] is widely used in many problems. Lin's five-layered FNN is also employed in this paper to investigate nonlinear diagnostic problem of moving-coil loudspeakers.

The expert diagnostic system in this paper is implemented on a notebook computer, with the aid of a nonlinear

[a]Author to whom correspondence should be addressed. Electronic mail: msbai@mail.nctu.edu.tw

TABLE I. Nomenclature of nonlinear loudspeaker model.

State variables	Unit	Interpretation
$u(t)$	V	The driving voltage at loudspeaker terminals
$x(t)$	mm	Displacement of the voice coil
$v(t)$	mm/s	Velocity of the voice coil
$i(t)$	A	The electric input current
$i_2(t)$	A	The current through L_2
Electrical parameters	**Unit**	**Interpretation**
$R_E(T_V)$	Ω	dc resistance of voice coil
$L_E(x)$	mH	Part of the voice coil inductance
$L_2(x)$	mH	Parainductance of the voice coil
$R_2(x)$	Ω	Electric resistance caused by eddy currents
Mechanical parameters	**Unit**	**Interpretation**
$Bl(x)$	N/A	The force factor
$F_m(x,i,i_2)$	N	The reluctance force
$C_{\mathrm{MS}}(x)$	mm/N	Mechanical compliance of driver suspension [theinverse of stiffness $K_{\mathrm{MS}}(x)$]
R_{MS}	kg/s	Mechanical resistance of driver suspension losses
M_{MS}	kg	Mechanical mass of driver diaphragm assembly including voice- coil and air load
Thermal model	**Unit**	**Interpretation**
R_{TV}	K/W	Thermal resistance of path from coil to magnet
R_{TM}	K/W	Thermal resistance of magnet structure to ambient air
C_{TV}	J/K	Thermal capacitance of voice coil and nearby surroundings
C_{TM}	J/K	Thermal capacitance of magnet structure
$P(t)$	W	Real electric input power
$T_V(t)$	K	Temperature of the voice coil
$\Delta T_V(t)$	K	$\Delta T_V(t)=T_V(t)-T_A$, increase in voice coil temperature
$T_M(t)$	K	Temperature of the magnet structure
$\Delta T_M(t)$	K	$\Delta T_M(t)=T_M(t)-T_A$
T_A	K	Ambient temperature
δ	K^{-1}	$\delta=0.003\,93\ \mathrm{K}^{-1}$ for copper and $\delta=0.003\,93\ \mathrm{K}^{-1}$ foraluminum.

analyzer equipped with a laser vibrometer. Two sample loudspeakers were employed in the case studies to validate the expert system. The experimental results will be discussed and summarized in Sec. V.

II. THE LARGE-SIGNAL MODEL OF MOVING-COIL LOUDSPEAKERS

As an electroacoustic transducer, a moving-coil loudspeaker [Fig. 1(a)] involves energy conversion and coupling among the electrical, mechanical, and acoustical domains. At low frequencies where the wavelength is large in comparison to the geometric dimensions, the state of a loudspeaker can be described simply by a lumped parameter model that can be represented by a generic equivalent circuit in Fig. 1(b) Note that, in contrast to the small signal model, the force factor $Bl(x)$, mechanical compliance $C_{\mathrm{MS}}(x)$, and voice coil inductance $L_E(x)$ are not constant, but varying with the displacement of the voice coil.[8]

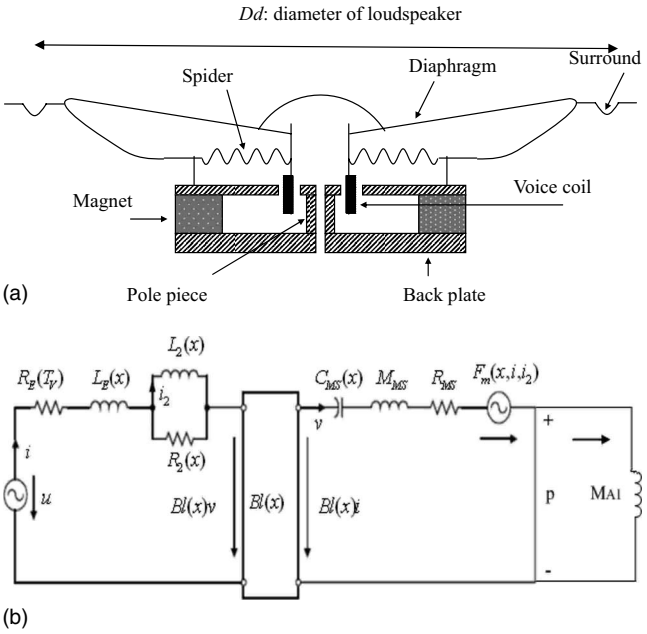


FIG. 1. Electroacoustic analogy of a moving-coil loudspeaker. (a) The physical structure. (b) The equivalent circuit.

A. Nonlinearities of moving-coil loudspeakers

1. Loudspeaker stiffness $K_{\mathrm{MS}}(x)$

Like a mechanical spring, a suspension system[9,10] is used to generate a restoring force for a loudspeaker, pulling the voice coil back to its rest position.[10,11]

2. Force factor $Bl(x)$

The force factor $Bl(x)$ (Ref. 9) serves as the coupling between the mechanical and electrical domains of a moving-coil loudspeaker, where B is flux density and l is the length of the voice coil in the magnet air gap. The force factor $Bl(x)$ may result in two nonlinear effects: (a) back electromotive force and (b) Lorentz force $F=Bl(x)i$.

3. Voice coil inductance $L_E(x)$

Inductance $L_E(x)$ (Refs. 9 and 12) that varies with displacement is another source of loudspeaker nonlinearity. The magnetic flux is dependent on the position of the coil and the magnitude of the current and frequency.[9,12]

4. Other nonlinearities

In addition to the preceding nonlinearities,[9,10] several other nonlinearities are included in this paper.

A. Doppler effect. Large cone excursions occur at low frequencies during loudspeaker operations. This mechanism may be described by the product of displacement and differentiated sound pressure and requires low and high frequency components at the same time. If the distance between the source and the listener can vary drastically to give rise to time delay modulation of sound pressure signals, the Doppler effect may lead to intermodulations in the high frequency regime.

B. Reluctance force F_m. The displacement-varying $L_E(x)$ may introduce an additional reluctance force

C. Parainductance and the resistance caused by eddy currents. Ideal linear inductance is not sufficient for modeling the phenomenon that electrical impedance increases at higher frequencies due to the inductance loss.

D. Rub and buzz (Refs. 25 and 26). This kind of distortion occurs when the voice coil rubs on the pole tips, the lead wire strikes the diaphragm, or a loose glue joint starts vibrating. This is especially a common cause of nonlinear distortion for small speakers or microspeakers, where overdriven operation tends to exceed the mechanical design limits such as the air gap and the excursion limit.

B. Large signal model

Traditional linear models[4] are insufficient for predicting large signal behavior of loudspeakers. Clearly, the large signal model[13–17] is a preferred way to simulate nonlinear effects such as nonlinear distortions and the dc-component of the cone displacement. The large signal model considers the loudspeaker as a coupled electrical-mechanical-acoustical system with nonlinear parameters, as shown in Fig. 1(b).

The dynamics of Fig. 1(b) can be described using the following differential equations:

$$u = iR_E(T_V) + \frac{d(L_E(x)i)}{dt} + \frac{d(L_2(x)i_2)}{dt} + Bl(x)v, \tag{1}$$

$$\frac{d(L_2(x)i_2)}{dt} = (i - i_2)R_2(x), \tag{2}$$

$$Bl(x)i - F_m(x,i,i_2) = (M_{\mathrm{MS}} + M_{A1})\frac{d^2x}{dt^2} + R_{\mathrm{MS}}\frac{dx}{dt} + K_{\mathrm{MS}}x. \tag{3}$$

Choosing $y_1 = x$, $y_2 = v$, $y_3 = i$, and $y_4 = i_2$ as state variables enables us to rewrite Eqs. (1)–(3) into the following state-space equation:

$$\begin{bmatrix} \dot{y}_1 \\ \dot{y}_2 \\ \dot{y}_3 \\ \dot{y}_4 \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ \dfrac{-1}{(M_{\mathrm{MS}} + M_{A1})C_{\mathrm{MS}}(x)} & \dfrac{-R_{\mathrm{MS}}}{(M_{\mathrm{MS}} + M_{A1})} & \dfrac{Bl(x) + \frac{1}{2}\frac{dL_E(x)}{dx}y_3}{(M_{\mathrm{MS}} + M_{A1})} & \dfrac{\frac{1}{2}\frac{dL_2(x)}{dx}y_4}{(M_{\mathrm{MS}} + M_{A1})} \\ 0 & \dfrac{-Bl(x) - \frac{dL_E(x)}{dx}y_3}{L_E(x)} & \dfrac{-(R_E(T_V) + R_2(x))}{L_E(x)} & \dfrac{R_2(x)}{L_E(x)} \\ 0 & 0 & \dfrac{R_2(x)}{L_2(x)} & \dfrac{-R_2(x) - \frac{dL_2(x)}{dx}y_2}{L_2(x)} \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ \dfrac{1}{L_E(x)} \\ 0 \end{bmatrix} u. \tag{4}$$

Numerical integration algorithms such as Runge–Kutta method[27] can readily be applied to solve the state-apace equation for the nonlinear responses. Once the velocity $y_2 = v$ is obtained, the far field sound pressure can be readily calculated using a baffled point source model

$$p(t,r) = \frac{\rho_0}{2\pi r}S_d\frac{dv}{dt}e^{-j2\pi f_2(r-x)/c}, \tag{5}$$

where r is the distance between the diaphragm and the listening position, ρ_0 is the density of air, and c is the speed of sound.

In the present work, we mainly focus on the nonlinear distortions that generally occur at low frequencies, where the wavelength is large as compared to the speaker dimension and Eq. (5) suffices to model the nonlinear system with acceptable accuracy.

C. Nonlinear measures of symptoms

A single tone is used as a stimulus in harmonic distortion measurements. The nth harmonic distortion associated with the excitation frequency f_1 is defined as

$$\mathrm{HD}_n = \frac{|P(nf_1)|}{P_t} \cdot 100\,\%, \tag{6}$$

where $P(nf_1)$ is the complex spectrum of the sound pressure p at the nth harmonic, and P_t is the rms value of the total signal within an averaging time T

$$P_t = \sqrt{\frac{1}{T}\int_0^T p(t)^2 dt}. \tag{7}$$

In addition, total harmonic distortion (THD) is a common nonlinear measure for a signal with n significant harmonics:

$$\mathrm{THD} = \frac{|P(2f_1)|^2 + |P(3f_1)|^2 + \ldots + |P(nf_1)|^2}{P_t} \times 100\,\%. \tag{8}$$

The dc component of displacement X_{dc} is also a useful measure for evaluating suspension asymmetries.

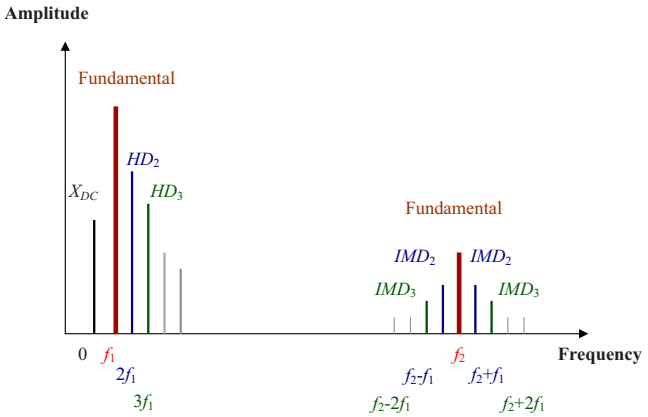


FIG. 2. (Color online) The generic frequency spectrum of a state variable, e.g., sound pressure p, displacement x, and input current i, generated by the two-tone stimulus.

$$X_{\mathrm{dc}} = \frac{|X(0)|}{X_t} \times 100\,\%\,, \tag{9}$$

where $X(0)$ is the dc-component in the complex spectrum of the cone displacement x and X_t is the rms value of the displacement x.

Apart from the preceding single tone measures, a two-tone (f_1 and f_2) stimulation recommended in the IEC standard 60268-5 (Ref. 18) is also utilized for assessing nonlinear loudspeaker symptoms. The excitation signal can be expressed as

$$u(t) = U_1 \sin(2\pi f_1) + U_2 \sin(2\pi f_2). \tag{10}$$

Therefore, an important nonlinear measure is the intermodulation distortion (IMD). IMD accounts for extra frequency components due to intermodulations occurring at $f_2 \pm nf_1$ ($n=1,2,\ldots$), as depicted in Fig. 2. The nth-order IMD associated with f_2 is defined as

$$\mathrm{IMD}_n = \frac{|P(f_2-(n-1)f_1)| + |P(f_2+(n-1)f_1)|}{|P(f_2)|} \times 100\,\%\,. \tag{11}$$

D. Creating a database of loudspeaker causes

A database documenting the relationships between causes and symptoms of loudspeaker nonlinearities is required for training the neural-fuzzy system. Instead of gathering data from real loudspeakers, which is extremely tedious and impractical, we opt for synthetic data obtained using large signal simulations with prescribed nonlinearities.

Among the frequently encountered loudspeaker nonlinearities,[9,10] six common defects are investigated in this paper: (a) asymmetry in $C_{\mathrm{MS}}(x)$, (b) dc-displacement of voice coil offset, (c) asymmetry in $L_E(x)$, (d) coil height (the height of voice coil exceed the air gap), (e) symmetrical limiting of suspension (the rate of change of suspension compliance with respect to cone displacement), and (f) Doppler effect.

Each nonlinear cause leads to unique symptoms of distortions, as summarized in Table II.[9] The entries marked with crosses in Table II indicate that a strong link exists between the nonlinearity and the distortion. The nonlinear causes in a loudspeaker's response can be subdivided into three categories, "critical nonlinearity variation," "asymmetric nonlinearity," and "Doppler effect," as follows:

1. Critical nonlinearity variation: "Coil height" and "symmetrical limiting of suspension"

Large cone excursion in general forces the voice coil to leave the air gap and the suspension to be overstretched. This gives rise to nonconstant symmetric $C_{\mathrm{MS}}(x)$ and $Bl(x)$ curves.[9,10] A symmetric curve usually produces the third- and other odd-order distortion components. To simulate the coil height and the "suspension limiting" defect, the second-degree term of the Bl and the C_{MS} series are usually multiplied by a factor $\beta > 1$ to increase variations, where x denotes the displacement of voice coil from the rest position.

$$Bl'(x) = b_0 + b_1 x + \beta b_2 x^2 + \sum_{i=3}^{n} b_i x^i, \tag{12}$$

$$C'_{\mathrm{MS}}(x) = c_0 + c_1 x + \beta c_2 x^2 + \sum_{i=3}^{n} c_i x^i. \tag{13}$$

2. Asymmetric nonlinearity: "Coil offset" and "asymmetry in $C_{\mathrm{MS}}(x)$ and $L_E(x)$"

Asymmetric nonlinearities generally result in the second- and other even-order distortion components. Asymmetric $Bl(x)$ and $C_{\mathrm{MS}}(x)$ curves can arise due to imperfection in loudspeaker manufacturing. These defects can be modeled by shifting the $Bl(x)$ and $C_{\mathrm{MS}}(x)$ curves by a small constant ε.

TABLE II. Relationship between nonlinear causes and symptoms of moving-coil loudspeakers.

Physical cause	X_{dc}	HD_2	HD_3	IMD_2	IMD_3	$IMD_{l,i}$
Coil offset		X		X		
Coil height			X		X	
Asymmetry in $C_{\mathrm{MS}}(x)$	X	X				
Symmetrical limiting of suspension			X			
Asymmetry in $L_E(x)$				X		X
Doppler effect				X		

$$Bl'(x) = Bl(x-\varepsilon) = \sum_{i=0}^{n} b_i (x-\varepsilon)^i, \tag{14}$$

$$C'_{\mathrm{MS}}(x) = C_{\mathrm{MS}}(x-\varepsilon) = \sum_{i=0}^{n} c_i (x-\varepsilon)^i. \tag{15}$$

The inductance $L_E(x)$ without shorting ring also has an asymmetric curve. To model this, we multiply the linear term of the L_E series by a factor $\beta > 1$ to yield

$$L'_E(x) = l_0 + \beta l_1 x + \sum_{i=2}^{n} l_i x^i. \tag{16}$$

3. Doppler effect

Doppler distortion generally arises because of large cone displacement in low frequencies at which the loudspeaker behaves as a moving source. The low frequency displacement is so large that the high frequency signal riding on the top of this low frequency signal is effectively changing its position. This is not very critical for low frequency component itself but causes intermodulation of high frequency signals with a short wavelength. This effect leads to frequency shift, which in turn increases IMD by 6 dB/octave toward higher frequencies.[9,13] This effect can be simulated by increasing the frequency of the second tone f_2 in Eq. (5). That is,

$$p(t,r) = \frac{\rho_0}{2\pi r} S_d \frac{dv}{dt} e^{-j2\pi(f_2+\Delta f_2)(r-x)/c}, \tag{17}$$

where Δf_2 denotes the frequency shift of f_2. The Doppler distortion arises because the second frequency is also present and its distance (between source and listener) is modulated. The cone excursion could be only a few centimeters. However, this distance between the source and the listener varies enough to give rise to Doppler effect. It is as if the high frequency source is moving back and forth, thus giving rise to the shifting in pitch.

In order to establish a database that incorporates all the aforementioned loudspeaker nonlinearities, a hybrid approach is undertaken. We begin with identifying the parameters of three real loudspeakers and then carrying out numerical simulations by varying β and ε in Eqs. (12)–(16) and input voltage to produce synthetic data. A total of 700 sets of nonlinear causes and symptoms are created for the database using this approach.

III. INFERRING NONLINEAR DEFECTS USING NEURAL-FUZZY SYSTEMS

In this paper, the FNN (Ref. 21) is utilized for the loudspeaker diagnostic problem. The six nonlinear causes and their associated symptoms summarized in Table II are selected for the input and output layers of the FNN, respectively. The FNN with a fuzzy rule-based model embedded in a NN carries out intelligent inference with combined advantages of both artificial NN and fuzzy logic. On the basis of the preceding database, the FNN is trained to deal with unknown nonlinear diagnostic problems of loudspeakers. Figure 3 illustrates the five layers in the FNN: the input layer, the input membership layer, the fuzzy rule layer, the output membership layer, and the conclusion layer. The function of each layer is explained next.

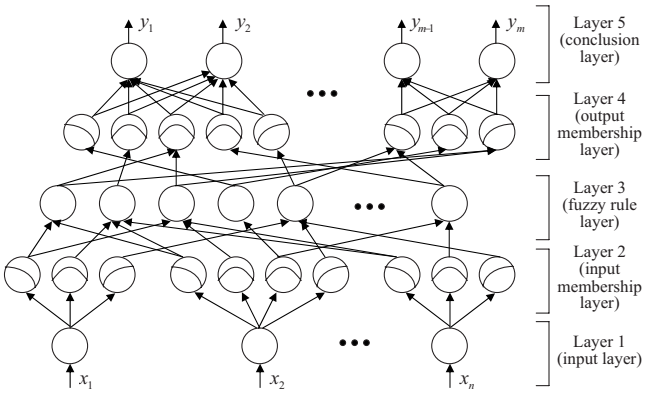


FIG. 3. The schematic of a McCulloch–Pitts neuron.

Unlike classical sets representing only binary states, *true* or *false*, the fuzzy sets use a membership function to determine the degree of "truth." Let $\tilde{A}$ be a fuzzy set and U be its *universe of discourse*

$$\tilde{A} = \{(x, \mu_{\tilde{A}}(x)) | x \in U\}, \tag{18}$$

where $\mu_{\tilde{A}}(x)$ is the membership function that represents the degree of x belonging to the fuzzy set $\tilde{A}$, and its value is always within the interval [0,1]. In this FNN, the input membership layer consists of several membership functions, "minor," "moderate," and "Severe," to indicate various levels of nonlinear distortions. The output membership layer consists of two membership functions representing "having this problem" and "not having this problem," respectively. The output membership function returns the probability ranging from 0% to 100% for each nonlinear cause. The core of the FNN is the fuzzy rule layer that associates the input membership layer and the output membership layer to realize the following "if-then" rule:

$$\text{Rule } i\text{: if } (x_1 \text{ is } A_{i1} \text{ and} \cdots \text{and } x_n \text{ is } A_{in})$$

$$\text{then } (y_1 \text{ is } B_{i1} \text{ and} \cdots \text{and } y_n \text{ is } B_{in}), \tag{19}$$

where A_{ij} and B_{ij} are the input and output fuzzy sets, respectively. Finally, the conclusion layer serves as a *defuzzifier* and makes an inference of fault type by using the mean of centric method.[21]

Figure 4 illustrates one basic element of the FNN, usually called an *M-P neuron*. This jth processing element computes a weighted sum of its inputs and then exports the output o_j:

$$o_j(t+1) = f\left(\sum w_{ij} x_i(t)\right), \tag{20}$$

where f is a threshold function such as the *sigmoid function*[22] in which the relationship between the input x and the output $f(x)$ is given by

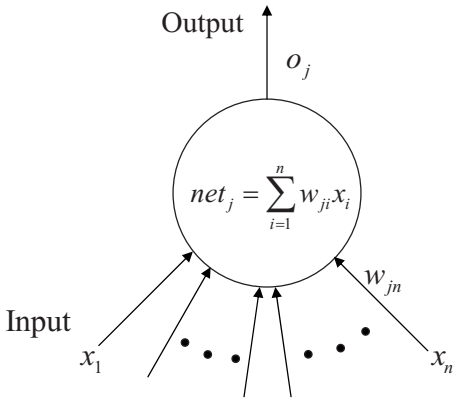


FIG. 4. The structure of the FNN, where $x's$ and $y's$ denote the input and output variables, respectively.

$$f(x) = \frac{1}{1 + e^{-x}}. \tag{21}$$

The neurons in the five layers of the FNN are interconnected by synapses.

At the initial stage, each synaptic weight in the FNN is set to be a random number. No rules exist between the input and output membership layers. The membership functions in the input and output membership layers are initially set by experience, as shown in Figs. 5(a) and 5(b). Next, at the training stage, a learning process is carried out to create

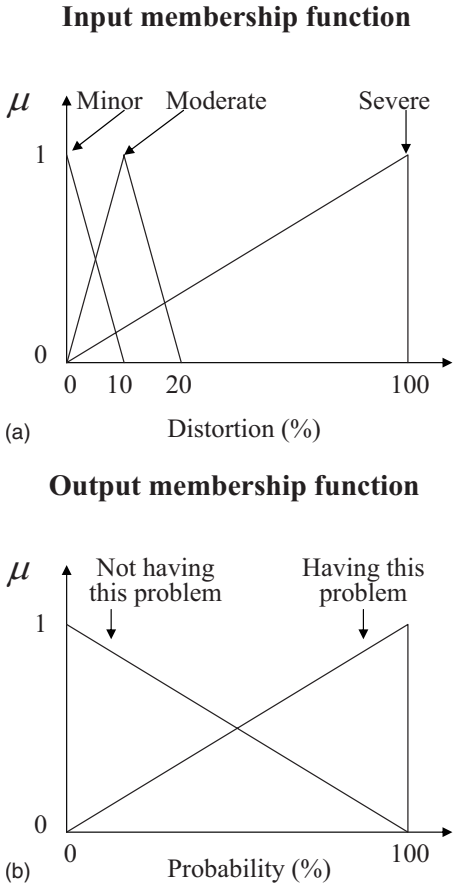


FIG. 5. The membership functions and the linguistic terms in the fuzzy logic system. (a) The input membership function. (b) The output membership function.

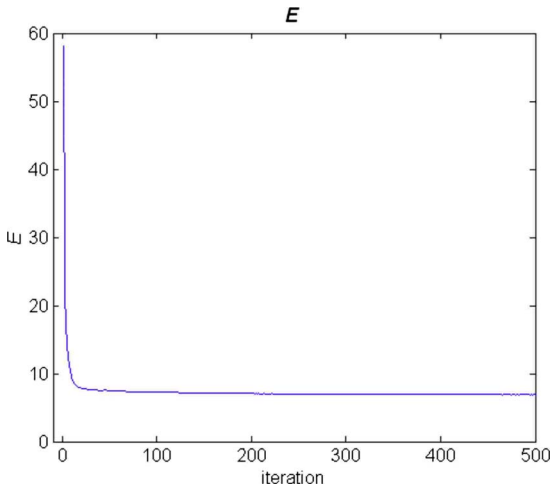


FIG. 6. (Color online) The membership functions of rule 1 and rule 2 in the fuzzy rule layer.

fuzzy rules and adjust the neural weights and the membership functions by using the preceding database. An error function E is defined for the network learning

$$E = \frac{1}{2} \sum_i (y_i - y_i')^2, \tag{22}$$

where y_i is the ith expected output and y' is the ith calculated output. A supervised learning algorithm with backpropagation[22] is employed to minimize the error function. In the algorithm, the increment of the weight w_{ij}^n at time t is updated according to

$$\Delta w_{ij}^n(t) = \eta \times \frac{\partial E}{\partial w_{ij}^n} + \alpha \times \Delta w_{ij}^n(t-1), \tag{23}$$

where $\Delta w_{ij}^n(t)$ is the increment of the weight between the ith neuron of the $(n+1)$th layer and the jth neuron of the nth layer at time t, η is the learning factor, and α is the momentum term for accelerating the convergence. As soon as the error decreases to an acceptable range, the training is terminated. To be specific, 500 and 200 cases out of the preceding database are used for training and verification, respectively. Figure 6 shows the learning curve of this FNN, where the system error reduces from 58 to 6.9 after 500 iterations. A judicious choice of 70% threshold indicates whether or not a loudspeaker under test has a particular nonlinear cause. To verify the FNN, in-group tests of the remaining 200 cases were conducted, with results summarized in Table III. A very successful detection rate was achieved using the network (average rate of correct inference=97.0%).[28,29]

An example for the FNN system is given as follows. Assume that two rules exist in training fuzzy network.

(1) If HD3 is moderate and IMD3 is minor, then it may have "coil height" problem.
(2) If HD3 is moderate and IMD3 is moderate, then it may not have coil height problem.

Two weighting factors W_1 and W_2 are given to those two rules in the FNN system. The membership functions of HD3 and IMD3 obtained after training are shown Fig. 7. For in-

TABLE III. Results of the in-group test of the FNN diagnostic system.

Nonlinear cause	Number of case	Rate of correct inference	Rate of incorrect inference
Normal	36	100%	0%
Asymmetry in $C_{MS}(x)$	14	100%	0%
Coil offset	13	100%	0%
Asymmetry in $L_E(x)$	12	100%	0%
Coil height	13	100%	0%
Symmetrical limiting of suspension	19	100%	0%
Doppler effect	14	100%	0%
Asymmetry in $C_{MS}(x)$+ coil offset	26	88.4%	11.6%
Asymmetry in $C_{MS}(x)$+ Asymmetry in $L_E(x)$	12	100%	0%
Asymmetry in $L_E(x)$+ coil offset	23	95.7%	4.3%
Symmetrical limiting of suspension+coil height	18	88.9%	11.1%
Total number of cases=200			
Average rate of correct inference=97.0%			

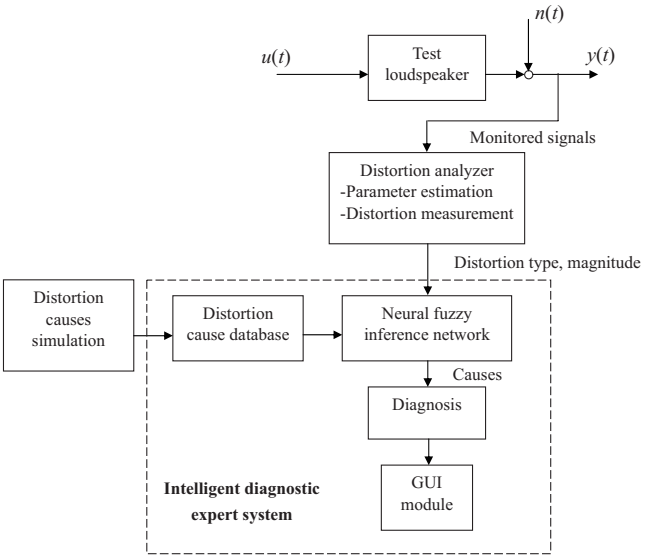


FIG. 8. The block diagram of the expert diagnostic system. The intelligent diagnostic system comprises a defect database and an inference engine marked by the dotted line.

stance, the distortion data HD3=10% and IMD3=15% correspond to 0.5 minor and 0.5 moderate in HD3. The same data correspond to 0.25 minor and 0.75 moderate in IMD3. Next, calculate the "degree of detonation" for each rule:

$$\text{Rule } 1 = W_1 \times 0.5 \times 0.25, \tag{24}$$

$$\text{Rule } 2 = W_2 \times 0.5 \times 0.75. \tag{25}$$

Then, according to the degree of detonation, inference of the defect coil height can be drawn, or "defuzzified," using the certainty equivalent principle.[22]

IV. EXPERIMENTAL INVESTIGATIONS

Figure 8 illustrates the architecture of the expert diagnostic system proposed in this paper. Two loudspeakers denoted as Drivers A and B were employed in the case study to examine the diagnostic performance of this expert diagnostic system. The parameters of these two loudspeakers are summarized in Table IV. Figure 9 shows the experimental arrangement inside an anechoic room, where the test conditions follow the IEC 60268-5, i.e., $U_1=4U_2$, $f_1=f_S$, $f_2=8.5f_1$. The bass tone voltage U_1 was chosen according to the rated power of the loudspeaker. A laser vibrometer and a microphone were used to measure cone displacement and sound pressure, respectively.

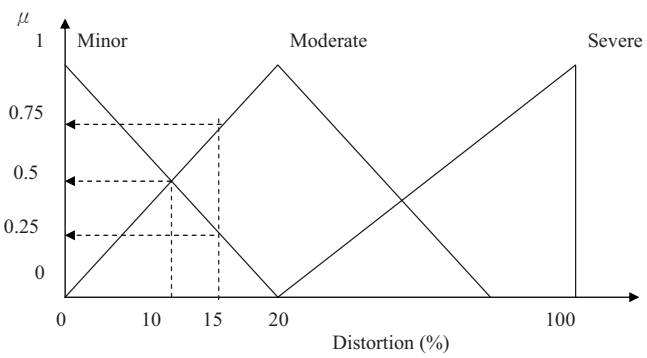


FIG. 7. The learning curve (error vs iteration) obtained in the training phase of the FNN system.

Nonlinear measurements obtained by the Klippel Distortion Analyzer 2 (Rev. 2.0)6 (Refs. 6 and 30) are summarized in Table V. Hence, the proposed FNN diagnostic system was applied to infer the nonlinear defects of the loudspeaker under test. The results are summarized in Table VI. To further explore the inference capability of the proposed FNN loudspeaker diagnostic system, a numerical simulation is undertaken. Although the finite element modeling is an alternative approach,[31] it is simply too time consuming to produce sufficient number of sample data. For the present multidomain

TABLE IV. Thermal and nonlinear parameters at the rest position.

Parameter	Driver A	Driver B	Unit
f_S	60	140	Hz
C_{MS} $(x=0)$	1.22	4.21	mm/N
Bl $(x=0)$	8.51	2.81	N/A
L_E $(x=0)$	1.20	0.25	mH
L_2 $(x=0)$	1.41	0.49	mH
M_{MS}	13.19	1.78	gm
Q_{MS}	1.91	4.21	
R_E	7.24	9.04	Ω
R_2 $(x=0)$	5.64	1.05	Ω
R_{TV}	14.05	14.96	K/W
R_{TM}	30.75	9.44	K/W
C_{TV}	2.56	0.97	J/K
C_{TM}	68.56	74.96	J/K
Normal rated power	5	3	W
S_D	78.54	28.27	cm^2

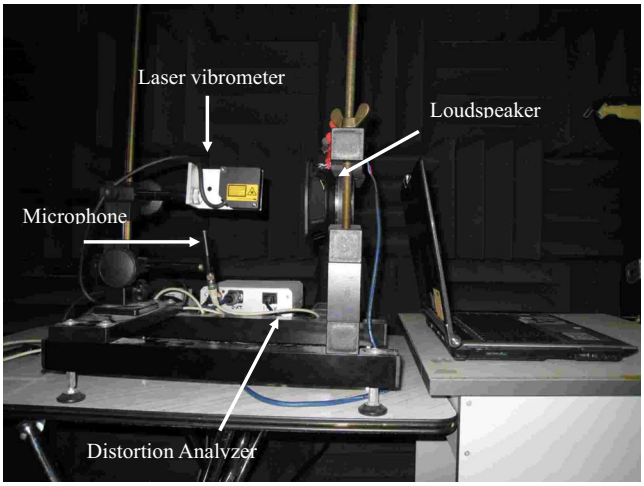


FIG. 9. (Color online) The experimental arrangement of the FNN loudspeaker diagnostic system.

problem that involves the electrical, mechanical, and acoustical subsystems, coupling these different physical domains results in a complicated set of matrix equations that are not trivial to solve simultaneously. We thus opt for more efficient analogous circuit approach in this paper. In the simulation, only one kind of nonlinearity is taken into account at a time, while keeping the remaining parameters the same as the rest position. By doing so, it is easier to validate the inference result and to assess the sensitivity of each nonlinear parameter.

A. Loudspeaker A

Driver A is a 10 cm diameter woofer with a resonance frequency of 60 Hz. Figure 10 shows the frequency response of the diaphragm displacement under the condition $U_1=6V$, where the displacement at 40 Hz is about 3 mm. The large signal parameters $Bl(x)$ and $C_{\mathrm{MS}}(x)$ measured by the distortion analyzer are shown in Figs. 11(a) and 11(b). In contrast to the nearly constant characteristics of $Bl(x)$ and $C_{\mathrm{MS}}(x)$, the voice coil inductance $L_E(x)$ shown in Fig. 11(c) has a distinct asymmetrical shape increasing toward the negative displacement.

Measured data in Table V reveals that Driver A suffers from high second-order intermodulation both in sound pressure and current. For this loudspeaker, the proposed FNN system inferred that an asymmetric $L_E(x)$ problem could exist with probability nearly 100% (Table VI). To validate this result, nonlinear simulations were conducted next.

In Fig. 12(a), the nonlinear simulation reveals that harmonic distortions are almost negligible ($<7\%$ above 50 Hz). On the other hand, a two-tone signal comprising a tone fixed at the frequency $f_1=60$ Hz and a tone with varying frequency f_2 is employed to excite intermodulations, as shown in Fig. 12(b). The third-order intermodulation components in sound pressure and current of Driver A are both very small ($<8\%$) in the wide frequency range. However, very high second-order intermodulations ($>20\%$) is clearly visible in Fig. 12(b). The fact that the intermodulation in current is comparable with that in sound pressure or the IMD increases at high frequencies suggests that the primary cause of nonlinearity could be asymmetry in $L_E(x)$.

Figure 12(c) shows a simulation result for the second-order intermodulation. Obviously, the asymmetry of the $L_E(x)$ (represented as a solid line marked with asterisks) is the dominant source of the second-order IMD because this curve reaches a level close to the response when all nonlinearities are involved.

In summary, the preceding analysis reveals that asymmetry in $L_E(x)$ is the dominant nonlinear cause, which agrees very well with the result inferred by the FNN system. A remedy to such problem is to place conductive materials such as conducting rings or caps made of aluminum or copper on the pole piece of the loudspeaker.[18]

B. Loudspeaker B

The driver B is a 6 cm woofer with a deviating voice coil. Figure 13 shows the frequency response of the diaphragm displacement under the condition $U_1=5$ V, where the maximum displacement is about 1.4 mm at 140 Hz. The $Bl(x)$ curve shown in Fig. 14(a) implies a distinct asymmetry, for the force factor decreases much faster at negative displacement than at positive displacement. Also, the compliance $C_{\mathrm{MS}}(x)$ shown in Fig. 14(b) is limited at the negative displacement because the variation of the compliance at $x=-1.5$ mm reduces to almost 55% of the maximum value. Comparing with the inductance of driver A reveals that the inductance of driver B is not that high but also has an asymmetrical characteristic, as shown in Fig. 14(c). Table V reveals that driver B suffers from high third-order harmonic distortion and the second- and third-order intermodulations. With FNN inference, high probabilities associated with coil offset, coil height, and symmetrical limiting of suspension summarized in Table VI indicate that there could be design flaws in the suspension and voice coil of driver B. To validate this result, nonlinear simulations were conducted next.

TABLE V. Distortion measures. Boldface indicates a significant distortion.

	X_{DC}	HD_2	HD_3	IMD_2	IMD_3	$\mathrm{IMD}_{2,i}$
Driver A	7.6%	3.8%	0.7%	**13.1%**	6.2%	**18.8%**
Driver B	8.5%	4.1%	**20.2%**	**26.0%**	**11.2%**	0.8%

The harmonic distortions measured at input voltage $U_1 = 5$ V are shown in Fig. 15(a). Clearly, the second-order harmonic distortion is not significant, whereas the third-order harmonic distortion exceeding 20% below the resonance frequency could degrade sound quality at low frequencies. Figure 15(a) shows a simulation result for the third-order harmonic distortion. At low frequencies, comparison between the measured result, the prediction considering all nonlinearities, and the prediction considering only nonlinearity of $C_{MS}(x)$ reveals that large third-order harmonic distortion is predominantly caused by symmetrical limiting of suspension. Above 200 Hz, disagreement between prediction and the measurement starts to show because the lumped parameter model fails to model the flexural modes of the diaphragm at high frequencies.

A two-tone stimulus with $f_1 = 140$ Hz, $f_2 = 900-3000$ Hz, $U_1 = 5$ V, $U_2 = U_1/4$ is utilized for measuring the intermodulations. In Fig. 15(b), the measured second-intermodulation and the third-intermodulation well exceed the 10% threshold. Figure 16(b) shows a simulation result for the third-order intermodulation. The measured data (dashed line) are nearly constant (10%) independent of the varying frequency f_2, which is a unique feature of intermodulations resulting from $Bl(x)$ nonlinearity. After switching off all nonlinearities and considering only the force factor $Bl(x)$, we obtain a prediction almost identical to the measurement. Thus, it can be concluded that the symmetric $Bl(x)$ nonlinearity is the primary source of the third-order IMD.

The simulation results of the second-order intermodulation are presented in Fig. 16(c). Again, the $Bl(x)$ nonlinearity contributes significantly to the second-order intermodulation by its critical asymmetry. Although the distortions due to nonlinear $L_E(x)$ and Doppler effect rise by 6 dB/octave, their importance is not as high as that of $Bl(x)$ nonlinearity below 3000 Hz.

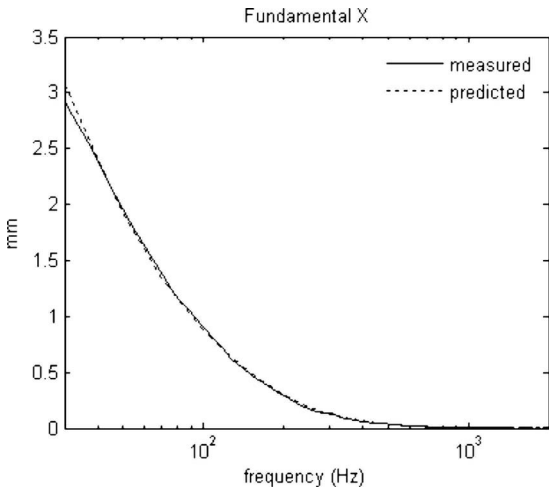


FIG. 10. Fundamental displacement x of driver A predicted (dotted line) and measured (solid line) at an input voltage of 6 V.

In summary, the diagnostic results of numerical simulation agree well with the results inferred by the FNN system. It can also be seen from this result that the FNN system is capable of dealing with not only single-cause but also multiple-cause problems.

V. CONCLUSIONS

A neural-fuzzy-based expert system has been developed for moving-coil loudspeaker diagnostics. The system was established in two stages. First, a cause-symptom database of loudspeaker nonlinearity was created on the basis of a large signal model. Second, the nonlinear causes and symptoms in the database were used as the input and output in training the neural-fuzzy system. A number of distortion indices were measured for the loudspeaker under test using a distortion analyzer. With the distortion indices as the input, the fault types of the loudspeaker can be determined using the FNN system, much like a human expert.

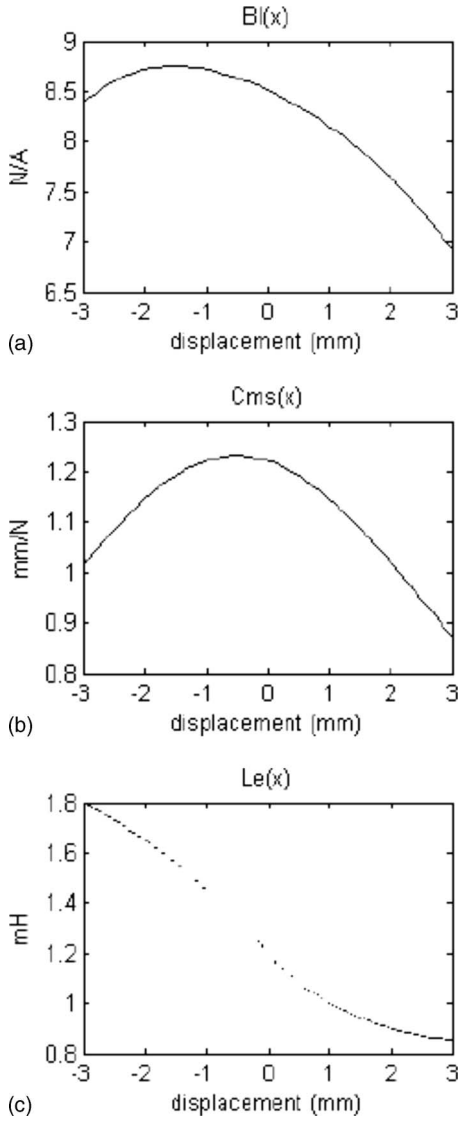


FIG. 11. Nonlinearities of driver A plotted versus displacement x. (a) Force factor $Bl(x)$. (b) Compliance $C_{MS}(x)$. (c) Inductance $L_E(x)$.

TABLE VI. Probability of defect inferred by the FNN diagnostic system. Boldface indicates a highly possible cause of nonlinearity.

	Coil offset	Coil height	Asymmetry in $C_{MS}(x)$	Suspension limiting	Asymmetry in $L_E(x)$	Doppler effect
Driver A	2%	0%	0%	0%	**99%**	2%
Driver B	**98%**	**100%**	26%	**98%**	0%	2%

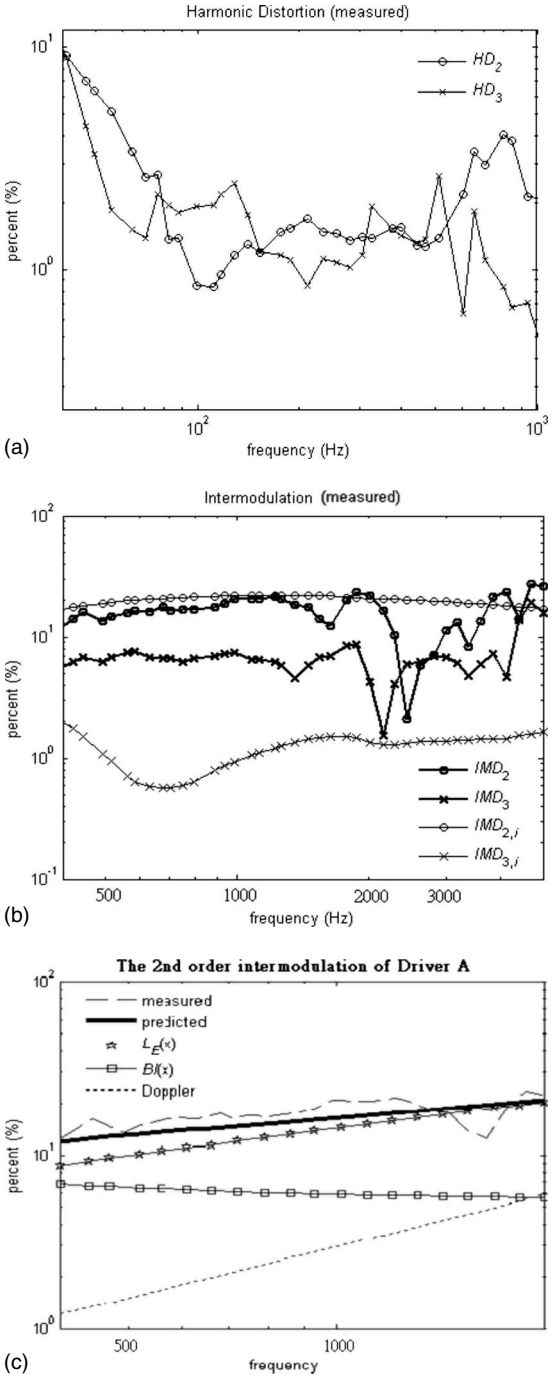


FIG. 12. The nonlinear distortions of driver A measured using the distortion analyzer. (a) The second- and third-order harmonic distortions. (b) The second- and third-order intermodulations. (c) The second-order intermodulations of driver A: measured data (dashed line), predicted data with all nonlinearities considered (solid line), and separate nonlinearities.

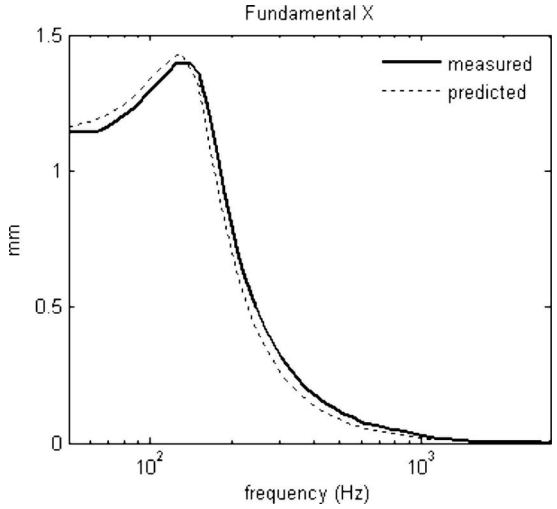


FIG. 13. Fundamental displacement x of driver B predicted (dotted line) and measured (solid line) at an input voltage of 5 V.

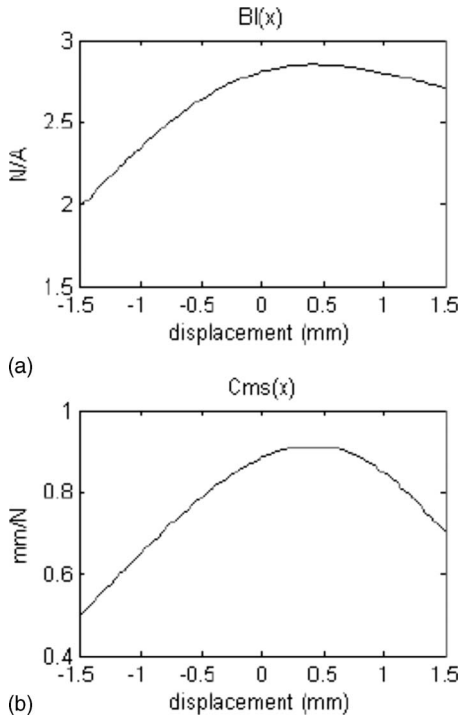


FIG. 14. Nonlinearities of driver B plotted vs displacement x. (a) Force factor $Bl(x)$. (b) Compliance $C_{MS}(x)$. (c) Inductance $L_E(x)$.

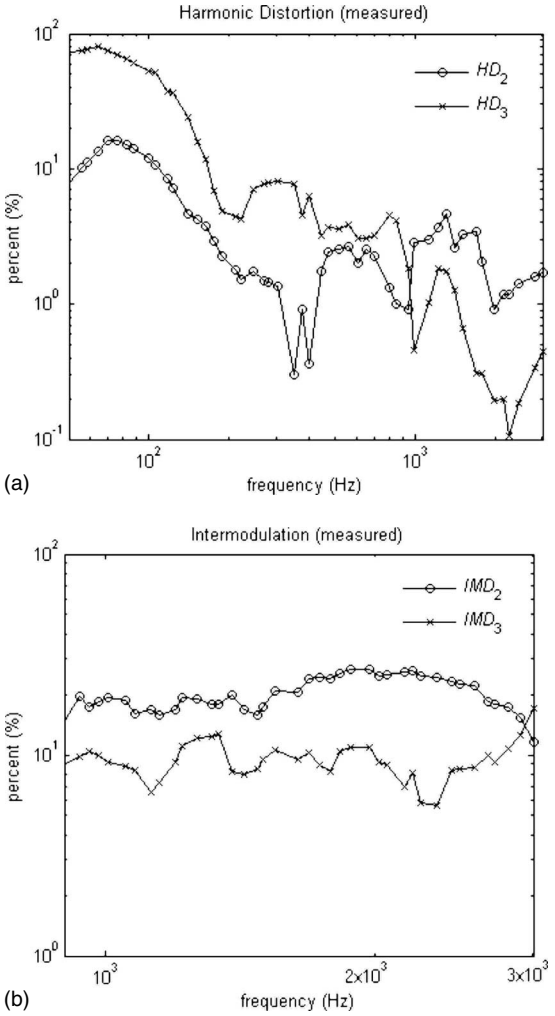


FIG. 15. The nonlinear distortions of driver B measured by the distortion analyzer. (a) The second- and third-order harmonic distortions. (b) The second- and third-order intermodulations.

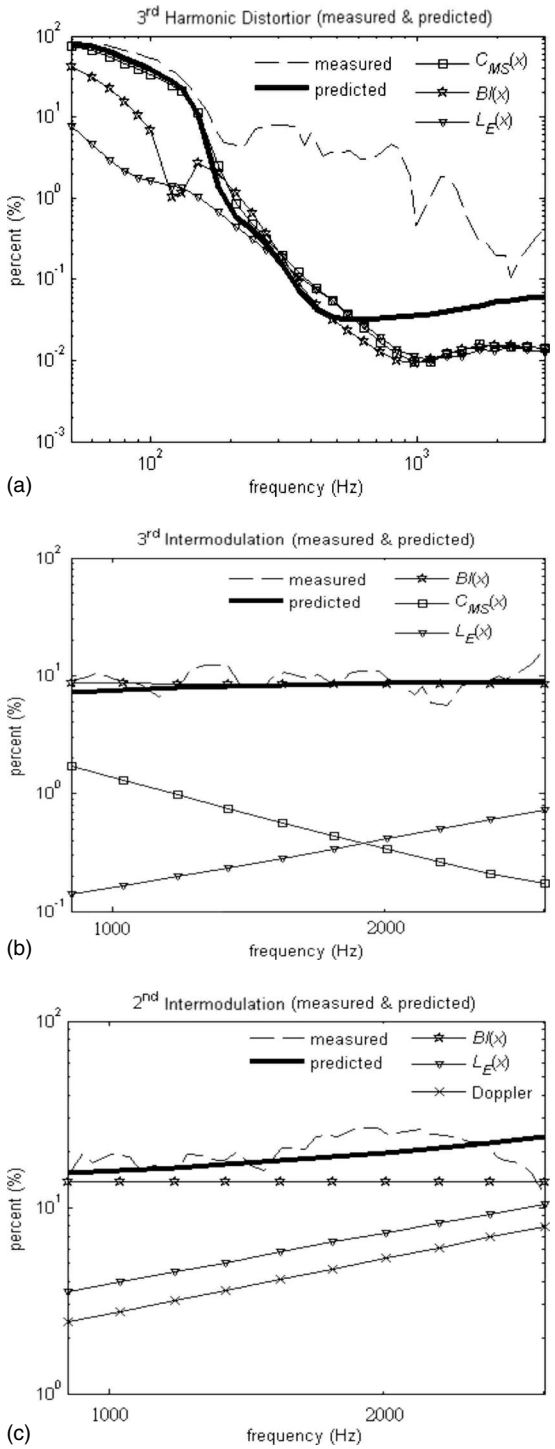


FIG. 16. Distortions of driver B measured (dashed line), predicted with all nonlinearities considered (solid line) and separate nonlinearities. (a) The third-order harmonic distortions. (b) The third-order intermodulations. (c) The second-order intermodulations.

This neural-fuzzy diagnostic system has been justified by experiments for two sample loudspeakers. The results inferred by the FNN diagnostic system was in good agreement with those obtained using numerical prediction. The results revealed that the proposed diagnostic system was capable of identifying single cause but also multiple causes of loudspeakers. This FNN system provides a more accurate and cost-effective solution of loudspeaker diagnostics than human experts. However, the present system is more cost effective than human experts and is only true if the diagnostics performed by a human expert is based on distortion measurements. However, the proposed methodology could still be of some value since systems for measuring the nonlinear loudspeaker parameters such as $B_l(x)$ and $C_{\mathrm{MS}}(x)$ are not widely available, as compared to distortion analyzers.

Several extensions of the present work are under way. The system is being converted to a fully on-line test bench, as required in quality control of mass production. The diagnostic capability of the system shall be enhanced by incorporating more fault types such as the rub and buzz problem and thermal buildup problem, etc., into the present system.

ACKNOWLEDGMENT

The work was supported by the National Science Council in Taiwan, Republic of China, under the Project No. NSC91-2212-E009-032.

[1]J. Borwick, *Loudspeaker and Headphone Handbook* (Focal, Oxford, UK, 1994).
[2]B. Elliott, "On the measurement of the low-frequency parameters of moving-coil piston transducers," 58th Convention of the Audio Engineer-

ing Society, New York, November, 1977.
[3]M. H. Knudsen, "Low-frequency loudspeaker models that include suspension creep," J. Audio Eng. Soc. **41**, 3–18 (1993).
[4]R. H. Small, "Direct-radiator loudspeaker system analysis," J. Audio Eng. Soc. **20**, 383–395 (1972).
[5]W. Klippel, "Nonlinear large-signal behavior of electrodynamic loudspeakers at low frequencies," J. Audio Eng. Soc. **40**, 483–496 (1992).
[6]W. Klippel, "Distortion analyzer—A new tool for assessing and improving electrodynamic transducer," presented at the 108th Convention of the Audio Engineering Society, Paris, 19–22 February, 2000.
[7]H. Huang, "Analysis of nonlinear behavior of loudspeakers using the instantaneous frequency," J. Acoust. Soc. Am. **113**, 2202 (2003).
[8]W. Klippel, "Measurement of loudspeaker parameters by inverse nonlinear control," J. Acoust. Soc. Am. **105**, 1359 (1999).
[9]W. Klippel, "Loudspeaker nonlinearities—Causes, parameters, symptoms," presented at 119th Convention of the Audio Engineering Society, New York, 7–10 October, 2005.
[10]W. Klippel, "Assessment of voice-coil peak displacement Xmax," J. Audio Eng. Soc. **51**, 307–324 (2003).
[11]D. Clark, "Precision measurement of loudspeaker parameters," J. Audio Eng. Soc. **45**, 129–141 (1997).
[12]J. D'Appolito, *Testing Loudspeakers* (Audio Amateur, New York, 1998).
[13]W. Klippel, "Prediction of speaker performance at high amplitudes," presented at the 111th Convention of the Audio Engineering Society, New York, 21–24 September, 2001.
[14]A. J. M. Kaizer, "Modeling of the nonlinear response of an electrodynamic loudspeaker by a Volterra series expansion," J. Audio Eng. Soc. **35**, 421–433 (1987).
[15]W. Klippel, "Nonlinear modeling of the heat transfer in loudspeakers," J. Audio Eng. Soc. **52**, 3–25 (2004).
[16]W. Klippel, "Dynamic measurement and interpretation of the nonlinear parameters of electrodynamic loudspeakers" J. Audio Eng. Soc. **38**, 944–955 (1990).
[17]E. R. Olsen and K. B. Christensen, "Nonlinear modeling of low frequency loudspeakers—A more complete model," The 100th Convention Audio Engineering Society, Copenhagen, 11–14 May, 1966.
[18]W. Klippel, "Diagnosis and remedy of nonlinearities in electrodynamical transducers," presented at the 109th Convention of the Audio Engineering Society, Los Angeles, 21–25 September, 2000.
[19]P. Jackson, *Introduction to Expert Systems* (Addison-Wesley, Reading, MA, 1986).
[20]L. A. Zadeh, "Fuzzy sets as a basis for a theory of possibility," Fuzzy Sets Syst. **1**, 3–28 (1978).
[21]T. Kohonen, "An introduction to neural computing," Neural Networks **1**, 3–16 (1988).
[22]C. T. Lin and C. S. George Lee, *Neural Fuzzy Systems* (Prentice-Hall, Englewood Cliffs, NJ, 1966).
[23]M. Becraft and R. Isermann, "Neuro-fuzzy systems for diagnosis," Fuzzy Sets Syst. **89**, 289–307 (1997).
[24]M. R. Bai, I. L. Hsiao, H. H. Tsai, and C. T. Lin, "Development of an on-line diagnosis system for rotor vibration via model-based intelligent inference," J. Acoust. Soc. Am. **107**, 315–323 (2000).
[27]C. F. Juang and C. T. Lin, "An on-line self-constructing neural fuzzy inference network and its applications," IEEE Trans. Fuzzy Syst. **6**, 12–13 (1998).
[28]J. H. Mathews and K. K. Fink, *Numerical Methods Using Matlab* (Prentice-Hall, Upper Saddle River, NJ, 1966).
[25]W. Klippel, "Measurement of impulsive distortion, rub and buzz and other disturbances," Klippel GmbH, www.klippel.de (Last viewed June 2008).
[26]S. Temme, "Are you shipping defective loudspeakers to your customers?," http://www.listeninc.com/site/notes.html (Last viewed September 2008).
[29]"Sound system equipment. Part 5: Loudspeakers," IEC Publication No. 60268-5.
[30]Specification of the Klippel Analyzer System, Klippel GmbH, www.klippel.de (Last viewed June 2008).
[31]M. Rausch, "Optimization of electrodynamic loudspeaker-design parameters by using a numerical calculation scheme," Acust. Acta Acust. **85**, 412–419 (1999).

The eigenspectra of Indian musical drums

G. Sathej
Department of Biotechnology, Indian Institute of Technology Madras, Chennai 600036, India

R. Adhikari
The Institute of Mathematical Sciences, CIT Campus, Tharamani, Chennai 600113, India

(Received 29 September 2008; revised 10 November 2008; accepted 5 December 2008)

In a family of drums used in the Indian subcontinent, the circular drum head is made of material of nonuniform density. Remarkably, and in contrast to a circular membrane of uniform density, the low eigenmodes of the nonuniform membrane are harmonic. In this work the drum head is modeled as a nonuniform membrane whose density varies smoothly between two prescribed values. The eigenmodes and eigenvalues of the drum head are obtained using a high-resolution numerical method. The mathematical model and the numerical method are able to handle both concentric and eccentric nonuniformities, which correspond, respectively, to the *dayan* and the *bayan* drums. For a suitable choice of parameters, which are found by optimizing the harmonicity of the drum, the eigenspectra obtained from the model are in excellent agreement with experiment. The model and the numerical method should find application in numerical sound synthesis.
 [DOI: 10.1121/1.3058632]

PACS number(s): 43.40.Dx, 43.75.Hi, 43.75.Wx [TDR] Pages: 831–838

I. INTRODUCTION

The eigenvalue problems for a string and a uniform circular membrane are classical problems in mathematical physics. The eigenvalues of a string are determined by the zeros of the sine function and so form a harmonic series. The large number of harmonic overtones give the vibrations of a string its musicality. The eigenvalues of a uniform membrane, on the other hand, are determined by the zeros of Bessel functions. The overtones are not integer multiples of the fundamental. Consequently, the vibrations do not have a strong sense of pitch and, therefore, lack the musicality of string vibrations.

Several musical traditions have devised means of restoring musicality to the vibrations of circular drums. The Western tympani achieves this by coupling the vibrations of the membrane with the large mass of air enclosed in the kettle below the drum head. For a judicious choice of modes, the combined membrane-air system has harmonic vibrations.[1] A different strategy is used in a whole family of drums used in the Indian subcontinent, where harmonic overtones are obtained by loading the central part of the membrane with material of heavier density. These drums have a strong sense of pitch and, in performance, are tuned to match the tonic of the vocalist or the instrumentalist.

The two most popular drums of this family are the South Indian *mridangam* and the North Indian *tabla*. The mridangam is a single drum covered on both sides with drum heads made of leather, while the tabla is a pair of drums, the *dayan* and the *bayan*, each of which have a single drum head (Fig. 1). The loading in the dayan is concentric to the membrane, while in the bayan the loading is eccentric (Fig. 2).

Raman[2,3] made the first scientific study of this family of drums. In a series of experiments, Raman and co-workers obtained the eigenmodes and eigenvalues of the mridangam, showing that the first nine normal modes gave five very nearly harmonic tones. The higher overtones were noticeably anharmonic, but Raman noted features in the construction of design to suppress the higher overtones. Subsequently, Ramakrishna and Sondhi[4] modeled the drum head as a composite membrane of two distinct densities, with the caveat that "the density of the loaded region is not constant… but decreases gradually." With this simplification, and for concentric loading, the eigenvalue problem could be solved analytically in terms of Bessel and trigonometric functions. The eigenvalues of the composite membrane model agree with Raman's experimental values to within 10%. Solving the composite membrane model for eccentric loading is considerably more difficult due to lack of circular symmetry. An exact solution for the eigenmodes in terms of known functions is not available. Two approximate solutions[5,6] have been presented, but the agreement with experimental values is generally poor. Little, therefore, is known about the eigenspectrum of the eccentrically loaded drum head.

The purpose of this work is twofold. The first is to present a mathematical model for the loaded drum head of the Indian musical drums, using the tabla as the prototypical example. The second is to present a high-resolution numerical method, based on Fourier–Chebyshev collocation, which may be used to obtain the eigenvalues and eigenmodes of a nonuniform circular membrane with an arbitrary variation in the mass density. Using the numerical method we obtain the eigenspectrum of our model drum head for both concentric and eccentric loadings. For concentric loading, our results are in excellent agreement with Raman's experimental values and offer an improvement over the composite membrane model of Ramakrishna and Sondhi. For the eccentric case, our numerical results give an accurate solution for the eigenvalues and eigenmodes and do not require the uncontrolled simplifying assumptions of previous work. We compare the eigenspectra of the concentric and eccentric drum heads and show that the eccentricity lifts the degeneracy of pairs of

FIG. 1. (Color online) The *tabla* is a pair of drums consisting of the left drum, *bayan*, and the right drum, *dayan*. The cords running along the length of the drums are used to adjust the tension in the membrane, allowing the tonic of the drum to be raised or lowered.

concentric eigenmodes. With further refinement, which is part of ongoing work, we believe that our model will find application for numerical sound synthesis of the tabla and other Indian musical drums.

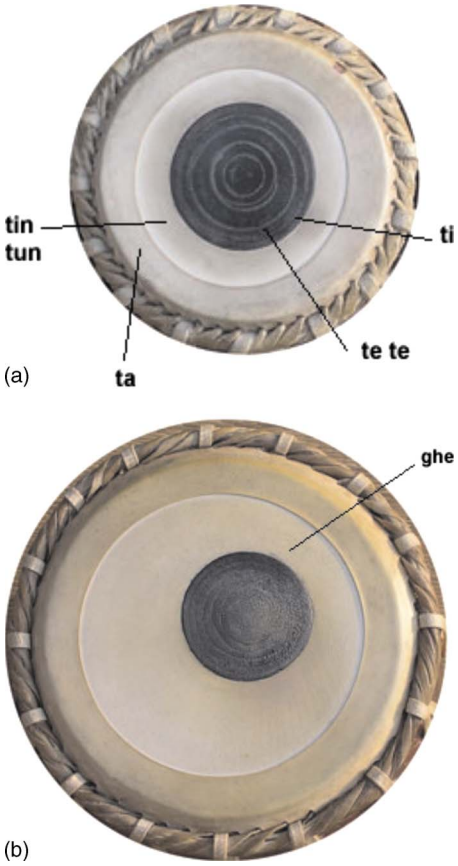


FIG. 2. (Color online) The drum head of the dayan (a). The drumhead is made with goatskin and loaded at the center. The central loading patch, the *sihai*, is made of a paste of soot, iron filings, and flour and applied layer by layer to the goatskin membrane. The sihai is cracked with a heavy stone after the paste dries to reduce the rigidity of the material. The thin flap at the outer edge of the drumhead, the *kinar*, serves to damp the higher harmonics. Distinct sounds, indicated by the mnemonic syllables *tin, tun, ta, ti*, and *tete*, are produced when the drum is struck at different parts. The bayan (b) is of similar construction, but crucially, the sihai is placed eccentrically on the membrane. The bayan has a smaller range of sounds, the principal one being the mnemonic syllable *ghe*. The pitch of this syllable is modulated by changing the position of the heel of the hand along the broadest part of the unloaded region of the drum head.

The remainder of this paper is organized as follows. In Sec. II we present our mathematical model for the continuous loading and the boundary value problem that must be solved to obtain the eigenvalues and the eigenfunctions. In Sec. III we discuss in detail our numerical method, discussing, in particular, how it leads to a generalized eigenvalue problem. Our results for concentric and eccentric loadings are presented in Sec. IV. We conclude with a summary and discussion of further work.

II. MATHEMATICAL MODEL

The drum head of the tabla is made of leather with the central patch (the *sihai*) made of a complex mixture of materials, as explained in the caption to Fig. 2. The sihai is approximately eight times as dense as the leather and covers approximately a quarter of the area of the membrane. The sihai is applied in layers, with each layer made to dry completely before the application of the next layer. This allows a control of the effective mass density of the sihai. The variation in the harmonicity of the drums with each layer of application of the paste has been studied carefully by Rossing and Sykes.[7]

It should be clear that the construction of the sihai is a complex art. However, the most crucial effect of the sihai is to increase the density in the central region of the drum head. *Effectively*, it is possible to think of the drum head, then, as a membrane of nonuniform density. This forms the basis of our mathematical model. We approximate the tabla drum head as a circular membrane of unit radius, with a nonuniform areal density $\rho(\mathbf{r})$, where $\mathbf{r}=(r,\theta)$ is a point on the membrane. Our specific model, which includes both the concentric and eccentric situations, is

$$\frac{\rho(r,\theta)}{\rho_2} = 1 + \frac{(\sigma^2-1)}{2}\left[1 - \tanh\left(\frac{R(r,\theta)-k}{\xi}\right)\right], \tag{1}$$

where

$$R(r,\theta) = \sqrt{(r\cos\theta-\epsilon)^2 + (r\sin\theta)^2}. \tag{2}$$

This function changes smoothly from a value ρ_2 at $r=1$ to a value $\rho_1=\rho_2\sigma^2$ at the center of the loaded region. The change occurs over a region of width ξ along the circle whose equation in polar coordinates is $r=R(r,\theta)$. For $\epsilon=0$ the loading is radially symmetric and represents the concentric loading of the dayan. For $\epsilon>0$, the loading is displaced from the center by a distance ϵ and then represents the eccentric loading of the bayan. For $\xi \ll 1$, k is the ratio of the radii of the loaded region and the complete circular membrane, requiring $0<k<1$. For the concentric case, the variation in density is centrally symmetric, and we display the variation as a function of the radius in Fig. 3. In the eccentric case, the density depends on both the radius and the polar angle, and Fig. 4 illustrates this case. Notice that $\sigma=1$ corresponds to the uniform membrane, while for fixed σ, k, and ϵ, $\xi\to 0$ recovers the composite membrane model. Our model, which is smoothly nonuniform, allows a gradual decrease in density of the loaded region and avoids the abrupt change in density of the composite membrane model. Previous attempts at modeling the drum head by continuous densities have all

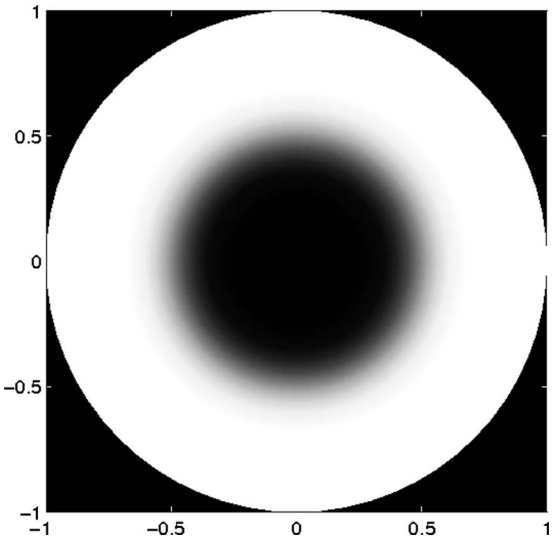


FIG. 3. Variation in areal density of the membrane for $\epsilon=0$. The density is plotted for $\sigma=2.57$, $k=0.492$, and $\xi=0.091$. We use this form of the density to model the dayan.

focused on the concentric case. Some of these use unphysical models for the mass density,[8,9] while others need parameters k and σ which do not agree with experiment.[10]

The equation of motion governing a membrane with spatially varying density $\rho=\rho(r,\theta)$ and uniform tension T is

$$\rho\ddot{u} = T\nabla^2 u. \tag{3}$$

Here, $u=u(r,\theta,t)$ is the transverse displacement of the membrane at time t. For a circular membrane of unit radius clamped at the boundary, the eigenvalue problem is obtained by seeking solutions of Eq. (3) which satisfy the Dirichlet condition

$$u(r=1,\theta,t)=0. \tag{4}$$

The initial-boundary value problem represented by Eqs. (3) and (4) has exact analytical solutions in only a handful of special cases. Of these, the most relevant for the present work are the uniform membrane $\rho(r,\theta)=\rho_0$ and the composite membrane model. To the best of our knowledge, there are no exact analytical solutions when the density varies with both the radius and the angle, as is in the case of Fig. 4. Due to the lack of circular symmetry, the usual strategy of separation of variables in polar coordinates fails, and the eigenfunctions cannot be obtained by a Fourier–Bessel expansion.

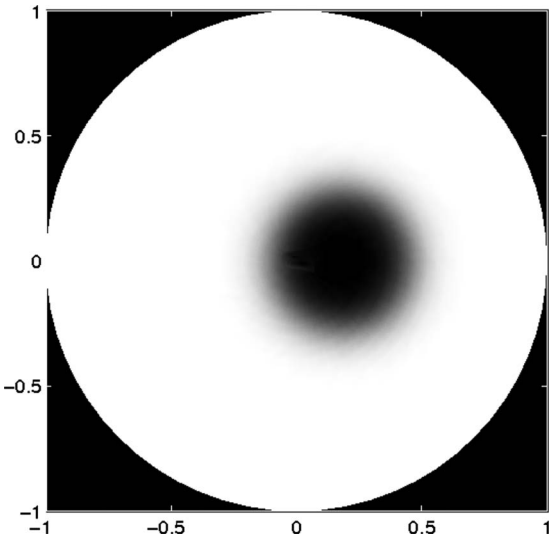


FIG. 4. Variation in areal density of the membrane for $\epsilon>0$. The density is plotted for $\epsilon=0.18$, $\sigma=2.57$, $k=0.29$, and $\xi=0.091$. We use this form of the density to model the bayan.

This motivates the use of a high-resolution numerical method which we describe in Sec. III. The advantage of the method, besides its accuracy, is that it solves with equal ease the eigenvalue problem for both the concentric and eccentric cases, facing no difficulty with density variations which depend on both radius and angle. The numerical method also opens the way to a time domain solution for Eq. (3) which should find application in numerical sound synthesis.

III. NUMERICAL METHOD

The eigenvalue equation for the normal modes of the loaded drum is obtained by assuming a solution $u(r,\theta,t)=\Psi_{mn}(r,\theta)\exp(i\omega_{mn}t)$ which transforms Eq. (3) into

$$-\omega_{mn}^2\rho(r,\theta)\Psi_{mn}(r,\theta) = T\nabla^2\Psi_{mn}(r,\theta). \tag{5}$$

Here, Ψ_{mn} is the eigenmode with m nodal lines and n nodal contours. This generalizes the labeling used for the uniform circular membrane, where the nodal lines are diameters and the nodal contours are circles. There is an important difference in the mathematical structure of the eigenvalue problem for the uniform and nonuniform membranes. Since the density variation is dependent on position, it cannot be scaled out as in the case of the uniform membrane. This leads to a *generalized* eigenvalue problem. The eigenfunctions and eigenvalues are functionals of the nonuniform density $\rho=\rho(r,\theta)$.

A direct numerical solution of Eq. (5) is possible using methods of varying degrees of accuracy and sophistication. Spectral collocation methods appear to offer the greatest accuracy for the least computational expense for this class of problems. We have therefore used a Fourier–Chebyshev spectral collocation technique, which we describe below, to study the generalized eigenvalue problem in Eq. (5).

A spectral collocation method proceeds by choosing a set of orthogonal functions and approximating the solution in terms of a linear combination of the orthogonal functions. The approximating function is in the form of an interpolant and matches the solution exactly at a specially chosen set of nodes, the so-called collocation points. For example, for a function $v(x)$ which is periodic, the appropriate spectral basis is the set of trigonometric functions. The function $v(x)$, sampled at the N points $x_1,\ldots,x_N$ a distance h apart, can be interpolated by

$$p_N(x) = \frac{1}{2\pi}\sum_{k=-N/2}^{N/2}{}'' \exp(ikx_j)\hat{v}_k, \tag{6}$$

where the primed summation indicates that terms $k=\pm N/2$ are multiplied by $1/2$ and $\hat{v}_k$ is given by

$$\hat{v}_k = h\sum_{j=1}^{N} \exp(-ikx_j)v(x_j). \tag{7}$$

Then, the spectral derivative of $v(x)$, as estimated from its values $v(x_j)$ at the sample points, is simply the derivative of the spectral interpolant, evaluated at the sample points,

$$\left(\frac{dv}{dx}\right)_{\text{sp}} = p_N'(x)|_{x_j}. \tag{8}$$

The set of samples of $v(x_j)$ can be thought of as a vector $\mathbf{v} = (v_1, \ldots, v_N)$. Likewise, the spectral derivative can also be thought of as a vector with N components. The derivative vector can then be obtained from the function vector by a matrix multiplication, where the matrix entries are constructed from the derivative of the interpolant function $p_N(x)$. This idea of constructing a spectral interpolant and differentiating it to obtain an estimate of the derivative generalizes to two or more independent variables. Partial derivatives are obtained by Kronecker products of the differentiation matrices corresponding to each of the independent variables. In Fourier–Chebyshev spectral collocation, a Fourier expansion, as above, is used for the angular coordinate $\theta \in [0, 2\pi]$ and a Chebyshev expansion is used for the radial coordinate $r \in [0, 1]$. The usual Chebyshev expansion is for functions in $[-1, 1]$ and several methods exist for using the Chebyshev expansion for the radial coordinate. Here, we follow the method proposed by Fornberg,[11] using the implementation of Trefethen.[12] The Laplacian in polar coordinates,

$$\nabla^2 = \partial_r^2 + r^{-1}\partial_r + r^{-2}\partial_\theta^2, \tag{9}$$

is then replaced by the Fourier–Chebyshev differentiation matrix

$$L = (D_1 + RE_1) \otimes I_l + (D_2 + RE_2) \otimes I_r + R^2 \otimes D_\theta^{(2)}. \tag{10}$$

For N_r (odd) Chebyshev collocation points and N_θ (even) Fourier collocation points, the matrices above are representations of the partial derivates on the grid. The two terms with the matrices D_1 and D_2 represent ∂_r^2, the two terms with the matrices E_1 and E_2 represent $r^{-1}\partial_r$, and the last term is the representation of $r^{-2}\partial_\theta^2$. R is the diagonal matrix $\text{diag}(r_j^{-1})$, $1 \leqslant j \leqslant (N_r - 1)/2$. The two identity matrices

$$I_l = \begin{pmatrix} I & 0 \\ 0 & I \end{pmatrix}, \tag{11}$$

$$I_r = \begin{pmatrix} 0 & I \\ I & 0 \end{pmatrix} \tag{12}$$

are formed out of the $N_{\theta/2} \times N_{\theta/2}$ identity matrix I. The somewhat complicated looking expression for the Laplacian arises from the use of Fornberg's prescription for handling the radial coordinate using a Chebyshev expansion. Further details are available in Ref. 12. The loading function $\rho(r, \theta)$ is itself represented by a matrix $\mathbf{B}$ and the generalized eigenvalue problem can then be formulated as a matrix equation

$$\mathbf{L} \cdot \Psi = -\lambda^2 \mathbf{B} \cdot \Psi. \tag{13}$$

We solve for the eigenvectors Ψ_{mn} and the eigenvalues λ_{mn} using the MATLAB function eigs which is based on a Cholesky decomposition algorithm.

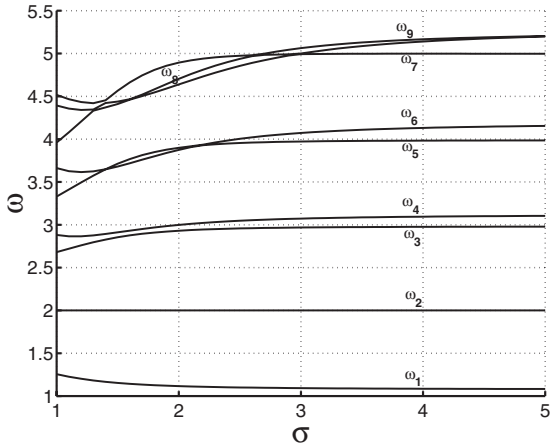


FIG. 5. Variation in eigenfrequencies with mass density ratio σ for the concentric case, $\epsilon = 0$, for fixed value of radius ratio $k = 0.4$ and fixed smoothness parameter $\xi = 0.091$. The frequencies are normalized by the first overtone. The frequencies are very close being in integer ratios around $\sigma = 3.0$. For large σ, the frequencies essentially remain constant, but there are several modes that are no longer harmonic.

We have benchmarked our code with the known eigenvalues of the uniform circular membrane and find spectral convergence with increase in the number of modes. For $N_r = 31$ and $N_\theta = 20$ the eigenvalues are accurate to ten decimal places. For the nonuniform membrane, higher number of modes is needed, especially in the radial direction, to capture the rapid variation in density for $\xi \ll 1$. The results reported below have $N_r = 65$ and $N_\theta = 30$ for the concentric case and $N_r = 65$ and $N_\theta = 56$ for the eccentric case unless stated otherwise. Note that with this choice, the accuracy of the numerical method is several decimal places more than the best reported experimental values. Thus when comparing with experiment, eigenvalues obtained from the numerical solution can be safely attributed to the model itself and not to numerical errors.

IV. RESULTS

We now present our results for the numerical solution of the generalized eigenvalue problem for the smoothly nonuniform membrane. Recall that our model density distribution has four parameters; the ratio of areal densities σ, the ratio of radii k, the eccentricity parameter ϵ, and the smoothness parameter ξ. We first present results for $\epsilon = 0$, which models the concentric loading of the dayan, followed by results for $\epsilon > 0$, which models the eccentric loading of the bayan.

A. Concentric loading

In Fig. 5 we show the variation in the eigenfrequencies of the first nine eigenmodes as the density contrast is increased at fixed values of k and ξ. All frequencies are normalized by the frequency of the first overtone. The frequency ratios rapidly depart from those of the uniform circular membrane ($\sigma = 1$) to attain harmonic ratios in the neighborhood of $\sigma = 3$. Our numerical results suggest that for very large σ the ratios do not depend on σ but have several modes which are

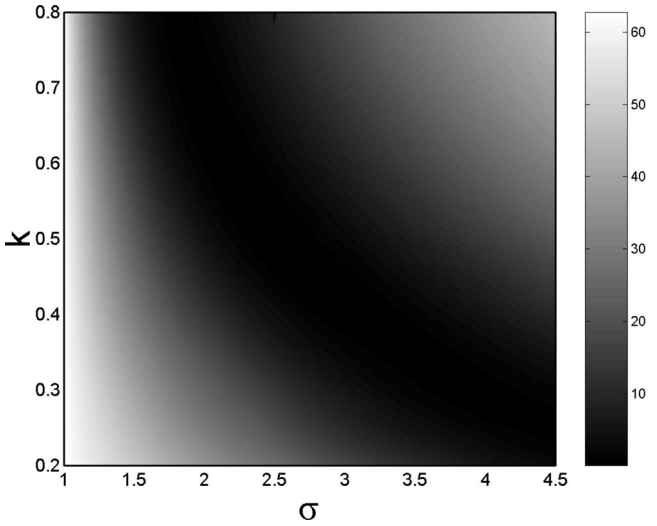


FIG. 6. The quality function $Q(\sigma,k)$, defined in the text, as a function of σ and k. Darker regions correspond to more harmonic vibrations. The most harmonic vibrations are obtained for $\sigma=2.57$ and $k=0.492$.

no longer harmonic. There is, then, an optimum value of σ around $\sigma=3$ which gives maximally harmonic vibrations. The absolute values of the frequencies decrease monotonically with an increase in the loading, as has been observed previously in analytical and experimental work.

To obtain the values of σ and k which produce a maximally harmonic drum, we define a quality function $Q(\sigma,k)$ which measures the squared deviation of the frequency of the ith eigenmode from its *nearest* integer value.

$$Q(\sigma,k)=\sum_{i=1}^{N_{\max}}(\omega_i(\sigma,k)-h_i)^2. \tag{14}$$

Here ω_i denotes the eigenvalue of the ith eigenmode, where i is the rank of the eigenmode when sorted in ascending order of eigenvalue. Eigenvalues from $i=1$ to $i=N_{\max}$ are used in calculating the quality. h_i denotes the nearest integer multiple of the fundamental corresponding to frequency ω_i. Smaller values of $Q(\sigma,k)$ correspond to more harmonic vibrations. Such an optimization has also been looked at by Gaudet *et al.*[13]

Thus, to find the best values of σ and k at fixed smoothness ξ, we are left with a two parameter optimization problem. We scan the (σ,k) parameter space over the range $1\leqslant\sigma\leqslant5$ and $0.2\leqslant k\leqslant0.8$ and obtain the values of $Q(\sigma,k)$ with $N_{\max}=15$. We show the result as a pseudocolor plot in Fig. 6. In this range, we find that Q has a minimum of $Q_{\min}=0.027$ for $\sigma_{\text{opt}}=2.57$ and $k_{\text{opt}}=0.492$. These values for the density and radii ratios are better when compared with $Q=0.074$ for $\sigma=3.125$, $k=0.4$ for the composite membrane model. Our optimum values, which are obtained without any fitting parameters, fall well within the range of $2.5<\sigma<4$, $0.45<k<0.55$ that are actually used in the construction of the dayan. Having obtained the optimal values of σ and k we further optimize on the smoothness parameter to find a value of ξ_{opt} of 0.091.

In Table I we compare the eigenvalues as determined from experiment, from the composite membrane model, and the present model with smooth nonuniformity. We restrict the comparison to the first nine eigenmodes which Raman identified as the being most harmonic. To make a like-for-like comparison with the composite membrane, we use values $\sigma=3.125$ and $k=0.4$. Apart from mode ψ_{03}, the present model provides a better fit to experimental values than the composite membrane model.

TABLE I. Comparison of eigenfrequencies of the first nine eigenmodes of the dayan, the composite membrane model (Ref. 4), and the smoothly nonuniform membrane model presented in this work for $\sigma=3.125$ and $k=0.4$. The frequencies are normalized by the first overtone. The figures in parentheses indicate the deviation, in cents, from the experimental value.

Mode	Experimental	Composite	Smooth
ψ_{01}	1.03	1.0309 (+1.51)	1.0345 (+7.55)
ψ_{11}	2.00	2.0000 (0.00)	2.0000 (0.00)
ψ_{21}	3.00	3.0412 (+23.61)	3.0393 (+22.53)
ψ_{02}	3.00	3.1546 (+86.99)	3.0534 (+30.54)
ψ_{31}	4.00	4.0928 (+39.70)	4.0086 (+3.72)
ψ_{12}	4.00	4.2268 (+95.47)	4.1463 (+62.18)
ψ_{03}	5.04	4.9794 (−20.94)	4.7784 (−92.27)
ψ_{41}	5.03	5.1134 (+28.47)	5.0023 (−9.56)
ψ_{22}	5.08	5.3093 (+76.43)	5.2491 (+56.69)

Going beyond the first nine modes, we find that our model continues to produce harmonic overtones while the composite membrane model shows significant deviations from harmonicity. This is seen clearly in Fig. 7 where we compare the first 15 eigenvalues for our model and the composite membrane model for $\sigma=2.57$ and $k=0.492$. We can, therefore, conclude that the gradual change in density of the loaded region, included here but absent from the composite membrane model, does have an appreciable effect on the musicality of the drum. In Fig. 8 we show the nodal lines for the first 20 eigenmodes, including the degenerate ones. These are similar to the nodal lines of a uniform circular membrane.

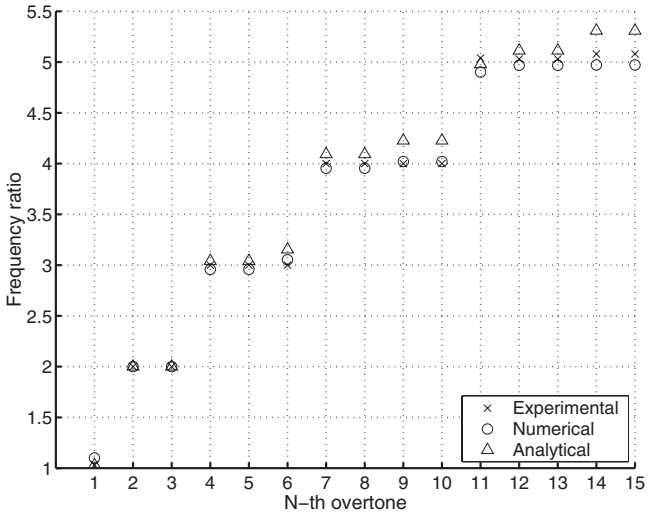


FIG. 7. Comparison of eigenfrequencies of the first 15 eigenmodes of the dayan, the composite membrane model (Ref. 4), and the smoothly nonuniform membrane model presented in this work. The higher overtones are clearly more harmonic when a gradual variation in density of the loaded region is allowed. The parameters used for the smoothly nonuniform membrane are the optimum values of $\sigma=2.57$, $k=0.492$, and $\xi=0.091$.

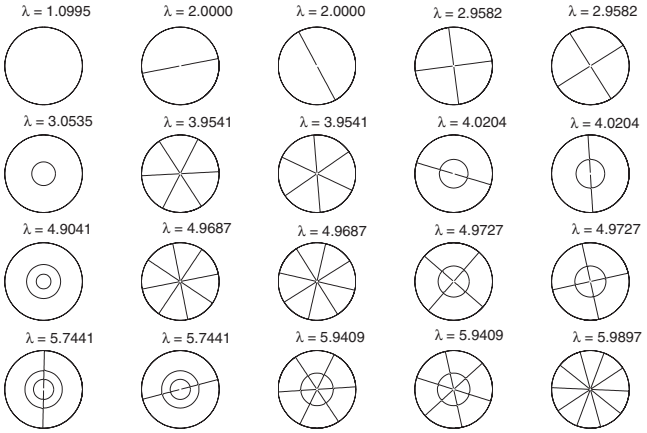


FIG. 8. Nodal contours for the first 20 eigenmodes of the dayan. The plots are with $\sigma=2.57$, $k=0.492$, and $\xi=0.091$.

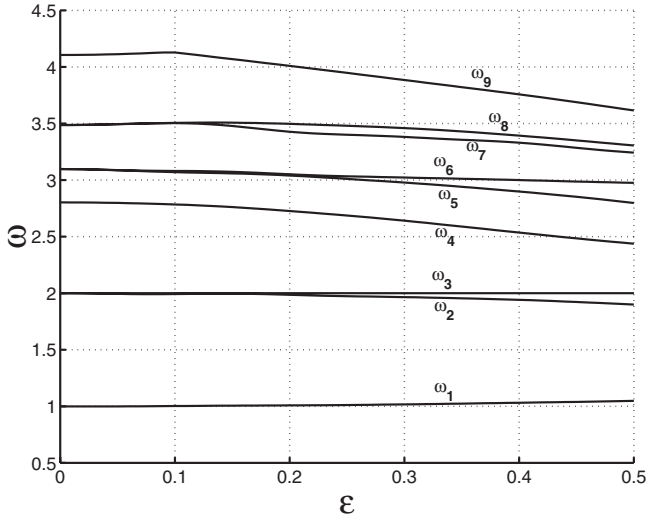


FIG. 9. Variation in eigenvalues with eccentricity for $k=0.29$, $\sigma=3.125$, and $\xi=0.091$. The lower eigenvalues remain unchanged for moderate eccentricities.

Returning to Fig. 6, we note that there is a distinct valley of small values in the pseudocolor plot of $Q(\sigma,k)$. This implies that there are many pairs of values of σ and k which allow for vibrations that are by and large harmonic. There is, indeed, a wide variation in σ and k in the dayan from different instrument makers, typically in the range $2.5<\sigma<4$ and $0.45<k<0.55$. It is tempting to speculate if this could have helped the early makers of the Indian drums, no doubt proceeding by empirical trial-and-error effort, at reaching the optimal values of σ and k. It should also be mentioned that several Indian musical drums, notably the mridangam, often use a temporary loading of flour paste (which is applied at the start of the performance and removed afterward) to ensure harmonic vibrations. It is likely that the principle of central loading was discovered by such temporary application of a heavier material, and later evolved into the more elaborate permanent loading of the sihai.

B. Eccentric loading

The main function of the eccentric loading in the bayan is to allow for modulations in the pitch of the drum dynamically, that is, while it is being played. The heel of the hand is moved back and forth along the diameter passing through the centers of the membrane and the eccentrically placed sihai to modulate the pitch. In this way, the bayan can produce a distinct sound, not found in any of the other Indian musical drums. In their experimental measurement of the eigenfrequencies of the tabla, Banerjee and Nag[14] noted that only the first few modes are excited by the player's action. The requirement appears to be, then, to allow an eccentric placement of the sihai and yet retain the harmonicity of the lower modes of vibrations.

In Fig. 9, we show how eigenvalues depend on the eccentricity ϵ at fixed values of $\sigma=3.125$, $k=0.29$, and $\xi=0.091$. We see that for eccentricities up to 0.1, there is hardly any variation in the eigenspectrum. For larger eccentricity the higher eigenmodes become anharmonic faster than the lower eigenmodes. This is completely consistent with the observation of Nag *et al.* Our present model thus captures this important feature of the eccentric loading.

It is worthwhile comparing our numerical results with two prior approximate analytical calculations. In Table II we compare the eigenvalues as obtained from experiment, an approximate calculation based on the composite membrane model,[5] and the present model. The agreement with experimental values is very good, except for eigenmodes which have one nodal circle. We believe that our numerical results are accurate for these modes but are yet to understand why Ramakrishna's approximate calculation produces better fits to the experimental data. We note that a similar divergence between the model and experiment has been noted in the work by Sarojini and Rahman,[6] where a variational method was used to calculate the eigenvalues.

In Table III we compare our model with the variational calculation of the eigenvalues of the composite membrane. This comparison is of methodological interest only since the values used do not correspond to the actual values used in the construction of the bayan. We note that the agreement is generally not very good, indicating that the variational method possibly overestimates the eigenvalues in this case.

In Fig. 10, we show the nodal contours of the first 20

TABLE II. Comparison of eigenfrequencies of the first ten eigenmodes of the dayan, the composite membrane model (Ref. 5), and the smoothly nonuniform membrane model presented in this work for $k=0.29$ and $\epsilon=0.18$. The figures in parentheses indicate the deviation of frequencies, in cents, from the experimental value.

Mode	Experimental	Analytical	Numerical
ψ_{01}	0.54	0.49 (−168.2)	0.4846 (−187.38)
ψ_{11}	0.95	0.97 (+36.07)	0.9960 (+81.86)
ψ_{11}	1.00	1.00 (0.0)	1.0000 (0.00)
ψ_{21}	1.52	1.46 (−69.72)	1.5617 (+46.86)
ψ_{21}	1.54	1.47 (−80.53)	1.5628 (+25.44)
ψ_{02}	1.75	1.72 (−29.93)	1.3843 (−405.81)
ψ_{31}	2.06	1.94 (−103.9)	2.0903 (+25.28)
ψ_{31}	2.1	1.95 (−128.29)	2.1012 (+0.99)
ψ_{12}	2.32	2.34 (+14.86)	1.7341 (−503.88)
ψ_{12}	2.36	2.35 (−7.35)	1.7752 (−492.93)

TABLE III. A comparison of the variational (Ref. 6) and numerical eigenvalues for eccentric loading (k=0.4, ϵ=0.18).

Mode	Variational	Numerical
ψ_{01}	1	1
ψ_{11}	1.9	1.9199
ψ_{11}	1.96	1.9215
ψ_{02}	3.08	2.9026
ψ_{21}	2.98	2.9207
ψ_{21}	2.98	2.9215
ψ_{12}	4.04	3.7715
ψ_{12}	4.15	3.8323

eigenmodes for eccentric loading. It is interesting to see that modes, which were degenerate in the concentric case, are no longer degenerate. Furthermore, the nodal diameters now deform into nodal lines which are no longer straight. The nodal circles also deform into closed contours. The lifting of the degeneracies is also evident in Fig. 9 where we see that each of the lines fork out into two lines at large values of the eccentricity. Finally, in Figs. 11 and 12 we compare the first four eigenmodes for concentric and eccentric cases. As should be obvious, the lifting of the degeneracies is clearly seen. Certain eigenmodes no longer have circular symmetry.

V. SUMMARY

We have presented a mathematical model, consisting of a membrane of nonuniform density, for the vibrations of the drum head of Indian musical drums. We used a high-resolution numerical method, based on Fourier–Chebyshev collocation, to make an exhaustive study of the variation in the eigenvalues of the model as function of the model parameters. The eigenspectrum of the model agrees very well with the experimentally measured eigenvalues of the tabla.

There are several directions in which this present work needs to be extended to make it more realistic. First, we have completely neglected the role of the enclosed air inside the drums. The principal effect of this is to raise the pitch of those modes which, during vibration, appreciably change the volume of the enclosed air. The fundamental mode of vibration is most affected by this. Preliminary work studying this

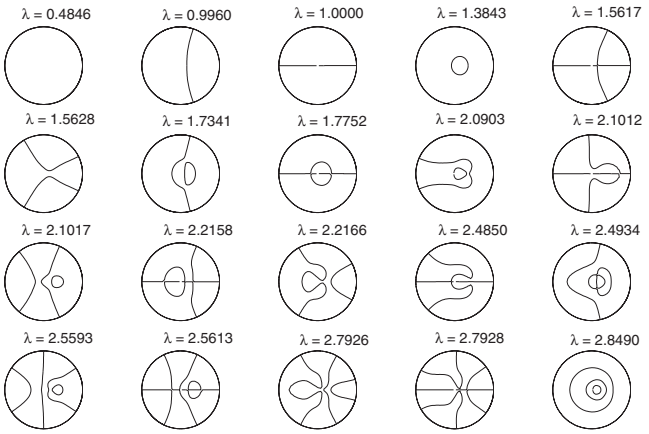


FIG. 10. Nodal contours for the first 20 eigenmodes of the bayan. The plots are with σ=3.125, k=0.29, ξ=0.091, and ϵ=0.18.

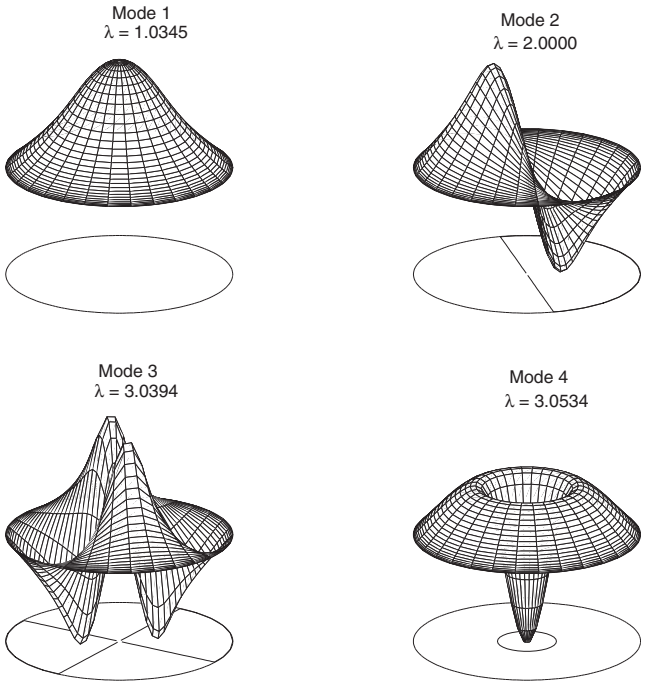


FIG. 11. Eigenmodes for the concentric case.

effect has been done by Bhat,[15] but a more systematic numerical study remains to be done. This is part of ongoing work.

Second, our present study focuses only on the real parts of the eigenvalues of the normal modes. We have completely neglected the role of acoustic damping due to the radiation of sound. The damping effects depend quite strongly on the symmetry of the vibrations. For example, the radiation damping of the fundamental, which is in the far field is an

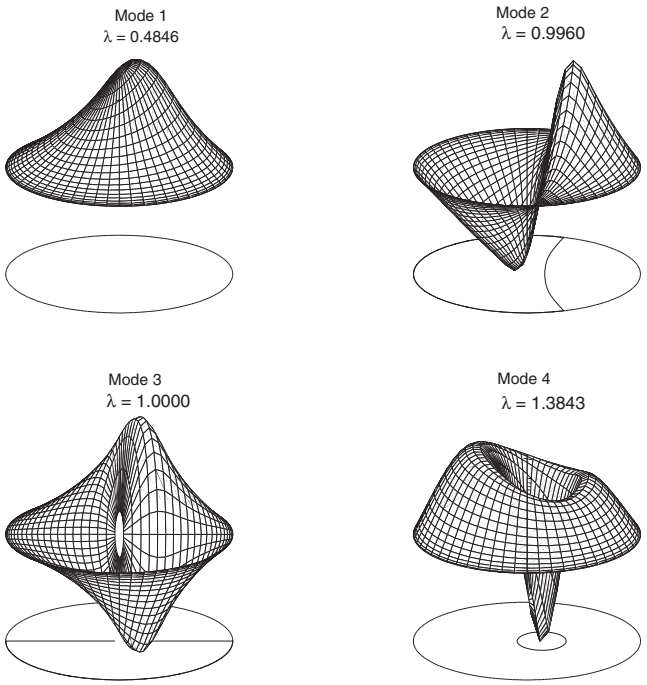


FIG. 12. Eigenmodes for the eccentric case.

acoustic monopole, is quite different from that of the first overtone, which in the far field is an acoustic dipole. We have seen no numerical study of the radiation damping problem for the Indian drums, though extensive work has been done for the uniform circular membrane. This is a problem for further study.

Third, with further refinement our model and the numerical method should find application in numerical sound synthesis. With the increasing power of computer hardware, it is now possible to simulate in real time, physical models, albeit simple ones, of musical elements like strings, membranes, and plates. Numerical sound synthesis will take advantage of growing computational power and we believe that it will be possible to have realistic numerical models of the Indian musical drums using the model and numerical method presented here.

Fourth, we note that our model of the nonuniform membrane is simplified. The sihai, as mentioned earlier, is actually a complex material made of several ingredients like soot, iron filings, flour, and other polymerizing substances. The making and application of the sihai is an art, and we believe that our smoothly nonuniform membrane model captures the subtle physics of the sihai in a gross manner. There remains considerable scope for improvement of mathematical models of the drum head.

Finally, we end with an interesting speculation. Kac[16] posed the isospectral problem for a two-dimensional Laplacian by wittily asking "Can one hear the shape of a drum ?" Very recently it has been shown that one cannot hear the shape of a drum,[17] that is, there exist distinct boundaries in which the Laplacian operator with Dirichlet boundary conditions has the same spectrum. One may now ask, Is the same true for a nonuniform membrane? In other words, "Can one hear the shape of an Indian drum?"

ACKNOWLEDGMENTS

R.A. wishes to thank Dhananjay Modak for useful discussions regarding the construction of the tabla. We thank deodesign.wordpress.com for permission to reproduce Figs. 1 and 2. This work was funded in part by EPSRC Grant No. GR/S10377 at the University of Edinburgh.

[1] T. D. Rossing, *Science of Percussion Instruments*, Popular Science Series (World Scientific, Singapore, 2000).
[2] C. V. Raman and S. Kumar, "Musical drum with harmonic overtones," Nature (London) **104**, 500–500 (1920).
[3] C. V. Raman, "Indian musical drums," Proc. Indian Acad. Sci., Sect. A **1A**, 179–188 (1934).
[4] B. S. Ramakrishna and M. M. Sondhi, "Vibrations of Indian musical drums regarded as composite membranes," J. Acoust. Soc. Am. **26**, 523–529 (1954).
[5] B. S. Ramakrishna, "Modes of vibration of the Indian drum dugga or left-hand thabala," J. Acoust. Soc. Am. **29**, 234–238 (1957).
[6] T. Sarojini and A. Rahman, "Variational method for the vibrations of the Indian drums," J. Acoust. Soc. Am. **30**, 191–196 (1958).
[7] T. D. Rossing and W. A. Sykes, "Acoustics of Indian drums," Percussive Notes **19**(3), 58–67 (1982).
[8] R. N. Ghosh, "Note on musical drums," Phys. Rev. **20**, 526–527 (1922).
[9] K. N. Rao, "Note on musical drums with harmonic overtones," Proc. Indian Acad. Sci., Sect. A **7A**, 75–84 (1938).
[10] S. Malu and A. Siddharthan, "Acoustics of the Indian drum," e-print arXiv:math-ph/0001030v1.
[11] B. Fornberg, *A Guide to Pseudospectral Methods* (Cambridge University Press, Cambridge, 1996).
[12] L. N. Trefethen, *Spectral Methods in MATLAB* (SIAM, Philadelphia, PA, 2000).
[13] S. Gaudet, C. Gauthier, and S. Léger, "The evolution of harmonic Indian musical drums: A mathematical perspective," J. Sound Vib. **291**, 388–394 (2006).
[14] B. M. Banerjee and D. Nag, "The acoustical character of sounds from Indian twin drums," Acustica **75**, 206–208 (1991).
[15] R. B. Bhat, "Acoustics of a cavity-backed membrane: The Indian musical drum," J. Acoust. Soc. Am. **90**, 1469–1474 (1991).
[16] M. Kac, "Can one hear the shape of a drum?," Am. Math. Monthly, **73**, 1–23 (1966).
[17] C. Gordon, D. Webb, and S. Wolpert, "Isospectral plane domains and surfaces via Riemannian orbifolds," Invent. Math. **110**, 1–22 (1991).

Acoustic metafluids

Andrew N. Norris[a)]
Mechanical and Aerospace Engineering, Rutgers University, Piscataway, New Jersey 08854

(Received 6 September 2008; revised 19 November 2008; accepted 21 November 2008)

Acoustic metafluids are defined as the class of fluids that allow one domain of fluid to acoustically mimic another, as exemplified by acoustic cloaks. It is shown that the most general class of acoustic metafluids are materials with anisotropic inertia and the elastic properties of what are known as pentamode materials. The derivation uses the notion of finite deformation to define the transformation of one region to another. The main result is found by considering energy density in the original and transformed regions. Properties of acoustic metafluids are discussed, and general conditions are found which ensure that the mapped fluid has isotropic inertia, which potentially opens up the possibility of achieving broadband cloaking.
 [DOI: 10.1121/1.3050288]

PACS number(s): 43.40.Sk, 43.20.Fn, 43.20.Tb, 43.35.Bf [AJMD] Pages: 839–849

I. INTRODUCTION

Ideal acoustic stealth is provided by the acoustic cloak, a shell of material that surrounds the object to be rendered acoustically "invisible." Stealth can also be achieved by "hiding under the carpet,"[1] as shown in Fig. 1. A simpler situation but one that displays the essence of the acoustic stealth problem is depicted in Fig. 2. The common issue is how to make one region of fluid acoustically mimic another region of fluid. The fluids are different as are the domains they occupy; in fact, the mimicking region is typically smaller in size, and it can be viewed as a compacted version of the original.

The subject of this paper is not acoustic cloaks, or carpets, or ways to hide things, but rather the type of material necessary to achieve stealth. We define these materials as *acoustic metafluids*, which as we will see can be considered fluids with microstructure and properties outside those found in nature. The objective is to derive the general class of acoustic metafluids, and in the process show that there is a closed set which can be mapped from one to another. Acoustic metafluids are defined as the class of fluids that (a) acoustically mimic another region as in the examples of Figs. 1 and 2, and (b) can themselves be mimicked by another acoustic metafluid in the same sense. The requirement (b) is important, implying that there is a closed set of acoustic metafluids. The set includes as a special case the "normal" acoustic fluid of uniform density and bulk modulus. Acoustic metafluids can therefore be used to create stealth devices in a normal fluid. But, in addition, acoustic metafluids can provide stealth in any type of acoustic metafluid. The reciprocal nature of these fluids make them a natural generalization of normal acoustic fluids.

The acoustic cloaks that have been investigated to date fall into two categories in terms of the type of acoustic metafluid proposed as cloaking material. Most studies, e.g., Refs. 2–7, consider the cloak to comprise fluid with the normal stress-strain relation but anisotropic inertia, what we call *inertial cloaking*. Particular realization of inertial cloaks are in principle feasible using layers of isotropic normal fluid,[8–11] the layers are introduced in order to achieve a homogenized medium that approximates a fluid with anisotropic inertia. An alternative and more general approach[12,13] is to consider anisotropic inertia combined with anisotropic elasticity. The latter is introduced by generalizing the stress strain relation to include what are known as pentamode[14] elastic materials.[12,15,13] Clearly, the question of how to design and fabricate acoustic metafluids remains open. The focus of this paper is to first characterize the acoustic metafluids as a general type of material. In fact, as will be shown, this class of materials contains broad degrees of freedom, which can significantly aid in future design studies.

The paper is organized as follows. The concept of acoustic metafluids is introduced in Sec. II through two "acoustic mirage" examples. The methods used to find the acoustic metafluid in these examples are simple but not easily generalized. An alternative and far more powerful approach is discussed in Sec. III: the transformation method. This is based on using the change in variables between the coordinates of the two regions combined with differential relations to identify the metafluid properties of the transformed domain. Leonhardt and Philbin[16] provided an instructive review of the transformation method in the context of optics. The transformation method does not, however, define the range of material properties capable of being transformed. This is the central objective of the paper and it is resolved in Sec. IV by considering the energy density in the original and transformed domains. Physical properties of acoustic metafluids are discussed in Sec. V, including the unusual property that the top surface is not horizontal when at rest under gravity. The subset of acoustic metafluids that have isotropic inertia is considered in Sec. VI, and a concluding summary is presented in Sec. VII.

II. ACOUSTIC MIRAGES AND SIMPLE METAFLUIDS

The defining property of an *acoustic metafluid* is its ability to mimic another acoustic fluid that occupies a different domain. The simplest type of acoustic illusion is what may

a)Electronic mail: norris@rutgers.edu

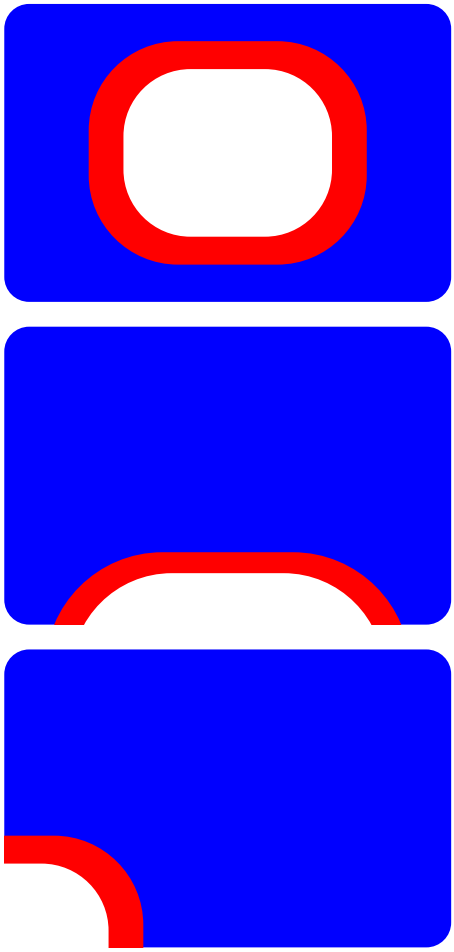


FIG. 1. (Color online) Three ways to acoustically hide something: envelope it with a cloak (top), hide it under a carpet (middle), or hide the object in the corner. In each case the acoustic metafluid in the cloaking layer emulates the acoustic properties of uniform fluid occupying the layer plus the hidden region.

be called an *acoustic mirage* where an observer hears, for example, a reflection from a distant wall, but in reality the echo originates from a closer boundary. Two examples of acoustic mirages are discussed next.

A. 1D mirage

Consider perhaps the simplest configuration imaginable, a one- dimensional (1D) semi-infinite medium. The upper picture in Fig. 2 shows the left end of an acoustic half-space $x \geqslant 0$ with uniform density ρ_0 and bulk modulus κ_0. The wave speed is $c_0=\sqrt{\kappa_0/\rho_0}$. Now replace the region $0 \leqslant x < b$ with a shorter section $0 < b-a \leqslant x < b$ filled with an acoustic metafluid. The acoustic mirage effect requires that

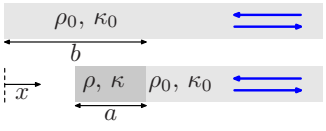


FIG. 2. (Color online) The top shows waves in a semi-infinite medium $x \geqslant 0$. The wave incident from the right reflects from a perfectly reflecting boundary at $x=0$. The lower figure shows the same medium in $x > b$ with the region $0 \leqslant x < b$ replaced by a shorter region of acoustic metafluid. Its properties are such that it produces a perfect acoustic illusion or "mirage."

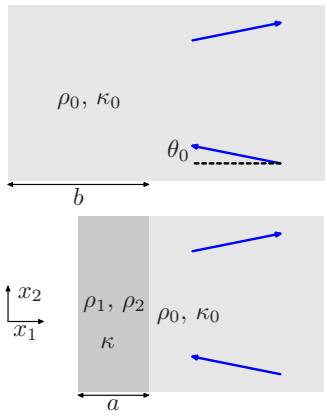


FIG. 3. (Color online) The same as Fig. 2 but now for oblique incidence. The same strategy used for the 1D case is no longer adequate; however, a solution may be found if the metafluid in the layer of thickness a is allowed to have anisotropic inertia defined by the inertias ρ_1 and ρ_2 in the x_1 and x_2 directions.

an observer in $x > b$ hears a response as if the half-space is as shown in the top of Fig. 2. This occurs if the metafluid region is such that (i) no reflection occurs at the interface $x=b$, i.e., the acoustic impedance in the modified region is the same as before; and (ii) the round trip travel time of a wave incident from the right is unchanged. The impedance condition and the travel time requirement ensure that the amplitude and phase of any signal is exactly the same as in the original half-space, and hence the mirage is accomplished.

Let the acoustic metafluid have material properties ρ and κ, with speed $c=\sqrt{\kappa/\rho}$. Conditions (i) and (ii) are satisfied if

$$\rho c = \rho_0 c_0, \quad \frac{a}{c} = \frac{b}{c_0}, \tag{1}$$

respectively, implying the density and bulk modulus in the shorter region are

$$\rho = \frac{b}{a}\rho_0, \tag{2a}$$

$$\kappa = \frac{a}{b}\kappa_0. \tag{2b}$$

In this example the acoustic metafluid is another acoustic fluid, although with greater density and lesser bulk modulus. Note that the total mass of the metafluid region is unchanged from the original: $\rho_0 b = \rho a$.

B. 2D mirage

Consider the same problem in two dimensional (2D) under oblique wave incidence, Fig. 3. A wave incident at angle θ_0 from the normal has travel time $b/(c_0 \cos\theta_0)$ in the original layer. If the shortened region has wave speed c, then the modified travel time is $a/(c\cos\theta)$, where θ is defined by the Snell–Descartes law, $(1/c)\sin\theta=(1/c_0)\sin\theta_0$. At the same time, the reflectivity of the modified layer is $R=(Z-Z_0)/(Z+Z_0)$, where $Z=\rho c/\cos\theta$. The impedance and travel time conditions are

$$\frac{\rho c}{\cos\theta} = \frac{\rho_0 c_0}{\cos\theta_0}, \quad \frac{a}{c\cos\theta} = \frac{b}{c\cos\theta_0}. \tag{3}$$

Solving for the modified parameters implies κ is again given by Eq. (2b) but the density is now

$$\rho = \frac{b}{a}\frac{\rho_0}{2\cos^2\theta_0}\left[1 + \sqrt{1 - \frac{a^2}{b^2}\sin^2 2\theta_0}\right]. \tag{4}$$

The mirage works only for a single direction of incidence, θ_0, and is therefore unsatisfactory. The underlying problem here is that three conditions need to be met: Snell–Descartes' law, matched impedances, and equal travel times, with only two parameters, ρ and κ. Some additional degree of freedom is required.

1. Anisotropic inertia

One method to resolve this problem is to introduce the notion of anisotropic mass density, see Fig. 3. The density of the metafluid region is no longer a scalar, but becomes a tensor: $\rho \to \boldsymbol{\rho}$. The equation of motion and the constitutive relation in the metafluid are

$$\boldsymbol{\rho}\dot{\mathbf{v}} = -\nabla p, \quad \dot{p} = -\kappa\nabla\cdot\mathbf{v}, \tag{5}$$

where $\mathbf{v}$ is the particle velocity and p is the acoustic pressure. Assuming 2D dependence with constant anisotropic density of the form

$$\boldsymbol{\rho} = \begin{bmatrix} \rho_1 & 0 \\ 0 & \rho_2 \end{bmatrix} \tag{6}$$

and eliminating $\mathbf{v}$ imply that the pressure satisfies a scalar wave equation

$$\ddot{p} - c_1^2 p_{,11} - c_2^2 p_{,22} = 0, \tag{7}$$

where $c_j = \sqrt{\kappa/\rho_j}$, $j = 1, 2$. Equations (5) and (7) are discussed in greater detail and generality in Sec. V, but for the moment we cite two results necessary for finding the metafluid in Fig. 3: the phase speed in direction $\mathbf{n} = n_1\mathbf{e}_1 + n_2\mathbf{e}_2$ is v, and the associated wave or group velocity vector is $\mathbf{w}$, where

$$v^2 = c_1^2 n_1^2 + c_2^2 n_2^2, \tag{8a}$$

$$\mathbf{w} = v^{-1}(c_1^2 n_1 \mathbf{e}_1 + c_2^2 n_2 \mathbf{e}_2). \tag{8b}$$

2. Solution of the 2D mirage problem

The travel time is $a/\mathbf{w}\cdot\mathbf{e}_1$, and the impedance is now $Z = \rho_1 v/\cos\theta$, where, referring to Fig. 3, $n_1 = \cos\theta$ and $n_2 = \sin\theta$. Hence, the conditions for zero reflectivity and equal travel times are

$$\frac{\rho_1 v}{\cos\theta} = \frac{\rho_0 c_0}{\cos\theta_0}, \tag{9a}$$

$$\frac{av}{c_1^2\cos\theta} = \frac{b}{c_0\cos\theta_0}. \tag{9b}$$

Dividing the latter two relations implies κ is given by Eq. (2b). Snell–Descartes' law, $v^{-1}\sin\theta = c_0^{-1}\sin\theta_0$, combined with Eq. (9a), yields that

$$\frac{c_0^2}{v^2} = \frac{\rho_1^2}{\rho_0^2} + \left(1 - \frac{\rho_1^2}{\rho_0^2}\right)\sin^2\theta_0, \tag{10}$$

while Snell–Descartes' law together with Eq. (8a) implies

$$\frac{c_0^2}{v^2} = \frac{c_0^2\rho_1}{\kappa} + \left(1 - \frac{\rho_1}{\rho_2}\right)\sin^2\theta_0. \tag{11}$$

Comparison of Eqs. (10) and (11) implies two identities for ρ_1 and ρ_2. In summary, the three parameters of the modified region are

$$\kappa = \frac{a}{b}\kappa_0, \quad \rho_1 = \frac{b}{a}\rho_0, \quad \rho_2 = \frac{a}{b}\rho_0. \tag{12}$$

These give the desired result: no reflection and the same travel time for any angle of incidence. The metafluid layer faithfully mimics the wave properties of the original layer as observed from exterior vantage points.

3. Anisotropic stiffness

An alternative solution to the quandary raised by Eq. (4) is to keep the density isotropic but to relax the standard isotropic constitutive relation between stress $\boldsymbol{\sigma}$ and strain $\boldsymbol{\varepsilon} = [\nabla\mathbf{u} + (\nabla\mathbf{u})^t]/2$ to allow for material anisotropy. Thus, the standard relation $\boldsymbol{\sigma} = -p\mathbf{I}$ with $p = \kappa\,\mathrm{tr}\,\boldsymbol{\varepsilon}$ is replaced by the stress-strain relation for *pentamode materials*[10]

$$\boldsymbol{\sigma} = \kappa(\mathbf{Q}:\boldsymbol{\varepsilon})\mathbf{Q}, \tag{13a}$$

$$\mathrm{div}\,\mathbf{Q} = 0. \tag{13b}$$

The physical meaning of the symmetric second order tensor $\mathbf{Q}$ is discussed later within the context of a more general constitutive theory. The requirement $\mathrm{div}\,\mathbf{Q} = 0$ was first noted by Norris[12] and is discussed in Sec. V. Standard acoustics corresponds to $\mathbf{Q} = \mathbf{I}$.

Rewriting Eq. (13a) as $\boldsymbol{\sigma} = -p\mathbf{Q}$ and using the divergence free property of $\mathbf{Q}$, the equation of motion and the constitutive relation in the metafluid are now

$$\rho\dot{\mathbf{v}} = -\mathbf{Q}\nabla p, \quad \dot{p} = -\kappa\mathbf{Q}:\nabla\mathbf{v}. \tag{14}$$

Eliminating $\mathbf{v}$ yields the scalar wave equation

$$\ddot{p} - \kappa\mathbf{Q}:\nabla(\rho^{-1}\mathbf{Q}\nabla p) = 0. \tag{15}$$

General properties of this equation have been discussed by Norris[12] and will be examined later in Sec. V. For the purpose of the problem in Fig. 3, $\mathbf{Q}$ is assumed constant of the form

$$\mathbf{Q} = \begin{bmatrix} Q_1 & 0 \\ 0 & Q_2 \end{bmatrix}, \tag{16}$$

then it follows that the phase speed and group velocity in direction $\mathbf{n}$ are

$$v^2 = C_1^2 n_1^2 + C_2^2 n_2^2, \quad \mathbf{w} = v^{-1}(C_1^2 n_1 \mathbf{e}_1 + C_2^2 n_2 \mathbf{e}_2), \tag{17}$$

where $C_j = Q_j\sqrt{\kappa/\rho}$, $j = 1, 2$. Proceeding as before, using Snell–Descartes' law and the conditions of equal travel time and matched impedance yields

$$Q_1^2=\frac{a}{b}\frac{\kappa_0}{\kappa},\quad Q_2^2=\frac{b}{a}\frac{\rho}{\rho_0},\quad Q_1Q_2=\frac{b}{a}.$$

Since the important physical quantity is the product of κ with $\mathbf{Q}\otimes\mathbf{Q}$, any one of the three parameters κ, Q_1, and Q_2, may be independently selected. A natural choice is to impose $\operatorname{div}\mathbf{Q}=0$ at the interface, which means that the "traction" vector $\mathbf{Qn}$ is continuous, where $\mathbf{n}$ is the interface normal. In this case $\mathbf{n}=\mathbf{e}_1$ so that $Q_1=1$ and

$$\rho=\frac{b}{a}\rho_0,\quad \kappa=\frac{a}{b}\kappa_0,\quad \mathbf{Q}=\begin{bmatrix}1 & 0\\ 0 & \frac{b}{a}\end{bmatrix}. \tag{18}$$

Comparing the alternative metafluids defined by Eqs. (12) and (18) note that in each case the density and the stiffness associated with the normal $\mathbf{e}_1$ direction both equal their 1D values given by Eq. (2). The first metafluid of Eq. (12) has a smaller inertia in the transverse direction $\mathbf{e}_2(\rho_2<\rho_1)$. The second metafluid defined by Eq. (18) has increased stiffness in the transverse direction $(Q_2>Q_1)$. The net effect in each case is an increased phase speed in the transverse direction as compared with the normal direction $(c_2>c_1, C_2>C_1)$.

The 1D and 2D mirage examples illustrate the general idea of acoustic metafluids as fluids that replicate the wave properties of a transformed region. However, the methods used to find the metafluids are not easily generalized to arbitrary regions. How does one find the metafluid that can, for instance, mimic a full spherical region by a smaller shell? This is the cloaking problem. The key to the generalized procedure are the related notions of transformation and finite deformation, which are introduced next.

III. THE TRANSFORMATION METHOD

A. Preliminaries

Let Ω and ω denote the original and the deformed domains (the regions $0\leqslant X<b$ and $b-a\leqslant x<b$ in the examples of Sec. II. The coordinates in each configuration are $\mathbf{X}$ and $\mathbf{x}$, respectively; the divergence operators are Div and div, and the gradient (nabla) operators are $\bar{\nabla}$ and ∇. The upper and lower case indices indicate components, X_I, x_i and the component form of $\operatorname{div}\mathbf{A}$ is $\partial A_i/\partial x_i=A_{i,i}$ or $\partial A_{iJ}/\partial x_i=A_{iJ,i}$ when $\mathbf{A}$ is a vector and a second order tensorlike quantity, respectively, and repeated indices imply summation (case sensitive). Similarly $\operatorname{Div}\mathbf{B}=B_{\mathrm{lj},I}$.

The finite deformation or transformation is defined by the mapping $\Omega\rightarrow\omega$ according to $\mathbf{x}=\boldsymbol{\chi}(\mathbf{X})$. In the terminology of finite elasticity $\mathbf{X}$ describes a particle position in the Lagrangian or undeformed configuration, and $\mathbf{x}$ is particle location in the Eulerian or deformed physical state. The transformation $\boldsymbol{\chi}$ is assumed to be one-to-one and invertible.[17] The deformation gradient is defined $\mathbf{F}=\bar{\nabla}\mathbf{x}$ with inverse $\mathbf{F}^{-1}=\nabla\mathbf{X}$, or in component form $F_{iI}=\partial x_i/\partial X_I$, $F_{Ii}^{-1}=\partial X_I/\partial x_i$. The Jacobian of the finite deformation is $\Lambda=\det\mathbf{F}=|\mathbf{F}|$, or in terms of volume elements in the two configurations, $\Lambda=dv/dv$. The polar decomposition implies $\mathbf{F}=\mathbf{VR}$, where the rotation $\mathbf{R}$ is proper orthogonal $(\mathbf{RR}^t=\mathbf{R}^t\mathbf{R}=\mathbf{I}, \det\mathbf{R}=1)$ and the left stretch tensor $\mathbf{V}\in\mathrm{Sym}^+$ is the positive definite solution of $\mathbf{V}^2=\mathbf{FF}^t$. Note for later use the kinematic identities[18]

$$\operatorname{div}(\Lambda^{-1}\mathbf{F})=0, \tag{19a}$$

$$\operatorname{Div}(\Lambda\mathbf{F}^{-1})=0, \tag{19b}$$

and the expression for the Laplacian in $\mathbf{X}$ in terms of derivatives in $\mathbf{x}$, i.e., the chain rule[8]

$$\bar{\nabla}^2 f=\Lambda\operatorname{div}(\Lambda^{-1}\mathbf{V}^2\nabla f). \tag{20}$$

B. The transformation method

The undeformed domain Ω is of arbitrary shape and comprises a homogeneous acoustic fluid with density ρ_0 and bulk modulus κ_0. The goal is to mimic the scalar wave equation in Ω,

$$\ddot{p}-(\kappa_0/\rho_0)\bar{\nabla}^2 p=0,\quad \mathbf{X}\in\Omega, \tag{21}$$

by the wave equation of a metafluid occupying the deformed region ω. The basic result[6,8,19] is that Eq. (21) is exactly replicated in ω by the equation

$$\ddot{p}-\kappa\operatorname{div}(\boldsymbol{\rho}^{-1}\nabla p)=0,\quad \mathbf{x}\in\omega, \tag{22}$$

where the bulk modulus and inertia tensor are

$$\kappa=\kappa_0\Lambda,\quad \boldsymbol{\rho}=\rho_0\Lambda\mathbf{V}^{-2}. \tag{23}$$

The equivalence of Eq. (21) with Eqs. (22) and (23) is evident from the differential equality (20). The idea is to use the change in variables from $\mathbf{X}$ to $\mathbf{x}$ to identify the metafluid properties. Equation (23) describes a metafluid with anisotropic inertia and isotropic elasticity. It can be used for modeling acoustic cloaks but has the unavoidable feature that the total effective mass of the cloak is infinite. This problem, discussed by Norris,[12] arises from the singular nature of the finite deformation in a cloak which makes $\Lambda\mathbf{V}^{-2}$ nonintegrable. This type of fluid, which could be called an inertial fluid, appears to be the main candidate considered for acoustic cloaking to date. The major exception is Milton *et al.*[13] who considered fluids with properties of pentamode materials, although as we will discuss in Sec. IV, their findings are of limited use for acoustic cloaking.

1. Pentamode materials

Norris[12] showed that Eq. (20) is a special case of a more general identity:

$$\bar{\nabla}^2 p=\Lambda\mathbf{Q}:\boldsymbol{\nabla}(\Lambda^{-1}\mathbf{Q}^{-1}\mathbf{V}^2\boldsymbol{\nabla}p), \tag{24}$$

where $\mathbf{Q}$ is *any* symmetric, invertible, and divergence-free $(\operatorname{div}\mathbf{Q}=0)$ second order tensor. The increased degrees of freedom afforded by the arbitrary nature of $\mathbf{Q}$ means that Eq. (21) is equivalent to the generalized scalar wave equation in ω,

$$\ddot{p} - \kappa \mathbf{Q}:\nabla(\boldsymbol{\rho}^{-1}\mathbf{Q}\,\nabla\, p) = 0, \quad \mathbf{x} \in \omega, \tag{25}$$

where the modulus κ and the inertia follow from a comparison of Eqs. (21), (24), and (25) as

$$\kappa = \kappa_0 \Lambda, \quad \boldsymbol{\rho} = \rho_0 \Lambda \mathbf{Q}\mathbf{V}^{-2}\mathbf{Q}. \tag{26}$$

As will become apparent later, these metafluid parameters describe a pentamode material with anisotropic inertia. For the moment we return to the acoustic mirages in light of the general transformation method.

C. Mirages revisited

The 2D mirage problem corresponds to the following finite deformation $x_1 = b - a + ab^{-1}X_1$, $x_2 = X_2$ for $0 \leqslant X_1 < b$, $-\infty < X_2 < \infty$. The deformation gradient is

$$\mathbf{F} = \begin{bmatrix} \dfrac{a}{b} & 0 \\ 0 & 1 \end{bmatrix}, \tag{27}$$

implying $\mathbf{R} = \mathbf{I}$, $\mathbf{V} = \mathbf{F}$ and $\Lambda = a/b$. Equation (23) therefore implies

$$\kappa = \frac{a}{b}\kappa_0, \quad \boldsymbol{\rho} = \begin{bmatrix} \dfrac{b}{a}\rho_0 & 0 \\ 0 & \dfrac{a}{b}\rho_0 \end{bmatrix}. \tag{28}$$

These agree with the parameters found in Sec. II, Eq. (12).

Using the more general formulation of Eqs. (25) and (26) with the arbitrary tensor chosen as $\mathbf{Q} = \Lambda^{-1}\mathbf{V}$ yields the metafluid described by Eq. (18). It is interesting to note that although ρ of Eq. (26) is, in general, anisotropic, it can be made isotropic in this instance by any $\mathbf{Q}$ proportional to $\mathbf{V}$. Keeping in mind the requirement seen above that $Q_1 = 0$, we consider as a second example of Eq. (25) the case $\mathbf{Q} = \Lambda^{-1}\mathbf{V}$. The mirage can then be achieved with material properties

$$\rho = \frac{b}{a}\rho_0, \quad \kappa = \frac{a}{b}\kappa_0, \quad \mathbf{Q} = \begin{bmatrix} 1 & 0 \\ 0 & \dfrac{b}{a} \end{bmatrix}. \tag{29}$$

This again corresponds to a pentamode material with isotropic inertia, equal to that of Eq. (18). These two examples illustrate the power associated with the arbitrary nature of the divergence-free tensor $\mathbf{Q}$. There appears to be a multiplicative degree of freedom associated with $\mathbf{Q}$ that is absent using anisotropic inertia. As will be evident later, this degree of freedom is related to a *gauge transformation*.

IV. THE MOST GENERAL TYPE OF ACOUSTIC METAFLUID

A. Summary of main result

In order to make it easier for the reader to assimilate, the paper's main result is first presented in the form of a theorem. In the following, $\mathbf{Q}_0$ and $\mathbf{Q}$ are arbitrary symmetric, invertible, and divergence-free second order tensors.

Theorem 1. *The kinetic and strain energy densities in* Ω *of the form*

$$T_0 = \dot{\mathbf{U}}^t \boldsymbol{\rho}_0 \dot{\mathbf{U}}, \quad W_0 = \kappa_0 (\mathbf{Q}_0 : \bar{\boldsymbol{\nabla}}\mathbf{U})^2, \tag{30}$$

respectively, are equivalent to the current energy densities in ω:

$$T = \dot{\mathbf{u}}^t \boldsymbol{\rho} \dot{\mathbf{u}}, \quad W = \kappa (\mathbf{Q}:\nabla \mathbf{u})^2, \tag{31}$$

where

$$\kappa = \Lambda \kappa_0, \tag{32a}$$

$$\boldsymbol{\rho} = \Lambda \mathbf{Q}\mathbf{F}^{-t}\mathbf{Q}_0^{-1}\boldsymbol{\rho}_0\mathbf{Q}_0^{-1}\mathbf{F}^{-1}\mathbf{Q}. \tag{32b}$$

Discussion of the implications are given following the proof.

B. Gauge transformation

The energy functions per unit volume in the undeformed configuration, T_0 and W_0, depend on the infinitesimal displacement $\mathbf{U}$ in that configuration. The kinetic energy is defined by the density $\boldsymbol{\rho}_0$, while the strain energy is $W_0 = C_{0IJKL}U_{J,I}U_{L,K}$, where C_{0IJKL} are elements of the stiffness. The density and stiffness possess the symmetries $\boldsymbol{\rho}_{0IJ} = \rho_{0JI}$, $C_{0IJKL} = C_{0KLIJ}$, $C_{0IJKL} = C_{0JIKL}$. The total energy is $E_0 = T_0 + W_0$ and the total energy $E = T + W$ per unit deformed volume is, using $E_0 \mathrm{d}V = E\mathrm{d}v$, simply $E = \Lambda^{-1}E_0$.

Our objective is to find a general class of material parameters $\{\boldsymbol{\rho}_0, \mathbf{C}_0\}$ that maintain the structure of the energy under a general transformation χ. Structure here means that the energy remains quadratic in velocity and strain. In order to achieve the most general form for the transformed energy, introduce a gauge transformation for the displacement. Let

$$\mathbf{U} = \mathbf{G}^t\mathbf{u}, \tag{33}$$

or $U_I = u_i G_{iI}$ in components, where $\mathbf{G}$ is independent of time but can be spatially varying. Thus, using the chain rule $U_{J,I} = U_{J,i}F_{iI}$ yields $E = T + W$, where the kinetic and strain energy densities are

$$T = \dot{\mathbf{u}}^t \boldsymbol{\rho} \dot{\mathbf{u}}, \tag{34a}$$

$$W = \Lambda^{-1} C_{IJKL} F_{iI} F_{kK} (G_{jJ}u_j)_{,i} (G_{lL}u_l)_{,k}, \tag{34b}$$

and the transformed inertia is

$$\boldsymbol{\rho} = \Lambda^{-1}\mathbf{G}\boldsymbol{\rho}_0\mathbf{G}^t. \tag{35}$$

The kinetic energy has the required structure, quadratic in the velocity; the strain energy, however, is not in the desired form. The objective is to obtain a strain energy of standard form

$$W = C_{ijkl}u_{j,i}u_{l,k}, \tag{36}$$

where $\mathbf{C}$ has the usual symmetries: $C_{ijkl} = C_{klij}$, $C_{ijkl} = C_{jikl}$.

The question of how to convert W of Eq. (34b) into the form of Eq. (36) for generally anisotropic elasticity stiffness $\mathbf{C}_0$ will be discussed in a separate paper. The objective here is to find the largest class of metafluids that includes all those previously found.

C. Pentamode to pentamode

In order to proceed assume that the initial stiffness tensor is of pentamode form[15]

$$\mathbf{C}_0 = \kappa_0 \mathbf{Q}_0 \otimes \mathbf{Q}_0, \tag{37a}$$

$$\mathrm{Div}\, \mathbf{Q}_0 = 0, \tag{37b}$$

that is, $C_{0IJKL} = \kappa_0 Q_{0IJ} Q_{0KL}$, where $\mathbf{Q}_0^t = \mathbf{Q}_0$ is a positive definite symmetric second order tensor. The tensor $\mathbf{Q}_0$ is necessarily divergence-free.[12]

The strain energy density in the physical space after the general deformation and gauge transformation is now

$$W = \kappa_0 \Lambda^{-1} [Q_{0IJ}(u_j G_{jJ})_{,I}]^2. \tag{38}$$

Consider

$$Q_{0IJ}(u_j G_{jJ})_{,I} = Q_{0IJ} G_{jJ} u_{j,I} + Q_{0IJ} G_{jJ,I} u_j. \tag{39}$$

In order to achieve the quadratic structure of Eq. (36) the final term in Eq. (39) must vanish. Since $\mathbf{u}$ is considered arbitrary this in turn implies that $Q_{0IJ} G_{jJ,I}$ must vanish for all j. With no loss in generality let

$$\mathbf{G}^t = \mathbf{Q}_0^{-1} \mathbf{P}, \tag{40}$$

or $G_{jJ} = Q_{0JM}^{-1} P_{Mj}$ in components. Then using the identity for the derivative of a second order tensor, $(\mathbf{A}^{-1})_{,\alpha} = -\mathbf{A}^{-1} \mathbf{A}_{,\alpha} \mathbf{A}^{-1}$, where α can be any parameter, gives

$$\begin{aligned} 0 = Q_{0IJ} G_{jJ,I} &= -Q_{0IJ} Q_{0JK}^{-1} Q_{0KL,I} Q_{0LM}^{-1} P_{MJ} + P_{Ij,I} \\ &= -Q_{0IL,I} Q_{0LM}^{-1} P_{Mj} + P_{Ij,I} = P_{Ij,I}, \end{aligned} \tag{41}$$

where the important property (37b) has been used. Hence, $\mathrm{Div}\, \mathbf{P} = 0$. Then using Eq. (39) yields

$$Q_{0IJ}(u_j G_{jJ})_{,I} = P_{Ij} u_{j,I} = P_{Ij} F_{iI} u_{j,i} = \Lambda Q_{ij} u_{j,i}, \tag{42}$$

where the tensor $\mathbf{Q}$ is defined by

$$\mathbf{P} = \Lambda \mathbf{F}^{-1} \mathbf{Q} \Leftrightarrow \mathbf{Q} = \Lambda^{-1} \mathbf{F} \mathbf{P}. \tag{43}$$

The condition $\mathrm{Div}\, \mathbf{P} = 0$ becomes, using Eq. (19b),

$$P_{Ij,I} = (\Lambda F_{Ii}^{-1} Q_{ij})_{,I} = \Lambda F_{Ii}^{-1} Q_{ij,I} = \Lambda Q_{ij,i}, \tag{44}$$

implying

$$\mathrm{div}\, \mathbf{Q} = 0. \tag{45}$$

It has therefore been demonstrated that if the gauge transformation is defined by

$$\mathbf{G}^t = \Lambda \mathbf{Q}_0^{-1} \mathbf{F}^{-1} \mathbf{Q}, \tag{46}$$

where $\mathbf{Q}$ is divergence free in physical coordinates, then the strain energy (38) is $W = \kappa_0 \Lambda (Q_{ij} u_{j,i})^2$. This is of the desired form, Eq. (36), with $C_{ijkl} = \kappa Q_{ij} Q_{kl}$, hence completing the proof of Theorem 1.

D. Discussion

1. Equivalence of physical quantities

Theorem 1 states that the pentamode material defined by stiffness κ_0 with stresslike tensor $\mathbf{Q}_0$ and anisotropic inertia $\boldsymbol{\rho}_0$ is converted into another pentamode material with anisotropic inertia. The properties of the new metafluid are defined by the original metafluid and the deformation-gauge pair $\{\mathbf{F}, \mathbf{G}\}$, where $\mathbf{F}$ is arbitrary and possibly inhomogeneous, and $\mathbf{G}$ is given by Eq. (46) with $\mathbf{Q}$ symmetric, positive definite, and divergence-free but otherwise completely arbitrary. The special case of a fluid with isotropic stiffness but anisotropic inertia, Eq. (23), is recovered from Theorem 1 when the starting medium is a standard acoustic fluid and $\mathbf{Q}$ is taken to be $\mathbf{I}$.

It is instructive to examine how physical quantities transform: we consider displacement, momentum, and pseudopressure. Eliminating $\mathbf{G}$, it is possible to express the new displacement vector in terms of the original,

$$\mathbf{u} = \Lambda^{-1} \mathbf{Q}^{-1} \mathbf{F} \mathbf{Q}_0 \mathbf{U}. \tag{47}$$

Physically, this means that as the metafluid in ω acoustically replicates that in Ω, particle motion in the latter is converted into the mimicked motion defined by Eq. (47).

Define the momentum vectors in the two configurations,

$$\mathbf{m}_0 = \boldsymbol{\rho}_0 \dot{\mathbf{U}}, \quad \mathbf{m} = \boldsymbol{\rho} \dot{\mathbf{u}}. \tag{48}$$

Equations (35) and (46) imply that they are related by

$$\mathbf{m} = \mathbf{Q} \mathbf{F}^{-t} \mathbf{Q}_0^{-1} \mathbf{m}_0. \tag{49}$$

The transformation of momentum is similar to that for displacement, Eq. (47), but with the inverse tensor, i.e., $\mathbf{u} = \mathbf{G}^{-t} \mathbf{U}$ while $\mathbf{m} = \Lambda^{-1} \mathbf{G} \mathbf{m}_0$.

Stress in the two configurations is defined by Hooke's law in each:

$$\boldsymbol{\sigma}_0 = \mathbf{C}_0 : \bar{\nabla} \mathbf{U}, \quad \boldsymbol{\sigma} = \mathbf{C} : \nabla \mathbf{u}, \tag{50}$$

where $\mathbf{C}_0$ and $\mathbf{C}$ are the fourth order elasticity tensors,

$$\mathbf{C}_0 = \kappa_0 \mathbf{Q}_0 \otimes \mathbf{Q}_0, \quad \mathbf{C} = \kappa \mathbf{Q} \otimes \mathbf{Q}, \tag{51}$$

that is, $C_{ijkl} = \kappa Q_{ij} Q_{kl}$, etc. Using Eqs. (37), (42), and (51) yields

$$\boldsymbol{\sigma}_0 = -p \mathbf{Q}_0, \tag{52a}$$

$$\boldsymbol{\sigma} = -p \mathbf{Q}, \tag{52b}$$

where p is the same in each configuration,

$$p = -\kappa \mathbf{Q} : \nabla \mathbf{u} = -\kappa_0 \mathbf{Q}_0 : \bar{\nabla} \mathbf{U}. \tag{53}$$

The quantity p is similar to pressure, and can be exactly identified as such when $\mathbf{Q}$ is diagonal, but it is not pressure in the usual meaning. For this reason it is called the pseudopressure. It is interesting to compare the equal values of p in Ω and ω with the more complicated relations (47) and (49) for the displacement and momentum.

2. Equations of motion

The equations of motion can be derived as the Euler–Lagrange equation for the Lagrangian $T - W$. A succinct form is as follows, in terms of the the momentum density $\mathbf{m}$ and the stress tensor $\boldsymbol{\sigma}$:

$$\dot{\mathbf{m}} = \nabla \boldsymbol{\sigma}. \tag{54}$$

The constitutive relation may be expressed as an equation for the pseudopressure p,[12]

$$\ddot{p} = -\kappa \mathbf{Q}:\nabla\ddot{\mathbf{u}}, \tag{55}$$

while Eqs. (53), (48), and (54) imply that the acceleration is

$$\ddot{\mathbf{u}} = -\boldsymbol{\rho}^{-1}\mathbf{Q}\,\nabla p. \tag{56}$$

Eliminating p between the last two equations implies that the displacement satisfies

$$\kappa\mathbf{Q}\mathbf{Q}:\nabla\ddot{\mathbf{u}} - \boldsymbol{\rho}\ddot{\mathbf{u}} = 0. \tag{57}$$

This is, as expected, the elastodynamic equation for a pentamode material with anisotropic inertia. Alternatively, eliminating $\ddot{\mathbf{u}}$ between Eqs. (55) and (56) yields a scalar wave equation for the pseudopressure,

$$\ddot{p} - \kappa\mathbf{Q}:\nabla(\boldsymbol{\rho}^{-1}\mathbf{Q}\,\nabla p) = 0. \tag{58}$$

This clearly reduces to the standard acoustic wave equation when $\mathbf{Q}=\mathbf{I}$ and $\boldsymbol{\rho}$ is isotropic.

3. Relation to the findings of Milton et al.

The present findings appear to contradict those of Milton *et al.*[13] who found the negative result that it is not in general possible to find a metafluid that replicates a standard acoustic medium after arbitrary finite deformation. However, their result is based on the assumption that $\mathbf{G}\equiv\mathbf{F}$ [their Eq. (2.2)]. Equation (46) implies that $\mathbf{Q}$ must then be

$$\mathbf{Q} = \Lambda^{-1}\mathbf{F}\mathbf{Q}_0\mathbf{F}^t. \tag{59}$$

Using Eqs. (19a) and (37b) yields $Q_{ij,i}=\Lambda^{-1}Q_{0IJ}F_{jJ,I}$. Hence, this particular $\mathbf{Q}$ can only satisfy the requirement (45) that $\operatorname{div}\mathbf{Q}=0$ if

$$Q_{0IJ}\frac{\partial^2 x_i}{\partial X_I \partial X_J} = 0. \tag{60}$$

Milton *et al.*[13] considered $\mathbf{Q}_0$ isotropic (diagonal), in which case Eq. (60) means that the only permissible finite deformations are harmonic, i.e., those for which $\bar{\nabla}^2\mathbf{x}=0$. In short, the negative findings of Milton *et al.*[13] are a consequence of constraining the gauge to $\mathbf{G}\equiv\mathbf{F}$, which in turn severely restricts the realizability of metafluids except under very limited types of transformation deformation. The main difference in the present analysis is the inclusion of the general gauge transformation which enables us to find metafluids under arbitrary deformation.

V. PROPERTIES OF ACOUSTIC METAFLUIDS

The primary result of the paper, summarized in Theorem 1, states that the class of acoustic metafluids is defined by the most general type of pentamode material with elastic stiffness $\kappa\mathbf{Q}\otimes\mathbf{Q}$ where $\operatorname{div}\mathbf{Q}=0$, and anisotropic inertia $\boldsymbol{\rho}$. We now examine some of the unusual physical properties, dynamic and static, to be expected in acoustic metafluids. Some of the dynamic properties have been discussed by Norris,[12] but apart from Milton and Cherkaev[15] no discussion of static effects has been given.

A. Dynamic properties: plane waves

Consider plane wave solutions for displacement of the form $\mathbf{u}(\mathbf{x},t)=\mathbf{q}e^{ik(\mathbf{n}\cdot\mathbf{x}-vt)}$, for $|\mathbf{n}|=1$ and constants $\mathbf{q}$, k, and v, and uniform metafluid properties. Nontrivial solutions satisfying the equations of motion (57) require

$$(\kappa(\mathbf{Q}\mathbf{n})\otimes(\mathbf{Q}\mathbf{n}) - \boldsymbol{\rho}v^2)\mathbf{q} = 0. \tag{61}$$

The acoustical or Christoffel[20] tensor $\kappa(\mathbf{Q}\mathbf{n})\otimes(\mathbf{Q}\mathbf{n})$ is rank one and it follows that of the three possible solutions for v^2, only one is not zero, the quasilongitudinal solution

$$v^2 = \kappa\mathbf{n}\cdot\mathbf{Q}\boldsymbol{\rho}^{-1}\mathbf{Q}\mathbf{n}, \quad \mathbf{q} = \boldsymbol{\rho}^{-1}\mathbf{Q}\mathbf{n}. \tag{62}$$

The slowness surface is an ellipsoid. The energy flux velocity[20] (or wave velocity or group velocity) is

$$\mathbf{w} = v^{-1}\kappa\mathbf{Q}\boldsymbol{\rho}^{-1}\mathbf{Q}\mathbf{n}. \tag{63}$$

$\mathbf{w}$ is in the direction $\mathbf{Q}\mathbf{q}$, and satisfies $\mathbf{w}\cdot\mathbf{n}=v$, a well known relation for generally anisotropic solids with isotropic density.

B. Static properties

1. Five easy modes

The static properties of acoustic metafluids are just as interesting, if not more so. Hooke's law (52b) is

$$\boldsymbol{\sigma} = \mathbf{C}\boldsymbol{\varepsilon}, \tag{64}$$

where $\boldsymbol{\varepsilon}=\boldsymbol{\varepsilon}^t$ is strain and the stiffness $\mathbf{C}$ is defined by Eq. (50b). The strain energy is $W=\kappa(\mathbf{Q}:\boldsymbol{\varepsilon})^2$. Note that $\mathbf{C}$ is not invertible in the usual sense of fourth order elasticity tensors. If considered as a 6×6 matrix mapping strain to stress then the stiffness is rank one: it has only one nonzero eigenvalue. This means that there are five independent strains, each of which will produce zero stress and zero strain energy, hence the name *pentamode*.[15] The five "easy" strains are easily identified in terms of the principal directions and eigenvalues of $\mathbf{Q}$. Let

$$\mathbf{Q} = \lambda_1\mathbf{q}_1\mathbf{q}_1 + \lambda_2\mathbf{q}_2\mathbf{q}_2 + \lambda_3\mathbf{q}_3\mathbf{q}_3, \tag{65}$$

where $\{\mathbf{q}_1,\mathbf{q}_2,\mathbf{q}_3\}$ is an orthonormal triad. Three of the easy strains are pure shear: $\mathbf{q}_i\mathbf{q}_j+\mathbf{q}_j\mathbf{q}_i$, $i\neq j$ and the other two are $\lambda_3\mathbf{q}_2\mathbf{q}_2-\lambda_2\mathbf{q}_3\mathbf{q}_3$ and $\lambda_1\mathbf{q}_2\mathbf{q}_2-\lambda_2\mathbf{q}_1\mathbf{q}_1$. Any other zero-energy strain is a linear combination of these. Note that there is no relation analogous to Eq. (64) for strain in terms of stress because only the single "component" $\mathbf{Q}:\boldsymbol{\varepsilon}$ is relevant, i.e., energetic.

It is possible to write $\boldsymbol{\sigma}$ in the form (52) where $p=-\kappa\mathbf{Q}:\boldsymbol{\varepsilon}$. Under static load in the absence of body force choose κ such that $p=$const, or equivalently, Eq. (37b). The relevant strain component is then $\mathbf{Q}:\boldsymbol{\varepsilon}=-\kappa^{-1}p$ and the surface tractions supporting the body in equilibrium are $\mathbf{t}=\boldsymbol{\sigma}\cdot\mathbf{n}=-p\mathbf{Q}\cdot\mathbf{n}$. Figure 4 illustrates the tractions required to maintain a block of metafluid in static equilibrium. Note that the traction vectors act obliquely to the surface, implying that shear forces are necessary. Furthermore, the tractions are not of uniform magnitude. These properties are to be compared with a normal acoustic fluid which can be maintained in static equilibrium by constant hydrostatic pressure.

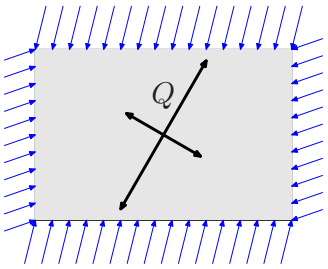


FIG. 4. (Color online) A rectangular block of metafluid is in static equilibrium under the action of surface tractions as shown. The two orthogonal arrows inside the rectangle indicate the principal directions of $\mathbf{Q}$ (30° from horizontal and vertical) and the relative magnitude of its eigenvalues (2:1). The equispaced arrows faithfully represent the surface loads.

2. div Q=0

The *gedanken* experiment of Fig. 4 also implies that $\mathbf{Q}$ has to be divergence-free. Thus, imagine a smoothly varying but inhomogeneous metafluid $\mathbf{C}=\kappa\mathbf{Q}\otimes\mathbf{Q}$ in static equilibrium under the traction $\mathbf{t}=-p\mathbf{Q}\cdot\mathbf{n}$ for constant p. The divergence theorem then implies $\operatorname{div}\mathbf{Q}=0$ everywhere in the interior. This argument is a bit simplistic, but it provides the basis for a more rigorous proof.[12] Thus, stress in the metafluid must be of the form $-f\overline{\mathbf{Q}}$ where $\overline{\mathbf{Q}}$ is a scalar multiple of $\mathbf{Q}$. Local equilibrium requires $\operatorname{div} f\overline{\mathbf{Q}}=0$, or $\nabla \ln f=-\overline{\mathbf{Q}}^{-1}\operatorname{div}\overline{\mathbf{Q}}$. This can be integrated to find f to within a constant. Now define $\mathbf{Q}=f\overline{\mathbf{Q}}$, and note that the tractions must be of the form $\mathbf{t}=-p\mathbf{Q}\cdot\mathbf{n}$ for constant p. The normalized $\mathbf{Q}$ is divergence-free.

3. Nonhorizontal free surface

Consider the same metafluid in equilibrium under a body force, e.g., gravity. Assuming the inertia is isotropic (see the comments about inertia at zero frequency in Sec. VI),

$$\operatorname{div}\boldsymbol{\sigma}+\rho\mathbf{g}=0. \tag{66}$$

Use Eq. (52b) with $\operatorname{div}\mathbf{Q}=0$ and the invertibility of $\mathbf{Q}$ implies

$$\nabla p=\rho\mathbf{Q}^{-1}\mathbf{g}. \tag{67}$$

For constant $\mathbf{Q}^{-1}\rho\mathbf{g}$ this can be integrated to give an explicit form for the pseudopressure,

$$p=(\mathbf{x}-\mathbf{x}_0)\cdot\mathbf{Q}^{-1}\rho\mathbf{g}, \tag{68}$$

where $\mathbf{x}_0$ is any point lying on the surface of zero pressure. Unlike normal fluids, the surface where $p=0$ does not have to be horizontal, see Fig. 5. The pseudopressure increases in the direction of $\mathbf{g}$, as in normal fluids. However, it is possible that p varies in the plane $\mathbf{x}\cdot\mathbf{g}=0$. For instance, the traction along the lower surface in Fig. 5 decreases in magnitude from left to right.

VI. METAFLUIDS WITH ISOTROPIC DENSITY

A. Necessary constraints on the finite deformation

The most practical case of interest is of course where the initial properties are those of a standard acoustic fluid with isotropic density and isotropic stress. The circumstances under which the mapped inertia $\boldsymbol{\rho}$ is also isotropic are now investigated. Acoustic metafluids with isotropic inertia are an important subset since it can be argued that achieving anisotropic inertia could be more difficult than the anisotropic elasticity. Indeed, the very concept of anisotropic inertia is meaningless at zero frequency, unlike anisotropic stiffness.

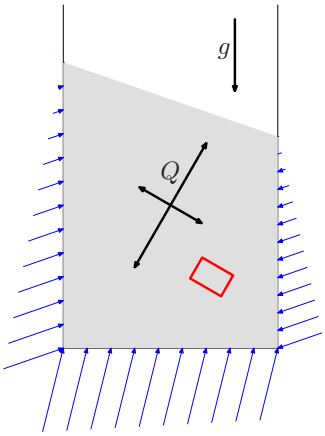


FIG. 5. (Color online) The same metafluid of Fig. 4 with isotropic density is in equilibrium under gravity. The upper surface is traction-free but nonhorizontal, an essential feature of metafluids. For this particular metafluid the top surface makes an angle of 19.11° with the horizontal. Also, the tractions on the horizontal lower boundary are inhomogeneous although parallel. The small rectangle is discussed in Fig. 6.

Assuming $\boldsymbol{\rho}_0=\rho_0\mathbf{I}$ and $\mathbf{Q}_0=\mathbf{I}$ then the current density becomes, using Eq. (32b) and the fact that $\mathbf{Q}$ is symmetric,

$$\boldsymbol{\rho}=\rho_0\Lambda\mathbf{Q}\mathbf{V}^{-2}\mathbf{Q}. \tag{69}$$

If $\boldsymbol{\rho}=\rho\mathbf{I}$ then Eq. (69) implies $\mathbf{Q}$ must be of the form

$$\mathbf{Q}=\rho^{1/2}(\rho_0\Lambda)^{-1/2}\mathbf{V}. \tag{70}$$

Thus, $\mathbf{Q}$ is proportional to the stretch tensor $\mathbf{V}$ and the coefficient of proportionality defines the current density.

It is not in general possible to choose $\mathbf{Q}$ in the form

$$\mathbf{Q}=\alpha\mathbf{V}, \tag{71}$$

where $\alpha\neq 0$. Certainly, $\mathbf{Q}$ of Eq. (71) is symmetric and invertible but not necessarily divergence-free. The latter condition requires $\operatorname{div}(\alpha\mathbf{V})=0$ which in turn may be expressed $\nabla\ln\alpha=-\mathbf{V}^{-1}\operatorname{div}\mathbf{V}$. The necessary and sufficient condition that $\mathbf{V}^{-1}\operatorname{div}\mathbf{V}$ is the gradient of a scalar function, and hence α can be found which makes $\mathbf{Q}$ of Eq. (71) possible, is that $\mathbf{V}$ satisfy

$$\operatorname{curl}\mathbf{V}^{-1}\operatorname{div}\mathbf{V}=0. \tag{72}$$

This condition is not very useful. It does, however, indicate that the possibility of achieving isotropic $\boldsymbol{\rho}$ depends on the underlying finite deformation; there is a subset of general deformations that can yield isotropic inertia. The deformation gradient $\mathbf{F}$ has nine independent elements, $\mathbf{V}$ has six, and $\mathbf{R}$ has three. The condition (72) is therefore a differential constraint on six parameters. We now demonstrate an alternative statement of the condition $\operatorname{div}(\alpha\mathbf{V})=0$ in terms of the rotation $\mathbf{R}$. This will turn out to be more useful, leading to general forms of potential deformation gradients.

Substituting $\mathbf{F}=\alpha^{-1}\mathbf{Q}\mathbf{R}$ into the identity (19a) and using Eq. (45) imply that

$$Q_{ij}(\Lambda^{-1}\alpha^{-1}R_{jK})_{,i}=0. \tag{73}$$

Using $\mathbf{Q}=\alpha\mathbf{F}\mathbf{R}^t$ and the identity $\mathbf{F}^t\nabla=\bar{\nabla}$ along with $\alpha\neq 0$ yield

$$R_{jM}(\Lambda^{-1}\alpha^{-1}R_{jK})_{,M}=0. \tag{74}$$

Then using the identity $(R_{jK}R_{jM})_{,M}=0$, Eq. (74) yields

$$\beta_{,K}=R_{jK}R_{jM,M}\leftrightarrow\bar{\nabla}\beta=\mathbf{R}^t\,\mathrm{Div}\,\mathbf{R}^t, \tag{75}$$

where $\beta=-\ln(\Lambda\alpha)$.

The necessary and sufficient condition that Eq.(75) can be integrated to find β is $\bar{\nabla}\wedge\bar{\nabla}\beta=0$, or using Eq. (75)

$$\mathrm{Curl}\,\mathbf{R}^t\,\mathrm{Div}\,\mathbf{R}^t=0. \tag{76}$$

The integrability condition (76) is in general not satisfied by $\mathbf{R}$, except in trivial cases. Norris[12] noted that isotropic density can be obtained if $\mathbf{R}=\mathbf{I}$. This corresponds to $\beta=$const, and it can be realized more generally if $\mathbf{R}$ is constant. Hence,

Lemma 1. *If the rotation* $\mathbf{R}$ *is constant then a normal acoustic fluid can be mapped to a unique metafluid with isotropic inertia:*

$$E_0=\kappa_0(\mathrm{Div}\,\mathbf{U})^2+\rho_0\dot{\mathbf{U}}\cdot\dot{\mathbf{U}}\quad in\ \Omega, \tag{77}$$

is equivalent to the current energy density

$$E=\lambda(\mathbf{V}:\nabla\mathbf{u})^2+\rho\dot{\mathbf{u}}\cdot\dot{\mathbf{u}}\quad in\ \omega, \tag{78}$$

where

$$\lambda=\Lambda^{-1}\kappa_0,\quad \rho=\Lambda^{-1}\rho_0. \tag{79}$$

The total mass of the deformed region ω *is the same as the total mass contained in* Ω.

The parameter λ is used to distinguish it from $\kappa=\Lambda\kappa_0$, because in this case $\mathbf{Q}=\Lambda^{-1}\mathbf{V}$. Also, the displacement fields are related simply by $\mathbf{u}=\mathbf{R}\mathbf{U}$.

As an example of a deformation satisfying Lemma 1: $\mathbf{x}=f(\mathbf{X}\cdot\mathbf{A}\mathbf{X})\mathbf{A}\mathbf{X}$ for any constant positive definite symmetric $\mathbf{A}$. This type of finite deformation includes the important cases of radially symmetric cloaks. Thus, Norris[12] showed that radially symmetric cloaks can be achieved using pentamode materials with isotropic inertia.

B. General condition on the rotation

The results so far indicate that isotropic inertia is achievable for transformation deformations with constant rotation. We would, however, like to understand the broader implications of Eq. (76). The rotation can be expressed in Euler form

$$\mathbf{R}=\exp(\theta\,\mathrm{axt}\,\mathbf{a}), \tag{80}$$

where θ is the angle of rotation, the unit vector $\mathbf{a}$ is the rotation axis, and the axial tensor $\mathrm{axt}(\mathbf{a})$ is a skew symmetric tensor defined by $\mathrm{axt}(\mathbf{a})\mathbf{b}=\mathbf{a}\wedge\mathbf{b}$. The vector $\theta\mathbf{a}$ encapsulates the three independent parameters in $\mathbf{R}$. The integrability condition (76) is now replaced with a more explicit one in terms of $\theta(\mathbf{X})$ and $\mathbf{a}(\mathbf{X})$. It is shown in the Appendix that

$$\mathbf{R}^t\,\mathrm{Div}\,\mathbf{R}^t=\mathbf{a}\wedge\bar{\nabla}\theta+\mathbf{Z}, \tag{81}$$

where the vector $\mathbf{Z}$ follows from Eq. (A6). In particular, $\mathbf{Z}$ vanishes if the axis of rotation $\mathbf{a}$ is constant. In general, for arbitrary spatial dependence, Eq. (81) implies that the integrability condition (76) is equivalent to the following constraint on the rotation parameters:

$$\mathbf{a}\bar{\nabla}^2\theta-(\mathbf{a}\cdot\bar{\nabla})\bar{\nabla}\theta+(\bar{\nabla}\theta\cdot\bar{\nabla})\mathbf{a}-(\bar{\nabla}\cdot\mathbf{a})\bar{\nabla}\theta+\mathrm{Curl}\,\mathbf{Z}=0. \tag{82}$$

In summary,

Lemma 2. *If the rotation* $\mathbf{R}$ *satisfies* Eq. (76) *or equivalently, if* θ *and* $\mathbf{a}$ *satisfy the condition* (82)*, then a normal acoustic fluid can be mapped to a unique metafluid with isotropic inertia according to* Eqs. (77) and (78) *with*

$$\lambda=\Lambda^{-1}e^{-2\beta}\kappa_0,\quad \rho=\Lambda^{-1}e^{-2\beta}\rho_0, \tag{83}$$

where the function β *is defined by* Eq. (75).

C. Simplification in 2D

The integrability condition (76) simplifies for the important general configuration of 2D spatial dependence. In this case $\mathbf{a}$ is constant, $\theta=\hat{\theta}(\mathbf{X}_\perp)$ where $\mathbf{X}_\perp\cdot\mathbf{a}=0$. Equation (82) then reduces to

$$\bar{\nabla}^2\hat{\theta}=0, \tag{84}$$

1. Example

Consider finite deformations with inhomogeneous rotation

$$\theta=\theta_0+\gamma X_1,\quad \mathbf{a}=\mathbf{e}_3, \tag{85}$$

for constants θ_0 and γ. This satisfies Eq. (84) and therefore Eq. (75) can be integrated. The metafluid in ω has isotropic density and pentamode stiffness given by Lemma 2, where $\beta=\gamma(X_2-X_{02})$. The constants θ_0 and X_{02} may be set to zero, with no loss in generality.

As an example of a deformation that has rotation of the form (85), consider deformation of the region $\Omega=[-\pi/2\gamma,\pi/2\gamma]\times[0,L]\times\mathbb{R}$ according to

$$\begin{pmatrix}x_1\\x_2\\x_3\end{pmatrix}=\begin{bmatrix}A_{11}&A_{12}&0\\A_{12}&A_{22}&0\\0&0&\alpha\end{bmatrix}\begin{pmatrix}\gamma^{-1}\sin\gamma X_1e^{-\gamma X_2}\\\gamma^{-1}(1-\cos\gamma X_1e^{-\gamma X_2})\\X_3\end{pmatrix}, \tag{86}$$

where $\alpha>0$ and the 2×2 symmetric matrix $\mathbf{A}$ with elements A_{ij} is positive definite. The deformation gradient is $\mathbf{F}=\mathbf{V}(X_2)\mathbf{R}(X_1)$ with $\mathbf{V}=\mathbf{A}e^{-\gamma X_2}+\alpha\mathbf{e}_3\mathbf{e}_3$, and

$$\mathbf{R}=\begin{bmatrix}\cos\gamma X_1&-\sin\gamma X_1&0\\\sin\gamma X_1&\cos\gamma X_1&0\\0&0&1\end{bmatrix}. \tag{87}$$

The mapped metafluid is defined by the energy density in ω

$$E = \frac{1}{\alpha \det \mathbf{A}}[\kappa_0[(\mathbf{A}:\nabla\mathbf{u})^2 e^{-2\gamma X_2} + (\alpha u_{3,3})^2] + \rho_0 \dot{\mathbf{u}} \cdot \dot{\mathbf{u}}].$$

In particular, it has isotropic inertia.

This example is not directly applicable in modeling a complete acoustic cloak. However, it opens up the possibility of patching together metafluids with different local properties, each with isotropic inertia so that the entire cloak has isotropic mass density.

VII. SUMMARY AND CONCLUSION

Whether it is the simple 1D acoustic mirage of Fig. 2 or a three-dimensional acoustic cloak, we have seen how acoustic stealth can be achieved using the concept of domain transformation. The fluid in the transformed region exactly replicates the acoustical properties of the original domain. The most general class of material that describes both the mimic and the mimicked fluids is defined as an acoustic metafluid. A general procedure for mapping/transforming one acoustic metafluid to another has been described in this paper.

The results, particularly Theorem 1 in Sec. IV, show that acoustic metafluids are characterized by as few as two parameters $(\rho, \sqrt{\kappa})$ and as many as 12 $(\boldsymbol{\rho}, \sqrt{\kappa}\mathbf{Q})$. This broad class of materials can be described as pentamode materials with anisotropic inertia. It includes the restricted set of fluids with anisotropic inertia and isotropic stress $(\mathbf{Q}=\mathbf{I})$.

The arbitrary nature of the divergence-free tensor $\mathbf{Q}$ adds an enormous amount of latitude to the stealth problem. It may be selected in some circumstances to guarantee isotropic inertia in cloaking materials, examples of which are given elsewhere.[12] In this paper we have derived and described the most general conditions required for $\boldsymbol{\rho}$ to be isotropic. The conditions have been phrased in terms of the rotation part of the deformation, $\mathbf{R}$. If this is a constant then the cloaking metafluid is defined by Lemma 1. Otherwise the condition is Eq. (76) with the metafluid given by Lemma 2. The importance of being able to use metafluids with isotropic inertia should not be underestimated. Apart from the fact that it resolves questions of infinite total effective mass[12] isotropic inertia removes frequency bandwidth issues that would be an intrinsic drawback in materials based on anisotropic inertia.

This paper also describes, for the first time, some of the unusual physical features of acoustic metafluids. Strange effects are to be expected in static equilibrium, as illustrated in Figs. 4 and 5. These properties can be best understood through realization of acoustic metafluids, and a first step in that direction is provided by the type of microstructure depicted in Fig. 6. The macroscopic homogenized equations governing the microstructure are assumed in this paper to be those of normal elasticity. It is also possible that the large contrasts required in acoustic metafluids could be modeled with more sophisticated constitutive theories, such as nonlocal models or theories involving higher order gradients. There is considerable progress to be made in the modeling, design, and ultimate fabrication of acoustic metafluids.

In addition to the degrees of freedom associated with the tensor $\mathbf{Q}$, the properties of metafluids depend on the finite deformation through $\mathbf{V}$. Even in the simple example of the 1D mirage, one could arrive at the lower picture in Fig. 2 through different finite deformations. This raises the question of how to best choose the nonunique deformation gradient $\mathbf{F}$. The present results indicate some strategies for choosing $\mathbf{F}$ to ensure the cloak inertia has isotropic mass, and the cloaking properties are in effect determined by the elastic pentamodal material. Li and Pendry[1] considered other optimal choices for the finite deformation. Combined with the enormous freedom afforded by the arbitrary nature of $\mathbf{Q}$, there are clearly many optimization strategies to be considered.

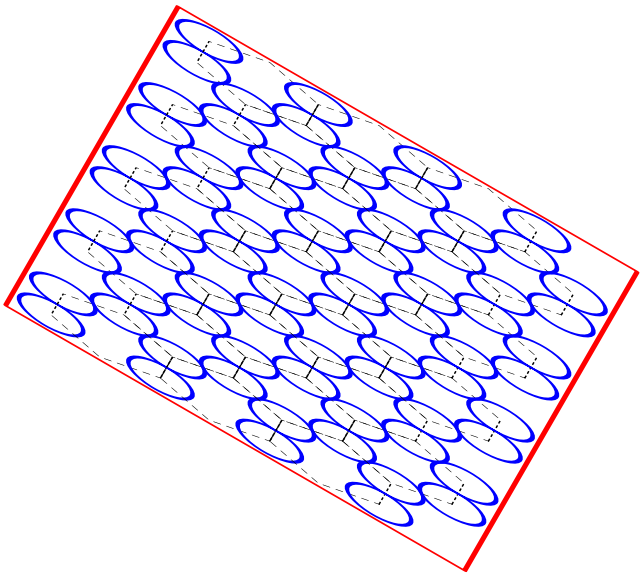


FIG. 6. (Color online) This shows a possible microstructure for a microscopic rectangular region of the metafluid in Figs. 4 and 5, see the latter. The sides of the rectangle are aligned with the principal axes of $\mathbf{Q}$. The metafluid is a pentamode material comprising a regular array of small beads such that each is in lubricated point contact with its three neighbors (2D). The dashed lines indicate the directions of the forces acting between the small oval-shaped beads. Although the structure as shown is unstable under shear, a realistic metafluid might contain some stabilizing mechanisms to enhance its rigidity. Details will be provided in a forthcoming paper.

ACKNOWLEDGMENTS

Thanks to the Laboratoire de Mécanique Physique at the Université Bordeaux 1 for hosting the author, in particular, Dr. A. Shuvalov, and to the reviewers.

APPENDIX: DERIVATION OF EQUATION (81)

Equation (80) can be written in Euler–Rodrigues form.[21]

$$\mathbf{R} = \mathbf{I} + \sin\theta\mathbf{S} + (1-\cos\theta)\mathbf{S}^2, \tag{A1}$$

where

$$\mathbf{S} \equiv \mathrm{axt}(\mathbf{a}) = \begin{bmatrix} 0 & -a_3 & a_2 \\ a_3 & 0 & -a_1 \\ -a_2 & a_1 & 0 \end{bmatrix}. \tag{A2}$$

Equation (A1) can be used, for instance, to find $\theta = \cos^{-1}\left[\frac{1}{2}(1-\mathrm{tr}\,\mathbf{R})\right]$ and hence $\mathbf{a}$ from $\mathbf{S}=(\mathbf{R}-\mathbf{R}^t)/(2\sin\theta)$. For the sake of notational brevity, in the remainder of the Appendix $c=\cos\theta$ and $s=\sin\theta$.

Explicit differentiation of Eq. (A1) yields

$$R_{iJ,J} = (cS_{iJ} - sP_{iJ})\theta_{,J} + sS_{iJ,J} - (1-c)P_{iJ,J}, \tag{A3}$$

where $\mathbf{P}=-\mathbf{S}^2=\mathbf{I}-\mathbf{aa}$. Noting that $S_{iJ,J}=-(\bar{\nabla}\wedge\mathbf{a})_i$ and $P_{iJ,J}=-a_i\bar{\nabla}\cdot\mathbf{a}-\mathbf{a}\cdot\bar{\nabla}a_i$, implies

$$\text{Div}\,\mathbf{R}^t = c\mathbf{a}\wedge\bar{\nabla}\theta - s\bar{\nabla}\wedge\mathbf{a} - s(\mathbf{I}-\mathbf{aa})\cdot\bar{\nabla}\theta + (1-c) \times[\mathbf{a}(\bar{\nabla}\cdot\mathbf{a}) + (\mathbf{a}\cdot\bar{\nabla})\mathbf{a}]. \tag{A4}$$

Multiplying by $\mathbf{R}^t$ using Eq. (A1) gives after some elimination and simplification,

$$\mathbf{R}^t\,\text{Div}\,\mathbf{R}^t = \mathbf{a}\wedge\bar{\nabla}\theta + \mathbf{Z}, \tag{A5}$$

where

$$\mathbf{Z} = -s[c\mathbf{I} + (1-c)\mathbf{aa}]\cdot\bar{\nabla}\wedge\mathbf{a} + (1-c)\mathbf{a}(\bar{\nabla}\cdot\mathbf{a}) - (1-c)[\mathbf{I} - (1-c)\mathbf{aa}]\cdot(\mathbf{a}\cdot\bar{\nabla})\mathbf{a} - s(1-c)\mathbf{a}\wedge(\mathbf{a}\cdot\bar{\nabla})\mathbf{a}. \tag{A6}$$

[1] J. Li and J. B. Pendry, "Hiding under the carpet: A new strategy for cloaking," Phys. Rev. Lett. **101**, 203901 (2008).
[2] S. A. Cummer and D. Schurig, "One path to acoustic cloaking," New J. Phys. **9**, 45 (2007).
[3] H. Chen and C. T. Chan, "Acoustic cloaking in three dimensions using acoustic metamaterials," Appl. Phys. Lett. **91**(18), 183518 (2007).
[4] L.-W. Cai and J. Sánchez-Dehesa, "Analysis of Cummer-Schurig acoustic cloaking," New J. Phys. **9**, 450 (2007).
[5] S. A. Cummer, B. I. Popa, D. Schurig, D. R. Smith, J. Pendry, M. Rahm, and A. Starr, "Scattering theory derivation of a 3D acoustic cloaking shell," Phys. Rev. Lett. **100**, 024301 (2008).
[6] A. Greenleaf, Y. Kurylev, M. Lassas, and G. Uhlmann, "Comment on "Scattering theory derivation of a 3D acoustic cloaking shell," e-print arXiv:0801.3279v1.
[7] A. N. Norris, "Acoustic cloaking in 2D and 3D using finite mass," e-print arXiv:0802.0701.
[8] Y. Cheng, F. Yang, J. Y. Xu, and X. J. Liu, "A multilayer structured acoustic cloak with homogeneous isotropic materials," Appl. Phys. Lett. **92**, 151913 (2008).
[9] D. Torrent and J. Sánchez-Dehesa, "Anisotropic mass density by two-dimensional acoustic metamaterials," New J. Phys. **10**, 023004 (2008).
[10] D. Torrent and J. Sánchez-Dehesa, "Acoustic cloaking in two dimensions: A feasible approach," New J. Phys. **10**, 063015 (2008).
[11] H.-Y. Chen, T. Yang, X.-D. Luo, and H.-R. Ma, "The impedance-matched reduced acoustic cloaking with realizable mass and its layered design," Chin. Phys. Lett. **25**, 3696–3699 (2008).
[12] A. N. Norris, "Acoustic cloaking theory," Proc. R. Soc. London, Ser. A **464**, 2411–2434 (2008).
[13] G. W. Milton, M. Briane, and J. R. Willis, "On cloaking for elasticity and physical equations with a transformation invariant form," New J. Phys. **8**, 248–267 (2006).
[14] The name pentamode is based on the defining property that the material supports five easy modes of infinitesimal strain. Sec. V for details.
[15] G. W. Milton and A. V. Cherkaev, "Which elasticity tensors are realizable?," J. Eng. Mater. Technol. **117**, 483–493 (1995).
[16] U. Leonhardt and T. G. Philbin, "Transformation optics and the geometry of light," e-print arXiv:0805.4778.
[17] The bijective property of the mapping does not extend to acoustic cloaks, where there is a single point in Ω mapped to a surface in ω, see Ref. 12 for details.
[18] R. W. Ogden, *Non-Linear Elastic Deformations* (Dover, New York, 1997).
[19] A. Greenleaf, Y. Kurylev, M. Lassas, and G. Uhlmann, "Full-wave invisibility of active devices at all frequencies," Commun. Math. Phys. **275**, 749–789 (2007).
[20] M. J. P. Musgrave, *Crystal Acoustics* (Acoustical Society of America, New York 2003).
[21] A. N. Norris, "Euler-Rodrigues and Cayley formulas for rotation of elasticity tensors," Math. Mech. Solids **3**, 243–260 (2008).

Semantic evaluations of noise with tonal components in Japan, France, and Germany: A cross-cultural comparison

Hans Hansen and Reinhard Weber
Acoustics Group, Institute of Physics, Carl von Ossietzky University, Carl-von-Ossietzky-Straße 9-11, 26111 Oldenburg, Germany

(Received 13 February 2008; revised 12 November 2008; accepted 18 November 2008)

An evaluation of tonal components in noise using a semantic differential approach yields several perceptual and connotative factors. This study investigates the effect of culture on these factors with the aid of equivalent listening tests carried out in Japan ($n=20$), France ($n=23$), and Germany ($n=20$). The data's equivalence level is determined by a bias analysis. This analysis gives insight in the cross-cultural validity of the scales used for sound character determination. Three factors were extracted by factor analysis in all cultural subsamples: pleasant, metallic, and power. By employing appropriate target rotations of the factor spaces, the rotated factors were compared and they yield high similarities between the different cultural subsamples. To check cross-cultural differences in means, an item bias analysis was conducted. The a priori assumption of unbiased scales is rejected; the differences obtained are partially linked to bias effects. Acoustical sound descriptors were additionally tested for the semantic dimensions. The high agreement in judgments between the different cultural subsamples contrast the moderate success of the signal parameters to describe the dimensions. © *2009 Acoustical Society of America.* [DOI: 10.1121/1.3050275]

PACS number(s): 43.50.Ba, 43.50.Cb [DOS] Pages: 850–862

I. INTRODUCTION

Knowledge of cultural differences is a valuable resource, be it for economic or social reasons. Differences between cultures have their very start on perceptual dimensions, and when it comes to auditory perception, comparing cultures on perceptual and connotative factors in sound quality investigations can lead to significant input on how to optimize product sound design for a heterogeneous and international market. This is, for instance, true for the automotive industry (Hussain *et al.*, 1998), but it also applies for products such as household appliances, multimedia technology, luxury articles, etc. (Guski, 1997; Blauert and Jekosch, 1997). Moreover, investigating how different cultures perceive sound is essential with respect to noise metrics, which are used to establish international noise evaluation standards of high objectivity. These standards recommend reporting and evaluating prominent discrete tones as well as impulsive aspects for a more in-depth noise characterization [e.g., ISO 7779:1999, (E) 1999]. But are, for instance, particular sounds evaluated equally across cultures? Sound evaluations are commonly carried out with sound-describing words, or adjectives. They are the basis for measuring perceptual and connotative factors in human beings, i.e., the affective meaning people attribute to objects or stimuli (Osgood *et al.*, 1957). The comprehension and comparability of such adjectives are crucial for assessing product (sound) quality as well as noise annoyance and related reactions to noise.

In the past, several studies on cross-cultural psychoacoustics have been conducted which deal with differences in noise perception (Kuwano *et al.*, 1986; Namba *et al.*, 1991a; Schick and Hoege, 1996) as well as with the sound character of products and musical excerpts (Iwamiya and Zhan, 1997; Kuwano *et al.*, 2006, 2007). In starting with noise issues, Kuwano *et al.* (1986) performed a study on neighborhood noise in which they evaluated the semantic profiles of loudness, annoyance, and noisiness concepts in three countries. They found that the profiles are stable over the years and seem to be equal with one notable difference: the Japanese and German concepts of loudness are rather affectively neutral compared to the English concept. Namba *et al.* (1991a) investigated the verbal expressions "loud," "noisy," and "annoying" in several countries such as Japan and the USA with the method of selective description, yielding different usages in each country. Schick and Hoege (1996) not only reviewed studies on neighborhood noise but specifically addressed problems commonly found in cross-cultural research such as stimuli selection. Focusing on differences within factorial dimensions of a semantic differential, these researchers found that the German subsample showed a more "negatively tuned judgmental structure" (p. 311) while the Japanese participants exhibited a more neutral structure.

To follow with examples dealing with sound character of products and musical excerpts, Iwamiya and Zhan (1997) studied the difference between Japanese and Chinese college students evaluating musical excerpts. Obtaining the factors sharpness, cleanness, and potency for their semantic differential, the Japanese students used the item "pleasant" independently from the three factors. They noted that the same literal meaning does not result in similar usage in the auditory domain. Kuwano *et al.* (2006) reported data from a cross-cultural car door sound evaluation. In congruence with the previously described findings, their semantic differential revealed a three-dimensional factorial space in both groups. Differences in the usage of adjectives were observed with regard to the adjectives noisy and powerful. In an experiment

implementing the semantic differential to evaluate auditory warning signals, Kuwano *et al.* (2007) found similar results in the USA, Germany, and Japan.[1]

A three-factor solution for the evaluation of acoustical stimuli is prevalent in most studies. Similarly, a three-factor solution has been reported in several countries regarding the affective meaning of concepts (Osgood *et al.*, 1975), and they have been labeled as evaluation, power, and activity (EPA) structure. The EPA structure can be related to the three-factor structure obtained by Namba *et al.* (1992) in a study to judge artificial sounds: pleasant, power, and metallic/timbre. The evaluation factor corresponds to the pleasant factor, while the power factors correspond directly to each other. The metallic/timbre factor corresponds to the activity factor because it describes modal features of a sound. In summarizing the studies mentioned above, major differences were obtained mainly with regard to the concept of loudness and noisiness. Also, all of the studies mentioned used culture as an independent variable.[2] Issues of cultural bias and equivalence, which describe the comparability of data across cultures, have yet to be explicitly included in methodological paradigms, although they have been discussed (e.g., Schick and Hoege, 1996). This study incorporates them and they will be introduced in more detail in Sec. II A.

In the current study, the effect of tonal components in noise on the sound character is investigated, i.e., annoyance, loudness, and timbre descriptions in Japan, France, and Germany. This class of sounds, i.e., broadband noise containing prominent tones, is widely encountered in the environment, e.g., machine noise, noise in passenger compartments of trains and cars, wind turbine noise, and fan noise in electrical appliances. Several national and international norms on noise emissions accommodate the fact that prominent tonal components have a large influence on sound character. The international standard advises to report these components separately [ISO 7779:1999 (E), 1999] in order to describe sound character more exhaustively or an additional penalty is added to the physical level for such sounds (DIN 45681:2005-03, 2005). In order to investigate this class of sounds and to systematize the effect of tonal components on sound character, a set of artificial sounds was generated for the current study. They were used to investigate their effect on sound character, i.e., annoyance, loudness, and timbre descriptions using a semantic differential which was carefully translated into every tested culture's language. Schick and Hoege (1996) argued that artificial stimuli in a cross-cultural context with a nonrecognizable sound source have two major advantages. First, the listener does not evaluate the source they attribute the sound to. Second, they are not attributed to different sources depending on culture. The stimuli are comprised of tonal components, one or two sine signals, which were added to Brownian noise (power density $\propto 1/f^2$). The noise was chosen due to its relatively low overall sharpness. The varying tonal content was implemented by varying the signal level of the sine(s), i.e., the level above masked threshold. The variation in a second parameter, the frequency of the second sine, allowed for the investigation of various ways to integrate tonal components in noise into a signal parameter, which is not covered by the current national and international standards [ISO 7779:1999 (E), 1999; DIN 45681:2005-03, 2005]. The stimuli therefore range from broadband noise with barely audible to highly prominent multiple tonal components.

To be able to cope with issues of cultural bias and equivalence in the current study, a paradigm developed in cross-cultural psychology by van de Vijver and Leung (1997) was introduced. A semantic differential, carefully translated into every tested culture's language, was used to explore the perceptual and connotative factors, i.e., the affective meaning, cross-culturally. Then, the method of van de Vijver and Leung (1997) was applied to identify bias and to determine the level of equivalence reached in the present investigation.

The current study was designed to research two main goals. The first aim is to test whether the three-factor solution for the affective meaning of sounds can be generalized cross-culturally. Given this basis, an analysis can subsequently be performed in order to relate the magnitude of possible cross-cultural differences on certain scales to the magnitude of bias. This issue is linked to the second goal of the study: the application of acoustical sound descriptors to noise with prominent tones in a cross-cultural context.

II. METHOD

A. Methodological background

Using a semantic differential, carefully translated into the respective mother languages, participants from Japan, France, and Germany judged the stimuli, consisting of tones in noise. A principal component analysis (PCA) was performed on the obtained semantic differentials and the paradigm reported by van de Vijver and Leung (1997) served as a guideline to compare the cross-cultural data from the listening tests. This paradigm is distinguished by two major opposing concepts: equivalence and bias of the experimental setting.[3] "Data are equivalent when an observed cross-cultural difference on a measurement scale is matched by a corresponding difference on the comparison scale" (Poortinga, 1989, p. 738). Equivalence is jeopardized by various forms of bias, and "scores are equivalent when they are unbiased" (van de Vijver and Leung, 1997, p. 7).

The concept of equivalence is characterized by different levels in ascending order: construct equivalence, measurement unit equivalence, and scalar equivalence. Analyzing equivalence is hierarchical, and the different levels must be established on the basis of the preceding ones. Optimally, scalar equivalence would be reached for the data to be compared. Once equivalence of the data has been proven in a cross-cultural comparison, the remaining differences can be characterized as valid cross-cultural differences. Hence the decisive reason to determine different levels of equivalence is the ability to identify valid cross-cultural differences thereafter. Proven bias destructs equivalence and bias is classified according to its influence on the level of equivalence reached: construct, method, and item bias (Poortinga, 1989, Table 2, P. 745). In the following, the different levels of equivalence along with their vulnerability to bias types are described.

1. Construct equivalence

Construct equivalence is achieved when the construct measured is identical across the cultures investigated. Deviating from construct equivalence leads to construct bias, e.g., the incomplete overlap of construct definitions, such as the construct *intelligence* in various cultures. The three independent factors obtained in sound evaluation studies, i.e., pleasant, metallic/timbre, and power (e.g., Solomom, 1958; Namba *et al.*, 1992), represent the construct investigated in this study. To test construct equivalence, the semantic differentials for all three countries were analyzed separately by PCA in order to reveal these independent factors. An initial comparison of the factorial structure should reveal any major differences in the construct formation. If the number of perceptual space dimensions is equal in all three cultural subsamples, target rotation toward an arbitrary factor solution will be performed. A comparison between the rotated and the target structure can show whether and how the perceptual dimensions differ (Watkins, 1989). The target rotation is a necessary step before comparing the perceptual dimensions because they are not necessarily congruent after factor analysis. The factors are to be named equally for all three cultural groups investigated after determining a common set of marker adjective items, i.e., the adjective items which represent the factor in the best way.

To quantify the relational agreement between the target and the rotated factor solutions, two coefficients are used; Tucker's ϕ and Pearson's correlation coefficient r_{xy}.[4] Tucker's ϕ, the congruence coefficient, while sensitive to a additive constant, is insensitive toward multiplications, i.e., if all factor loadings of one group have a proportional relationship to the factor loadings of the other group, no difference will be detected and Tucker's ϕ is 1.

For Tucker's ϕ no sampling distribution is known (Korth and Tucker, 1975). Nevertheless, the "significance level" can be estimated using Monte Carlo simulations. Korth (1978) reported a significance level (α=0.05) of 0.93 (four factors/ten variables). Korth (1978) stated that Tucker's ϕ is lower for less factors and more variables. A coefficient of 0.93 can be viewed as a conservative estimation for the significance level in this study.

The correlation coefficient r_{xy} indicates the strength and direction of a linear relationship between the factors, i.e., it is influenced neither by addition nor multiplication. In this way, Tucker's ϕ and the correlation coefficient quantitatively indicate the construct equivalence.

Another type of bias affecting construct equivalence, the fundamental level of equivalence, is method bias. It involves issues such as differential social desirability, stimulus familiarity, lack of sample comparability, etc., and therewith all issues concerning experimental methods. The different forms of method bias will be discussed throughout Secs. II B–II F.

2. Measurement unit equivalence and scalar equivalence

After focusing on the perceptual and connotative factors, the analysis will concentrate on the detection of valid cross-cultural differences based on the adjective items in the semantic differential. Therefore, the concepts of item bias and scalar equivalence, as opposites, will be explained in detail before the analysis will be introduced.

With *cross-cultural differences* we imply differences in item means across the three cultural subsamples tested. Therefore, measurement unit and scalar equivalence have to be established beforehand. After determining construct bias and ruling out method bias, item bias is the major bias source left. In other words, based on equivalent constructs, it will be tested whether the items are used in the same way.

Van de Vijver and Leung (1997) defined how to determine whether a measure is free of item bias: "… persons with an equal standing to the theoretical construct underlying the instrument should have the same expected score on the item, irrespective of group membership" (p. 69). This, however, does not imply equality of item means across cultures. Instead, on an individual level, this implies that people "with an equal standing" regarding a particular item just score the same. Therefore, a true difference between groups for an item on a bias free scale is the true difference in means for that particular item. Consequently, a bias analysis of the adjective items should precede the analysis of the "impact" because the items used in the cross-cultural context are not necessarily bias-free.

Here is an example to facilitate understanding. It is possible that the groups tested in this study use the adjective item pleasant/unpleasant in different ways. Imagine German participants express a high degree of pleasantness determined by a general high score on all adjectives associated with the pleasant factor. Yet concerning the adjective pair pleasing/unpleasing, they score a medium value, i.e., a 3 on a 1–7 rating scale. Furthermore, imagine Japanese participants also showing a high degree of pleasantness via the overall value, yet also exhibiting a high score on the pleasant/unpleasant item. In this case, comparing the scales at face value for this item would be misleading because the two groups have an "equal standing" although the Japanese group uses the pleasant/unpleasant item differently, i.e., by rating it higher than the German group.

In the following, an analysis capable of unraveling the depicted types of item biases will be described. The adjective items associated with the factors determined via PCA will be classified according to their strength in representing these factors. The corrected discriminatory power r or item-to-dimension total correlation will be calculated for each item. To generate the total score, all items of a specific dimension are summed, while the item under scrutiny is left out. For each factor, the corrected discriminatory power r is calculated by correlating the corrected total score with the item scores. Discriminatory power r of $(0.5>r>0.3)$ is classified as "medium" and $(r\geqslant 0.5)$ as "high" (e.g., Bortz and Doering, 2002). The homogeneity of the items representing a particular factor will be tested by calculating Cronbach's α in order to qualify the item composition. Values $\alpha>0.7$ denote a relatively high overall consistency (Bortz and Doering, 2002). If a strong correlation between the factor and the adjective items is found, i.e., marker items have high internal consistency and are appropriate manifestations of the under-

lying factors, further analyses will be conducted to test cultural differences between the scores of single adjective items.

To identify item bias in numerical scores, the equal standing of participants regarding one perceptual dimension must be operationalized. This is done by using the total factor score as an equal standing indicator. The total factor score is the sum of all respective marker items. For each stimulus, each participant, as well as each factor, one total factor score is calculated. The range of total factor scores is then divided into subranges containing approximately the same amount of total factor scores for each factor. These subranges are called score levels. They are indicators for equal standing of participants for a stimulus with respect to one particular factor. Therefore, two participants show an equal standing regarding one stimulus with reference to the perceptual factor if the factor levels are equal in both cases. Item biases can then be detected by running multivariate ANOVAS on item means with score levels and culture as the independent variables. The paradigm reported by van de Vijver and Leung (1997) is applied with these procedures, and culture as a main effect as well as the interaction between culture and score level indicate cross-cultural differences. On the one hand, culture as a main effect represents whether an item is generally used differently, i.e., higher or lower (uniform bias). On the other hand, the interaction effect indicates a change in the usage or interpretation of an item across score levels (nonuniform bias). Therefore, assuming that there is no construct and method bias, measurement unit equivalence only results when nonuniform bias is not found between cultures because the measurement unit is not affected by a constant offset on an item. However, scalar equivalence, i.e., full score comparability, is only reached if the uniform bias can be excluded, as only then a common scale origin can be assumed (Poortinga, 1989, Table 2, p. 745).

3. Testing differences between means: The impact of culture

If construct equivalence is confirmed and no item bias is found, an additional MANOVA can reveal the influence of culture on the unbiased adjective items. Furthermore, obtained differences on biased items will be related to the magnitude of bias discovered. After excluding all possible bias, the identifying valid cross-cultural differences will be the concluding step in analyzing differences in this cross-cultural study (Berry *et al.*, 2002).

4. Correlation of perceptual dimensions with acoustical sound descriptors

After identifying different perceptual dimensions, it is of special interest to investigate how they are related to calculated acoustic and psychoacoustic parameters. In this study, a measure describing the tonal contents ΔL, a sound pressure level (SPL) parameter $L_{Aeq,T}$, and the sharpness S was used. ΔL, the signal-to-noise (S/N) ratio above masked threshold, was investigated because a S/N approach is suggested by various international and national norms to describe the tonal contents [ISO 7779:1999(E), 1999; DIN 45681:2005-03]. Furthermore, the $L_{Aeq,T}$, a common level measure in noise evaluations (Namba and Kuwano, 1984; Marquis-Favre *et al.*, 2005), will be correlated with the perceptual dimensions. Sharpness S (von Bismarck, 1984) has been identified as a major source of auditory unpleasantness (Zwicker and Fastl, 1999; Zimmer *et al.*, 2004).[5] Relating and describing the factors with these signal parameters is expected to allow an improved understanding and interpretation of the perceptual dimensions.

TABLE I. Stimuli with one sine (frequency f_1=500 Hz) with varying S/N ratios embedded in Brownian noise (red noise).

S_i	f_1 (Hz)	ΔL[a] (dB)	$L_{Aeq,6\ s}$[b] (dB)	S[c] (acum)
1	…	…	44.5	1.25
2	500	−7	44.5	1.25
3	500	−3	44.7	1.25
4	500	1	44.7	1.24
5	500	5	44.9	1.22
6	500	13	46.7	1.17
7	500	23	53.2	1.06

[a]Difference between sine f_1 and noise level within the critical band corrected by masking threshold (DIN 45681:2005-03, 2005).
[b]A-weighted L_{eq}, average time 6 s.
[c]Sharpness as defined by von Bismarck (1984).

B. Stimuli

The stimuli were various sinusoidal sounds at different frequencies with Brownian noise [red noise, $L_{Aeq,6\ s}$ =44.5 dB(A)] added in order to investigate the effect of tonal components on the perceptual and connotative factors. Brownian noise has lower sharpness resulting in a more pleasant sensation compared with other broadband noises, e.g., uniform masking noise or pink noise (Zwicker and Fastl, 1999). As tonal components often add to annoyance, Brownian noise serves as an adequate starting level.

The first set is comprised of noise with a single component added (f_1=500 Hz). The exact stimuli configurations are shown in Table I. Their salience is described by the measure ΔL. This is the ratio between the sine level and the level of the respective critical band corrected by masking threshold according to DIN 45681:2005-03 (2005), i.e., the level above masking threshold, which is based on Zwicker and Feldtkeller (1967) (Chap. 18). They used the Békésy method of audiometry, i.e., the calculated threshold estimates the 50% point on the psychometric function. As the goal is to investigate the effect of the perceptual S/N ratio on sound character, a range of −7–23 dB was chosen for ΔL. 0 dB marks the threshold, and −7 dB can therefore be considered barely audible because it is at the very end of the psychometric function. 13 and 23 dB mark clearly prominent tones that were included to investigate the effect large level changes of dominant tones have on sound character.

Apart from the perceptual S/N ratio, the aim is to investigate the effect of multiple tones on sound character. Therefore, three additional frequencies were chosen. ISO 7779:1999(E) (1999) gives a proximity limit Δf_{prox} in which the energy should be integrated into a single tone. As in the first set, f_1 was set to 500 Hz. The frequency of the addi-

TABLE II. Stimuli with two sine signals with varying S/N ratios embedded in Brownian noise (red noise). Both sines have the same SPL.

S_i	f_1 (Hz)	f_2 (Hz)	ΔL_1[a] (dB)	ΔL_2[b] (dB)	$L_{Aeq,6\ s}$ (dB)	ΔL_{sum}[c] (dB)	S[d] (acum)
8	500	530	−11	−10.5	44.5	−7.8	1.25
9	500	530	−5	−4.5	44.6	−1.8	1.23
10	500	530	1	1.5	44.8	4.2	1.22
11	500	530	7	7.5	45.7	10.2	1.19
12	500	530	13	13.5	48.2	16.2	1.14
13	500	530	19	19.5	52.3	22.2	1.07
14	500	600	−11	−10	44.5	−7.3	1.25
15	500	600	−5	−4	44.6	−1.3	1.24
16	500	600	1	2	44.8	4.7	1.23
17	500	600	7	8	45.9	10.7	1.19
18	500	600	13	14	48.3	16.7	1.12
19	500	600	19	20	52.6	22.7	1.05
20	500	1000	−11	−7	44.5	−4.5	1.25
21	500	1000	−5	−1	44.6	0.5	1.23
22	500	1000	1	5	44.8	6.5	1.22
23	500	1000	7	11	45.7	12.5	1.19
24	500	1000	13	17	48.1	18.5	1.13
25	500	1000	19	23	52.2	24.5	1.05

[a]Difference between sine f_1 and noise level within the critical band corrected by masking threshold (DIN 45681:2005-03, 2005).
[b]Difference between sine f_2 and noise level within the critical band corrected by masking thershold (DIN 45681:2005-03, 2005).
[c]Total level of both sines.
[d]Sharpness as defined by von Bismarck (1984).

tional tone f_2 is varied. f_2 was set to 530, 600, or 1000 Hz. 530 Hz lies within the proximity limit (Δf_{prox}, (500 Hz) =34 Hz [ISO 7779:1999(E), 1999]). 600 Hz lies within one critical band (Zwicker and Fastl, 1999), while 1000 Hz has a harmonic relation and lies outside the critical band. Every frequency condition was presented with six different sine levels. f_1 and f_2 were equal in SPL. The SPL was also equal across frequency conditions, i.e., the level of $f_2(S_8)$ equaled $f_2(S_{14})$. The lowest SPL was chosen to match the signal level above the calculated masked threshold of the lowest SPL in the first set (S_{20}, f_2=−7 dB). The signal levels covered the whole range of perception of prominent tones. The stimuli configurations are summarized in Table II.

The stimuli were generated digitally using MATLAB (Version 6.0, Mathworks Inc.) with a sampling frequency of 44 100 Hz. Every stimulus lasted for 6 s each including a 10 ms fade in and fade out. Both the noise and the sines had a common onset and offset. The stimuli levels were described by $L_{Aeq,6\ s}$ measured with a sound level meter (Ono Sokki ONO-LA5110) directly reproduced from a Sony Digital Audio Tape-Coder (TCD-D8). The sharpness S was calculated according to von Bismarck (1984).

C. Semantic differential

The selection of the adjective compilation is based on semantic differentials reported by Namba *et al.* (1991a, 1992) and concepts by Schick and Hoege (1996). An example is shown in Fig. 1. This enabled an a priori sampling

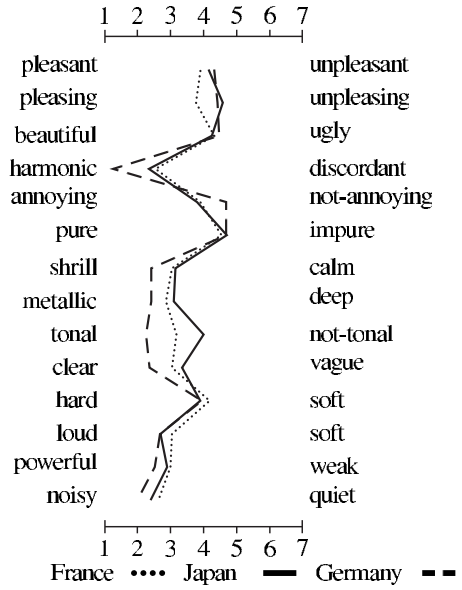


FIG. 1. Example of the semantic differential used in the experiment. Mean values of S_{17} are shown for every culture.

along the dimensions pleasant, metallic, and power, which guided adjective selection. The choice was carried out along these dimensions as the adjective choice alone is a source of method bias. The inclusion of Japanese semantic differentials mitigated this issue (Namba *et al.*, 1992; Kuwano *et al.*, 1994). The pleasant dimension was sampled a priori by adjectives such as *beautiful/annoying/clear* and *pleasant/pleasing*. The difference between pleasant and pleasing seems minimal in English because they represent an adjective and a verbal adjective from the same word stem. In the languages used in this study, however, they stem from different words. The metallic/timbre factor includes, apart from *metallic/shrill/harmonic/hard* (Namba *et al.*, 1992), adjectives which refer directly to the sound character such as *tonal*. The power factor is represented a priori by the adjectives *loud/powerful/noisy*. According to Schick and Hoege (1996) and Namba *et al.* (1991a), these adjectives cannot be grouped into a neutral power category in every culture, but in an a priori sampling this coarse approach not only suffices it is even a requirement for reducing method bias. Finally, the semantic differential chosen to evaluate sound perception consists of 14 adjective pairs each separated by a seven-step rating scale (Osgood *et al.*, 1957). The adjective pairs are presented in Table III in Japanese, French, and German, and an English translation was added for comprehension. The German version served as a reference for all other versions.

They were translated into Japanese, French, and German by native speaker professionals with experience in the semantic characterization of sounds. These translators were chosen according to individual proficiency. The translations were made with many precautions. Appropriate choices for translated adjectives were made based on their semantic and not their literal meaning in the target language.

D. Participants

Although a source of method bias, cross-cultural studies commonly treat culture as an independent variable for rea-

TABLE III. Adjective pairs and their translation as used in the experiments. (All Japanese transliterations are written according to the Hepburn system.)

English	Japanese	French	German
powerful/weak	hakuryo no aru/monotarinai	puissant/faible	kräftig/schwach
shrill/calm	kantakai/ochitsu ita	strident/calme	schrill/ruhig
clear/vague	hakkiri shita/bonyari shita	clair/flou	klar/vage
pure/impure	sunda/nigotta	pur/impur	rein/unrein
pleasant/unpleasant	kokosoyoi/fukanai	agréable/désagréable	angenehm/unangenehm
metallic/deep	kinsokuseino/fukami no aru	métallique/profond	metallisch/dumpf
noisy/quiet	yakamashii/shizukana	bruyant/doux	lärmend/still
hard/soft	katai/yawarakai	dur/mou	hart/weich
harmonic/discordant	chouwanotoreta/fuchouwana	harmonique/inharmonique	harmonisch/disharmonisch
tonal/not-tonal	pitchu/pitchu ga hakkirishinai	tonal/pas tonal	tonhaltig/nicht tonhaltig
loud/soft	ooki/chiisai	fort/doux	laut/leise
annoying/not-annoying	urusai/urusaikunai	gênant/pas gênant	lästig/nicht lästig
pleasing/unpleasing	monomashii/monomashikunai	plaisant/déplaisant	gefällig/ungefällig
beautiful/ugly	utsukushii/kitanai	beau/laid	schön/hässlich

sons of practicability (van de Vijver and Leung, 1997, pp. 2 and 3). Besides culture, sociodemographic and context variables were assessed: age, gender, experience in the field of acoustics, and educational status. 11 Japanese men and 9 Japanese women (average age: $22.9a \pm 1.7a$), 15 French men and 8 French women (average age: $20.5a \pm 1.8a$), as well as 12 German men and 8 German women (average age: $23.7a \pm 2.6a$) with normal hearing abilities participated in this study. Being *Japanese*, *French*, or *German* means participants have the respective mother tongue and are autochthon to the respective country. Furthermore, all were college students at the time of testing and approximately half of them majored in acoustics or related fields such as environmental engineering. As the cultures were matched according to these covariates, method bias was reduced at this stage.

The experiment was conducted in Japan from Nov. 2002 to Jan. 2003, in Germany from Nov. 2003 to Dec. 2003, and in France from Jun. to Oct. 2005.

E. Apparatus

As a necessary prerequisite for methodical consistency, stimuli presentation had to be equal in the different laboratories used. To guarantee the identical presentation of sound stimuli in all countries, the relevant equipment used was always the same. Failing to do so would have lead to major method bias. The stimuli were presented diotically via the same headphones (STAX SR-Lambda Pro, HA I.1) using the same amplifier (HEAD acoustics HPSIII.2) in sound-proof rooms at Osaka University, Japan, INSA in Lyon, France, and Oldenburg University, Germany. The transfer function of the headphones, measured with an artificial head, had a flat characteristic (± 3 dB) between 200 and 3000 Hz.

In Japan, the wav files were recorded on DAT (Sony DTC-ZE700) via an USB interface (Roland ED UA-30) and were reproduced with a DAT recorder (Pioneer D-05). In Germany and France the sounds were recorded on CD-ROM using a CD player (Philips CD618) for reproduction. The differences in the experimental setup relating to sound file recording and storage did not cause any significant changes in the stimulus presentation.

F. Procedure

The instructions and semantic differentials were handed out to participants as paper copies in their respective mother tongue. The participants were instructed to judge each stimulus as a whole according to the adjective pairs on the semantic differential. In order to get accustomed to the stimuli and to establish an adequate frame of reference, each participant first judged six test stimuli (S_2, S_6, S_{11}, S_{15}, S_{19}, and S_{22}), covering the perceptual range of all stimuli used. This phase also served for clarifying the usage of the adjective pairs in the context of this study. For the main experiment, both the stimulus presentation order and the semantic differential item order were randomized. The 25 stimuli were randomized in three different orders, and each participant listened to one order. Each participant evaluated the set with one of three randomized semantic differentials. This twofold randomization balances position effects, e.g., due to short-term memory. Presenting a stimulus consisted of repeating it three times, and repetitions were separated by a 4 s pause. The participants were allowed to start evaluating the stimulus as soon as its first presentation was over. Only after the participant finished evaluating a stimulus with the semantic differential, the next stimulus was presented. The entire experiment lasted for approximately 45 min. At the end of testing, three selected stimuli were replayed and judged again for a supplemental reliability analysis with the French and German samples. The stimuli chosen for repetition (S_{21}, S_1, and S_{25}) covered the entire evaluation range: Two extreme and one midrange stimuli were used.

Procedural differences can lead to method bias. In using the steps described here, the experimental procedure was equal in all cultures. The retest is one exception, but it did not lead to any bias because it was conducted after the experiment. By instructing participants to read the translated introduction without conveying additional information, the experimenter reduced the eminent sources of method bias at this stage.

III. RESULTS

The results are based on data obtained with the semantic differential described in Table III in Japan, France, and Ger-

many. To test reliability, a Pearson correlation was calculated for three stimuli which were evaluated twice by each participant in the French and German samples: 19 out of 20 German and 23 out of 24 French participants were considered reliable judges ($r_{crit,0.01}=0.40$). Participants who inconsistently judged repeated stimuli were excluded from further analysis. In Japan, no stimuli were repeated after initial testing, rendering a reliability analysis for this sample impossible. This may lead to additional unexplained variance in the Japanese data. Considering the solid reliabilities for the German and French samples as well as several successful semantic differential studies with Japanese participants by (e.g., Kuwano *et al.*, 1994, 2006), the fact that reliability cannot be calculated for the Japanese sample in this study may not jeopardize the overall interpretations of the results.

This section covers four issues: the equivalence of perceptual dimensions (Sec. III A), item bias analysis (Sec. III B), differences in means: the impact of culture (Sec. III C), and lastly acoustical sound descriptors in a cross-cultural research (Sec. III D).

A. The equivalence of perceptual dimensions

To analyze the 14 adjective pairs in the semantic differential, factor analyses were computed separately for the French, Japanese, and German samples. In all cases, three factors (Kaiser–Guttmann criterion, $EV>1$) were extracted using the PCA with Varimax rotation. The Kaiser–Meyer–Olkin measure of sampling adequacy equals 0.86 in the Japanese sample, 0.88 in the French sample, and 0.91 in the German sample. As such, the factor analysis was properly applied (Cureton and D'Agostino, 1983, p 389).

Differences between experimental samples are generally overestimated due to the various orientations of rotated solutions in factorial space. To avoid overestimation, the factor solutions for the French and the German samples were rotated toward the factor solution for the Japanese sample. The rotation procedure was chosen to achieve colinearity of each of the three dimensions, facilitating a comparison between factor loadings for each dimension. The target rotated factor solutions for the French sample and the German sample along with the Japanese Varimax-rotated solution are presented in Fig. 2.

For each factor, the following marker adjective pairs were identified as common descriptors for all samples:

Factor I: beautiful/ugly, harmonic/discordant, pleasing/unpleasing, pleasant/unpleasant, and not-annoying/annoying

Factor II: shrill/calm, clear/vague, metallic/deep, and tonal/not-tonal

Factor III: powerful/weak, loud/soft, and noisy/quiet

In considering the semantic quality of these adjectives, the factors were named: pleasant (I), metallic (II), and power (III). The same structure has been reported in other semantic differential studies from Japan (e.g., Namba *et al.*, 1992), which is why the terminology is adopted here.

The similarities and differences between the rotated factor loadings were quantified by computing congruency measures such as Tucker's ϕ and the correlation coefficients between the three matrices. These are relatively high: $\phi \geqslant 0.91$ and $r \geqslant 0.85$. The solutions for the Japanese and French samples concerning the pleasant factor loadings are almost proportional to each other ($\phi=0.98$). This also holds for the comparison between these loadings and the loadings for the German sample's pleasant factor ($\phi=0.94$). Again, the same holds for metallic factor loadings ($\phi=0.93$). The French power factor is less noteworthy ($\phi \leqslant 0.93$) than the Japanese and German power factors. As reported in Sec. II A 1, $\phi=0.93$ is a conservative estimation for the significance level. At this value, the power factor loadings can still be considered equivalent.

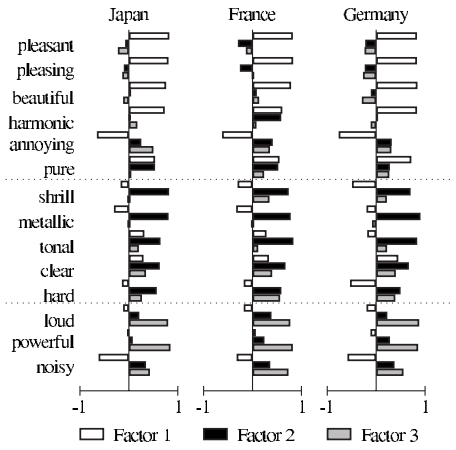


FIG. 2. Factor loadings of the Japanese, French, and German samples: Adjective items used in the semantic differential with a seven-step rating scale. The Japanese loadings were rotated with the Varimax rotation. This solution provides the target for the rotation of the French and German sample's factor space.

In commenting two major differences, the adjective pair harmonic/discordant in the French sample does not only load highly on the pleasant factor but, contrary to the other data, it is more closely related to the description of timbre, i.e., the metallic factor (II). Although the translation of this item is considered adequate by scientists with French as their native language, the different use of this item indicates bias. Second, the French participants associate the adjective pair noisy/quiet mostly with the powerful factor (III), while in the other groups the pair shows medium loadings on pleasant factor (I) and the powerful factor (III), indicating an evaluative connotation. Moreover, the concept of noisiness combines power and unpleasantness in this case. This may also be due to item bias, meaning that the correctly translated adjective pair is used differently in the French language. The differences measured on the adjective pairs hard/soft and pure/impure were not evaluated due to low communalities in all three groups.

As a major result, the equivalence of all three factors, pleasant, metallic and power, was established, even though there are some differences on certain items. This conceptual or construct equivalence is a necessary condition for the following item bias analysis.

B. Item bias analysis

After exploring the structure of the factorial space, further analyses were conducted to investigate the adjective pairs in more detail. After summarizing the results from the

discriminatory power analysis and the internal consistency analysis, i.e., traditional psychometric analyses, the item bias analysis will be reported, followed by a direct comparison of each mean item score between the samples.

As argued in Sec. II A 2, the factor-item relation is qualified by computing a discriminatory power analysis and an internal consistency analysis. The former is conducted to inspect how well single adjective pairs represent their associated factor. All characteristic items of the pleasant, metallic, and power factors were tested. Additionally, the items quiet/noisy and pure/impure are included in the analysis for the pleasant factor because they show high loadings in the German samples and medium loadings in the other samples.

The discriminatory power analysis yielded at least a medium ($0.5>r>0.3$) but mostly a high ($r\geqslant 0.5$) item-to-dimension correlation coefficient. To designate whether an adjective pair represents a factor for all cultures, only items with a high mean item-to-dimension correlation in all cultures and an individual correlation coefficient of at least medium size were selected. Most items are considered to be representative items for their associated factors because they show high mean discriminatory power. However, the following items had to be excluded: Noisy/quiet did not yield medium discriminatory power in the French sample, while pure/impure showed only medium mean discriminatory power. It turns out that the 12 representative pairs are identical to the 12 marker items identified in Sec. III A.

The internal consistency (Cronbach's α) is used as a measure to display how consistently the marker items can be aggregated to one scale. This is a necessary procedure for an item bias analysis. All adjective items yielded high internal consistency with their confounding factors pleasant, metallic, or power.

As the factor-item relation is within the set limits, the bias was assessed on item level (single adjective pairs). To provide indicators for equal standing (see Sec. II A), total factor score levels were calculated by summing the scores of all established marker items of each factor. Therefore, for each participant's judgment of a stimulus and for each factor, a total factor score was generated. These total factor scores for each factor were combined in a set comprising the contributions from all three cultures. This set of total factor score levels was subdivided into roughly equally large groups of *score levels*. These roughly equally sized groups of the combined distribution of total factor scores representing a factor were used to derive ten score levels each for the pleasant and metallic factors and five score levels for the power factor. Judgments aggregated on the same score level were assumed to have equal standing regarding the respective sound on the represented factor.[6]

MANOVAs were calculated consecutively for each factor. Score level and culture are the independent variables; the dependent variables; are the respective items belonging to a factor ($p=0.05$). To quantify the effect size, η^2_{partial} was calculated.[7] Table IV shows the results: culture as a main effect (first column) and the interaction effect between culture and score level (second column). These effects can be described as differences between means for the three cultures as measured on a seven-step rating scale. In other words, how large is the bias on significantly biased pairs? Adjective pairs showing bias with an effect size $\eta^2_{\text{partial}}\leqslant 0.6$ have difference between means smaller than 0.5 rating scale divisions between the cultural subsamples. Examining the item clear/vague in more detail, the Japanese sample shows uniform bias with regard to the French and German samples. At score levels >2, the French participants used the item about 1 scale division, and the German participants about 0.5 scale divisions lower than the Japanese participants. "Small" interaction is observed in the German sample. With increasing metallic perception, the German participants cease to describe the stimulus as vague and thus more inline with the Japanese sample, i.e., a nonuniform bias is observed here.

TABLE IV. Effect sizes for the significant ($p=0.05$) culture and of culture x score level effects on the item scales constituting the dimensions of perception. The main effect indicates uniform bias, the interaction nonuniform bias. (Please refer to Sec. III B for a score level definition. No significant changes are indicated by "—." No/weak bias is marked with a ⋆.)

		η^2_{partial}	
Dimension		Culture[a]	Culture× score level
	Pleasant		
pleasant/unpleasant	⋆	0.02	0.02
harmonic/discordant		0.05	0.02
pleasing/unpleasing	⋆	0.02	0.02
beautiful/ugly	⋆	0.01	0.03
not annoying/annoying		0.05	0.04
	Metallic		
shrill/calm	⋆		
clear/vague		0.07	0.04
metallic/deep	⋆	0.01	0.02
tonal/not tonal		0.01	0.05
	Power		
powerful/weak	⋆	0.01	0.04
noisy/quiet	⋆	0.01	0.02
loud/soft	⋆	0.01	0.03

[a]The respective interactions are *ordinal*; the main factor "culture" can therefore be interpreted.

In general, it can be concluded that there are several small biases on the adjective items used. These item scales must be scrutinized while interpreting the results at face value. Nevertheless, several items constituting the pleasant, metallic, and power factors seem to be bias-free, or they show a diminutive bias level and could be used for further level-oriented analysis, e.g., shrill/calm or pleasant/unpleasant (see Table IV, ⋆-marked).

C. Differences in means: The impact of culture

Apart from investigating the reliability of the semantic differential across cultures, a major goal is to investigate the potential effect of culture, i.e., the impact on the respective adjective items as described in Sec. II A. Do cultural groups differ in the kinds of judgments they make? Furthermore, do noise stimuli have different effects depending on the cultural group perceiving them?

In order to examine these questions, MANOVAs were calculated with tonal component frequency content f, tonal

TABLE V. Estimated effect size η^2 for the significant ($p=0.05$) main effect culture (cul) and the significant interaction culture x stimulus variation ($f \times$cul, $\overline{\Delta L} \times$cul) on item scales constituting the perceptual factors. (The largest effect sizes are in bold. No significant changes are indicated by"—.")

Adjective scales	cul	$f \times$cul	$\overline{\Delta L} \times$cul
pleasant/unpleasant	0.02	—	0.03
harmonic/discordant	0.02	—	**0.10**
pleasing/unpleasing	0.03	—	0.04
beautiful/ugly	0.01	—	0.05
annoying/not-annoying	0.06	—	0.03
shrill/calm	0.02	—	0.02
clear/vague	**0.06**	—	0.04
metallic/deep	0.01	—	0.02
tonal/not-tonal	0.04	—	**0.10**
powerful/weak	—	—	0.04
noisy/quiet	0.04	—	—
loud/soft	0.05	—	0.03

energy $\overline{\Delta L}$, and culture as independent factors. The S/N ratio increased in six steps. Stimulus S_1 was excluded because it lacks a tonal component. This analysis yielded *inter alia* the effects of culture and the interaction between culture and the other independent variables. At this stage, bias-free items must be assumed.

As a result, the tonal frequencies f have an influence on the metallic factor adjectives, which leads to increasing metallic factor scores with growing frequency. $\overline{\Delta L}$ correlates with increasing unpleasant, metallic, and power factor scores. Table V shows the effect of culture and the interactions between culture and the systematic stimulus variations. The items harmonic/discordant, clear/vague, and tonal/not-tonal have the largest effect sizes. The effects must be analyzed in relation to the findings in Sec. III B, i.e., are the reported changes in means of the same magnitude as the biases?

Using the example in Fig. 3, mean values for the adjective pair clear/vague are plotted against $\overline{\Delta L}$ of stimuli containing two frequencies for each cultural subsample at face value. The mean ratings of $\overline{\Delta L}=15$ dB and lower must be examined because they have significant differences. Following the concept in Sec. II A 3, the differences in means cannot be analyzed at face value. The differences in Fig. 3 can be explained due to the bias explicitly described in Sec. III B. Examining the other main and interaction effects in the light of specific bias, the effects are reduced to a minimum yielding no valid significant differences in means across the cultures investigated.

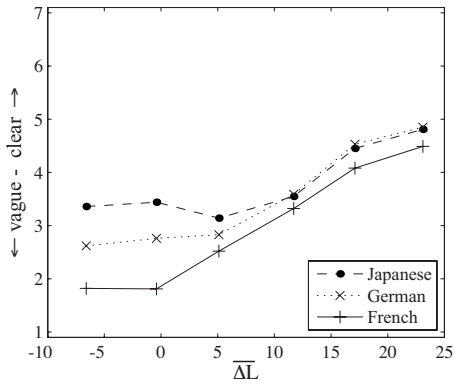


FIG. 3. Clear/vague item means for the judgments by the Japanese, French, and German subsamples vs the $\overline{\Delta L}$. $\overline{\Delta L}$ is the mean value of $\Delta L_{T,\mathrm{sum}}/\Delta L$ of the six steps of S/N-ratio increase. The obtained significant difference at lower values of $\overline{\Delta L}$ is mitigated by the fact that the scale is biased. (The standard deviation, $\sigma \approx 1$, was omitted for clarity.)

D. Acoustical sound descriptors in a cross-cultural context

Three objective parametric sound descriptors were calculated to objectively characterize the variance of the stimuli: a level parameter $L_{Aeq,6\ s}$, a parameter describing the tonal content ΔL_T (DIN 45681:2005-03, 2005), and the sharpness S (von Bismarck, 1984; Zwicker and Fastl, 1999). The background noise (stimulus S_1 in Table I) has the lowest level: 44.5 dBA. For the other stimuli, adding tonal energy increases the level $L_{Aeq,6s}$, and the difference between the lowest and the highest level (53.3 dBA) amounts to 8.7 dBA. Therefore, many of these stimuli have clearly distinguishable loudness differences. The parameter changing the most among the stimuli is the S/N ratio indicator ΔL_T, which was calculated according to DIN 45681:2005-03 (2005). This parameter ranges from $-\infty$ to 24.5 dB. With 1.25 acum, stimulus S_1, having no tonal content embedded in the noise, has the highest sharpness S. This sharpness is equivalent to the sharpness of a third octave band centered around 1.25 kHz with a level of 60 dBA. Apparently, adding tones with frequencies lower than 1.25 kHz, i.e., all tonal components in this study, to the background noise decreases the perceived sharpness of stimuli containing tonal energy when compared to stimulus S_1: The more the low frequency tonal content, the lower the sharpness. The addition of tones to the noise affects all three signal parameters, which is represented by high correlation coefficients r between them: $r(\Delta L_T, L_{Aeq,6\ s})=0.88$, $r(S, \Delta L_T)=-0.95$, and $r(S, L_{Aeq,6\ s})=-0.98$. All of these correlations are significant ($p=0.05$).

The relation between the objective sound descriptors and the perceptual dimensions is described by correlation coefficients. The marker item score sums for a particular factor dimension (see Sec. III B) were used as scale values of the corresponding stimuli to calculate. The results will be discussed separately for the three dimensions. In the following, all reported correlations are significant on a $p=0.05$ level.

(a) *Power*. A level parameter was expected to correlate the most highly with this factor. Yet due to the high correlation between all signal parameters, the $L_{Aeq,6\ s}$ is not better than the other parameters, all showing rather small coefficients. The correlation coefficients for the Japanese sample are significantly lower ($r=0.24$) than the other samples ($r \approx 0.5$). This difference is a result of the nonuniform bias on the items constituting this factor. As mentioned in Sec. III B, the nonuniform bias relates to the scale usage of the Japanese sample, i.e., it shows a smaller overall variance. Thus, the sample shows a smaller correlation. For an integral view, Fig. 4 shows the relationship between the $L_{Aeq,6\ s}$ and the average of all subjective scale values for the power factor.

(b) *Metallic*. Correlations with the objective signal param-

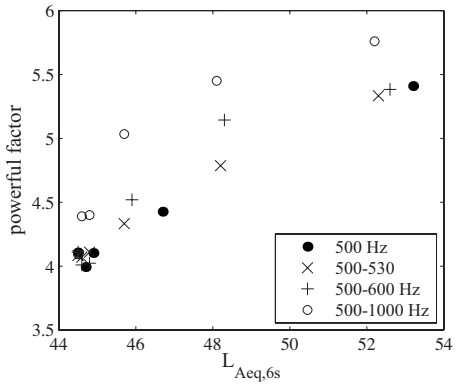


FIG. 4. $L_{Aeq,6\ s}$ is plotted against the power factor, i.e., the mean of the scale values for the underlying adjective items. (The standard deviation, $\sigma \approx 1$, is omitted for clarity.)

eters are highest for the metallic dimension ($r \approx 0.8$, Japanese sample $r \approx 0.6$). This factor is best represented by the parameter ΔL_T as the correlations with the level parameter $L_{Aeq,6\ s}$ and the calculated sharpness are lower. The parameter ΔL_T characterizes the relation between the tonal energy of a stimulus and its noise. The metallic dimension best represents how timbre fluctuates among the stimuli. Again, the difference between the cultures is a result of the lower overall variance in the Japanese sample, which leads to the nonuniform bias on the underlying scale (see Table IV).

Figure 5 illustrates a broader perspective on the relationship between the parameter ΔL_T and the metallic dimension. The average scale values for the three subsamples are reported there. A constant "metallic" perception was observed for ΔL_T below threshold, i.e., <0. No significant change in the sound character is expected for those values. The relationship becomes linear for higher values. These results are similar to the findings reported by Namba *et al.* (1992, $\Delta L_T \triangleq S/N$).

For our data, the correlation coefficient between the metallic or timbre factor and the calculated sharpness is negative in all cultural samples. This is because a more pronounced tonal character corresponds to a decrease in sharpness. Apparently, it is in line with the intercorrelations between the signal parameters.

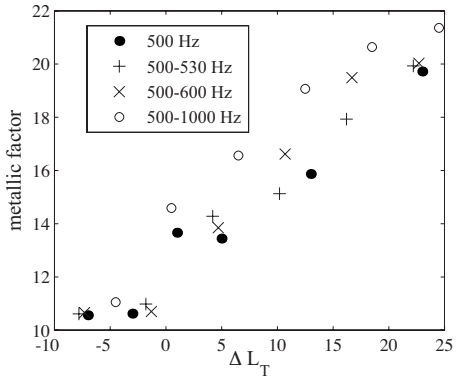


FIG. 5. Tonal energy ΔL_T being either ΔL or $\Delta L_{T,\mathrm{sum}}$ is plotted against the metallic factor, i.e., the mean scale values for the underlying adjective items. (The standard deviation, $\sigma \approx 0.8$, is omitted for clarity.)

(c) *Pleasant*. The correlation between the objective parameters and this factor is the lowest. Even insignificant correlations were found for the French participants, and the highest correlations were found for the German sample ($r \approx 0.4$), while the Japanese sample shows only a weak correlation ($r \approx 0.2$). Due to the fact that the factors are perpendicular to each other as well as to the high correlation between the objective parameters, the high correlation found for the metallic factor inevitably results in lower correlations for the other two subjective dimensions. Moreover, this pleasant factor can be interpreted as an evaluation factor. It varies noticeably among participants when they judge the same stimuli.

The parameter ΔL_T, i.e., the sum of all tonal energy above masked threshold, was not the only parameter calculated. Different summation methods for the tonal components were compared, such as taking only the maximal component, adding up only the frequencies lying in the proximity limit Δf_{prox} [ISO 7779:1999(E), 1994] or within a common critical band (DIN 45681:2005–03, 2005). The different approaches did not lead to significant differences in correlations for any case ($\Delta r_{\mathrm{crit}} \approx 0.13$). The parameter ΔL_T was prone to correlate most highly with the pleasant factor.

IV. DISCUSSION

A. Cross-cultural similarities and differences

According to the results presented in Sec. III A, the perceptual and connotative factors assessed for each culture are similar, i.e., they feature the same three-factor structure. These factors are labeled pleasant, metallic/timbre, and power after results reported by Namba *et al.* (1992). The invariance of this factorial solution among the Japanese, French, and German samples is comparable to the consistent structure reported by Osgood *et al.* (1975) regarding the affective meaning of concepts, which was tested in several cultures. In their study, concepts were presented without context to ensure that their meaning would not be associated to anything else, and the authors reported a structure they titled EPA. The factorial solution found in this study resembles the EPA structure: The evaluation factor corresponds to the pleasantness factor, representing the evaluation of sound, while the power factor coincides with the power factor. It is argued that the metallic/timbre factor represents the activity dimension as the latter is connected to the quality in the sense of modal features. Heise (1969) argued that this structure does not always have to be obtained as it depends on the stimuli investigated. Within certain sound corpora this structure might not be fully evolved as certain dimensions are fused. On the other hand, there may be more dimensions, e.g., Solomon (1958), if, for example, the sound corpus varies in many aspects of timbre. An example of fused dimension would be if the annoyance is highly coupled with certain characteristics of the sound, such as sharpness or roughness (Zwicker and Fastl, 1999). In this example, the pleasant and metallic/timbre factor would coincide.

Nevertheless, some items—noisy/quiet, harmonic/discordant, clear/vague—show different affective meanings,

although the denotative meaning was correctly translated. The Japanese and German samples used the item noisy/quiet in a similar way: Instead of employing it only in a nonevaluative manner to describe power, it is also used along with items loading on the pleasant factor. Concerning the concept of "noise," this is in line with Kuwano *et al.* (1986), who reported a high correlation between the concepts of noise found in Japan and Germany. Another difference was found with harmonic/discordant. Almost solely an indicator of pleasantness in the Japanese and German samples, it refers to the pleasant and timbre factors in the French sample, perhaps eliciting the musical connotation of harmony. By revealing culturally invariant dimensions, item bias is revealed as well. The item bias analysis, e.g., for clear/vague, showed that cross-cultural differences in means for this item correspond to different ways it is used.

The semantic differential used in this study assessed the affective meaning of sounds in a context-free setting. However, the broad similarity in judgments does not imply that potential reactions or annoyance are equal in every culture in a given context. Namba *et al.* (1991b), for instance, investigated how people deal with neighborhood noise problems by conducting noise evaluations in different cultures. The differences in the countermeasures against noise showed that the Japanese sample applied a more defensive approach, i.e., a high percentage responded that they could get used to noise depending on the situation. Hence, in a given context, these results imply differences in the interaction between noise and the context, which influences the formation of affective meanings. For stimuli in a context-free environment, this study shows that the initial judgment of sounds is very similar cross-culturally.

B. Bias analysis: Considering its relevance

This study employed the bias analysis described by van de Vijver and Leung (1997). It provides the possibility to determine the level of equivalence of the construct and the scales. Therefore, the bias analysis serves as more than a mere guideline for critically assessing every step of the experimental procedure regarding the exclusion of method bias: It yields a quantitative analysis to prevent overestimating any face-value difference. This study exemplifies the latter in two ways. At first, contrary to investigations of perceptual dimensions by Schick and Hoege (1996) as well as Iwamiya and Zhan (1997), the factor rotation reduces the differences between the factor solutions obtained in different cultural subsamples and after this, congruency is estimated quantitatively. Therefore, the bias analysis can draw a clearer picture of the actual similarities and by doing so, the differences in scale usage (e.g., harmonic/discordant in the French sample) are outlined precisely. Second, the item bias analysis yielded the prerequisite to analyze the face-value differences. The study showed that the cultural differences relate to the heterogeneous scale usage. Thus, the analysis offers a valuable test procedure indicating to what degree scales are biased. The semantic differential technique, i.e., using multiple scales which can be very close in meaning (e.g., pleasant and pleasing), seems to be a necessary precaution in order to explore the full affective meaning of sounds, although differences in usage are not guaranteed in the end. If semantic differentials, which are believed to cover the essential descriptions, are condensed a priori, it is not possible to obtain the perceptual and connotative factors which the item bias analysis is grounded upon. Hence, the cross-cultural application of single categorical scales, no matter how well translated to reflect their denotative meanings, should be avoided as this procedure is prone to bias errors.

C. Acoustical sound descriptors

Acoustical sound descriptors were correlated with the perceptual and connotative factors obtained by the factor analyses. The summation of the tonal energy above masked threshold explains the metallic/timbre factors. Therefore, the description of the tonal content can be realized by the tonal energy if the noise contains distinct tonal features (Hansen and Weber, 2008). However, the tonal energy is summed no matter how the frequencies relate to each other, i.e., whether they are located within one critical band or even within the proximity limit. This suggests that the level above threshold could be the parameter to describe the tonal content. This idea follows an approach by Vormann *et al.* (2000), in which a tone in noise adjusted to the same tonal energy of the stimulus was used as an adaptive subjective measure. Furthermore, concerning the items describing the change in tonal timbre, differences, which could not be attributed to differential item usage, were not obtained between the cultural samples.

Namba *et al.* (1992) showed that the power factor can be modeled by loudness. Here, too, the power dimension can be attributed to the overall loudness, but it is certainly influenced by the presence of the tonal components. The covariation between the $L_{Aeg,6\ s}$ and the parameter describing the tonal content makes a distinction impossible.

The pleasant/annoyance factor corresponds only very weakly to any of the sound descriptors, as the items on the factor show a relatively small total variance over all stimuli. In the French sample, the dimension is basically independent from the S/N ratio, contrary to, for example, Hellman (1982). They reported that annoyance increases with an increasing S/N ratio. Oppositely, Zimmer *et al.* (2004) reported in an indirect scaling experiment examining the auditory unpleasantness of environmental sounds using two predictors: sharpness and roughness. These predictors are valid in groups of similar loudness, and the latter serves as the greatest predictor between groups of different stimuli. These results are in line with a concept proposed by Berglund *et al.* (1994), in which perceived annoyance is based on three factors: loudness-based annoyance, quality-based intrusiveness (due to roughness and sharpness), and the distortion of information. In the present study, overall pleasantness does not vary much in the sense of auditory unpleasantness (Zimmer *et al.*, 2004), as the underlying parameters roughness/sharpness and loudness do not vary much. Therefore, in a context-free environment the annoyance ratings, which can only be based on the acoustical stimuli, should not vary to a great extent

either. If the tonal content had been directly related to auditory unpleasantness, the pleasant and metallic/timbre factor would have fused, which is not the case.

This result does, however, not argue for the independence of annoyance in the sense described by Berglund *et al.* (1994). Tonal components play a major role in sound identification. Enabling the formation of new emergent meanings, e.g., the sound of a lawn mower, is not annoying *per se* but becomes annoying if, for example, the neighbors are using it in the early morning. The current study has shown that there is little difference in the evaluation of tonal sounds within a context-free setting, but adding a context and therewith an emergent meaning might lead to evaluation differences across the investigated cultures. The sound of bells, a typical tonal sound, may illustrate this. Associated with a church in Germany, it is used only as a warning signal in Japan.

V. SUMMARY AND CONCLUSIONS

In this cross-cultural study the perception of noise with tonal components was examined in Japan, France, and Germany with the aid of a semantic differential. In order to uncover valid cross-cultural differences, a paradigm from cross-cultural psychology (Van de Vijver and Leung, 1997; Van de Vijver, 1998) was introduced and applied to the semantic differential data. Differences in cross-cultural data are not necessarily valid differences as they may be due to different forms of bias that jeopardize the equivalence of the data to be compared. Van de Vijver and Leung (1997) claimed that valid cross-cultural differences can be identified if and only if the different levels of equivalence for the data to be compared had been proven. This was equivalent to the request that the corresponding types of bias had been ruled out after a bias analysis.

The first goal was the evaluation of the semantic differentials in a cross-cultural context. It was necessary to apply a factor rotation in order to correctly assess the difference between the connotative and perceptual factors. After rotating the factors onto each other, almost congruent factors were obtained. These common factors were labeled: pleasant, metallic/timbre, and power. They correspond to three culturally invariant factors describing the affective meaning of concepts: evaluation, activity, and power (Osgood *et al.*, 1975). Small differences were observed regarding how some adjectives are situated within this three-dimensional space, most notably for the items harmonic/discordant and noisy/quiet. In part, this is attributed to different concepts of noisiness and loudness (Namba *et al.*, 1991a).

Furthermore, the items were examined on the basis of equal perceptual dimensions obtained. It was found that most of the face-value differences were due to bias on the respective adjective items. In conclusion, it is recommended to avoid employing single categorical scales when evaluating certain aspects of sound character, e.g., tonality. No matter how well the denotative meaning is established, a bias evaluation of single items is not possible. Hence, a source of systematic errors in a cross-cultural context would remain.

The second goal, the validation of psychoacoustic parameters regarding tonal contents in a cross-cultural context, leads to the conclusion that cultural bias is responsible for the observed differences in correlations between the metallic/timbre factor and the objective parameters for comparisons between both European subsamples with the Japanese subsample. The S/N-based parameters showed a fairly high correlation with the metallic factor. Thus, they are able to describe this important aspect of sound character. A difference was obtained on the pleasant factor. For the French sample there is no correlation between the pleasant factor and objective parameters investigated, while it is weak in the case of the Japanese and German samples. These results suggest that the tonal content might not have affected auditory unpleasantness (Zimmer *et al.*, 2004). Nevertheless, as the tonal content often confers source information to the recipient of noise, it is prone to have an important role in the context of transient affective meaning, i.e., the change of meaning due to a specific context.

ACKNOWLEDGMENTS

We would like to thank Professor Kuwano and Professor Namba as well as the work group at Osaka University for their helpful and encouraging discussions, their continuous support, and for the opportunity to use their testing facilities. Furthermore, we would like to thank Professor Parizet from INSA, Lyon for the invitation to conduct experiments at his facilities as well as for the wonderful support we received from him and the entire team at the LVA. We are especially grateful to all research partners for the valuable and indispensable help we received during the translation process of the semantic differential into the mother tongues Japanese and French. This contributed so much to the success of the investigations. Cara H. Kahl from the University of Hamburg, Germany deserves special thanks for her comments on earlier versions of this paper regarding its contents and style. At last, we want to thank the three anonymous reviewers for their valuable and helpful comments on an earlier version of this paper.

[1]Moreover, there are studies concerning musical pitch and how intervals, scales, and tuning are perceived in different cultures. Although they offer valuable insight, studies on music cognition are only marginally comparable to the issues in this study because artificial and stationary stimuli were investigated. For an overview of cross-cultural studies on musical pitch, see Stevens (2004).

[2]The authors are aware that "culture" and "country" are not necessarily the same as Hofstede (1991) pointed out. He admitted, however, that pragmatic reasons often enforce such an operationalization.

[3]"Bias refers to the presence of nuisance factors in cross-cultural research. Three types of bias are distinguished, depending on whether the nuisance factor is located at the level of the construct (construct bias), the measurement instrument as a whole (method bias), or the items (item bias or differential item functioning). Equivalence refers to the measurement level characteristics that apply to cross-cultural score comparisons; three types of equivalence are defined: construct (identity of constructs across cultures), measurement unit (identity of measurement unit), and scalar equivalence (identity of measurement unit and scale origin). Bias often jeopardizes equivalence" (van de Vijver, 1988, p. 41)

[4]Tucker's ϕ: $\phi_{xy}=\Sigma_i x_i y_i / \sqrt{\Sigma_i x_i^2 \Sigma_i y_i^2}$.

[5]The sharpness S is defined as $S=0.11\times\int_0^{24\text{bark}} N' g(z) z dz / \int_0^{24\text{bark}} N' g(z) dz$acum. The denominator is the total loudness N, while N' is the specific loudness of a critical band (von Bismarck, 1984). The weighting function $g(z)$ is 1 and increases with $z>16$ bark. Therefore, the sharp-

ness is the first moment of the weighted critical-band rate distribution of the specific loudness (Zwicker and Fastl, 1999).

[6]The MANOVA is robust against different score level numbers, yielding a similar result to those presented in Table IV, as reported by van de Vijver and Leung (1997). Furthermore, only five score levels were used for the power factor because a roughly equal number of scores on each level is mandatory.

[7]$\eta^2_{\text{partial}} = SS_i / SS_i + SS_{\text{error}}$, $SS_i = \Sigma_m (x_{im} - \bar{x}_i)^2$); i.e., the ratio of the variation induced by a main or interaction effect i (SS_i) to this variation (SS_i) plus the variation left to error (SS_{error}; Cohen, 1973).

Berglund, B., Harder, K., and Preis, A. (**1994**). "Annoyance perception of sound and information extraction," J. Acoust. Soc. Am. **95**, 1501–1509.

Berry, J., Poortinga, Y., Segall, M. H., and Dasen, P. R. (**2002**). *Cross-Cultural Psychology, Research and Applications*, 2nd ed. (Cambridge University Press, Cambridge, UK).

Blauert, J., and Jekosch, U. (**1997**). "Sound-quality evaluation—A multi-layered problem," Acust. Acta Acust. **83**, 747–753.

Bortz, J., and Doering, N. (**2002**). *Forschungsmethoden und Evaluation (Methods and Evaluation)*, 3rd ed. (Springer, Berlin).

Cohen, J. (**1973**). "Eta-squared and partial eta-squared in fixed factor anova designs," Educ. Psychol. Meas. **33**, 107–112.

Cureton, E. E., and D'Agostino, R. B. (**1983**). *Factor Analysis: An Applied Approach* (Lawrence Erlbaum Associates, Hillsdale, NJ).

DIN 45681:2005–03 (**2005**). "Acoustics—Detection of tonal components of noise and determination of tone adjustment for the assessment of noise immisions" (Beuth, Berlin, Germany), German industry standard.

Guski, R. (**1997**). "Psychological methods for evaluating sound quality and assessing acoustic information," Acust. Acta Acust. **83**, 765–774.

Hansen, H., and Weber, R. (**2008**). "The influence of tone length and S/N-ratio on the perception of tonal content," Acoust. Sci. & Tech. **29**, 156–166.

Heise, D. R. (**1969**). "Some methodological issues in semantic differential research," Psychol. Bull. **72**, 406–422.

Hellman, R. P. (**1982**). "Loudness, annoyance, and noisiness produced by single-tone-noise complexes," J. Acoust. Soc. Am. **72**, 62–73.

Hofstede, G. (**1991**). "Empirical models of cultural differences," in *Contemporary Issues in Cross-Cultural Psychology*, edited by N. Bleichrodt, and P. Drenth (Swets & Zeitlinger, Amsterdam), pp. 4–20.

Hussain, M., Pflüger, M., Brandl, F., and Biermayer, W. (**1998**). "Intercultural differences in annoyance response to vehicle interior noise," Proceedings of the Euronoise 98, Munich, Germany, pp. 521–526.

ISO 7779:1999(E) (**1999**). "Acoustics–Measurement of airborne noise emitted by information technology and telecommunications equipment," ISO, Genève, Switzerland, international standard.

Iwamiya, S., and Zhan, M. (**1997**). "A comparison between Japanese and Chinese adjectives which express auditory impressions," J. Acoust. Soc. Jpn. (E) **18**, 319–323.

Korth, B. (**1978**). "A significance test for congruence coefficients for Cattell's factors matched by scanning," Multivar. Behav. Res. **13**, 419–430.

Korth, B., and Tucker, L. (**1975**). "The distribution of chance congruence coefficients from simulated data," Psychometrika **40**, 361–372.

Kuwano, S., Fastl, H., Namba, S., Nakamura, S., and Uchida, H. (**2006**). "Quality of door sounds of passenger cars," Acoust. Sci. & Tech. **27**, 309–312.

Kuwano, S., Namba, S., Kurakata, K., and Kikuchi, Y. (**1994**). "Evaluation of broad-band noise mixed with amplitude-modulated sounds," J. Acoust. Soc. Jpn. (E) **15**, 131–142.

Kuwano, S., Namba, S., and Schick, A. (**1986**). "A cross-cultural study on noise problems," *Contributions to Psychological Acoustics* (BIS, Oldenburg, Germany).

Kuwano, S., Namba, S., Schick, A., Hoege, H., Fastl, H., Filippou, T., and Florentine, M. (**2007**). "Subjective impression of auditory danger signals in different countries," Acoust. Sci. & Tech. **28**, 360–362.

Marquis-Favre, C., Premat, E., Aubrée, D., and Vallet, M. (**2005**). "Noise and its effects—A review on qualitative aspects of sound. Part I: Notions and acoustic ratings," Acta. Acust. Acust. **91**, 613–625.

Namba, S., and Kuwano, S. (**1984**). "Psychological study on L_{eq} as a measure of loudness of various kinds of noises," J. Acoust. Soc. Jpn. (E) **5**, 135–148.

Namba, S., Kuwano, S., Hashimoto, T., Berglund, B., Rui, Z., Schick, A., Hoege, H., and Florentine, M. (**1991a**). "Verbal expression of emotional impression of sound: A cross-cultural study," J. Acoust. Soc. Jpn. (E) **12**, 12–29.

Namba, S., Kuwano, S., Kinoshita, K., and Kurakata, K. (**1992**). "Loudness and timbre of broad-band noise mixed with frequency-modulated sounds," J. Acoust. Soc. Jpn. (E) **13**, 49–58.

Namba, S., Kuwano, S., Schick, A., Aclar, A., Florentine, M., and Rui, Z. (**1991b**). "A cross-cultural study on noise problems: Comparison of the results obtained in Japan, West Germany, the U.S.A., China and Turkey," J. Sound Vib. **151**, 471–477.

Osgood, C. E., May, W. H., and Miron, M. S. (**1975**). *Cross-Cultural Universals of Affective Meaning* (University of Illinois Press, Champaign IL).

Osgood, C. E., Suci, G. J., and Tannenbaum, P. H. (**1957**). *The Measurement of Meaning*, new ed. (University of Illinois Press, Champaign IL).

Poortinga, Y. (**1989**). "Equivalence of cross-cultural data: An overview of basic issues," Int. J. Psychol. **24**, 737–756.

Schick, A., and Hoege, H. (**1996**). "Cross-cultural psychoacoustics," *Recent Trends in Hearing Research, Festschrift for Seiichiro Namba* (BIS, Oldenburg, Germany), Chap. 12, pp. 287–315.

Solomon, L. (**1958**). "Semantic approach to the perception of complex sounds," J. Acoust. Soc. Am. **30**, 421–425.

Stevens, C. (**2004**). "Cross-cultural studies of musical pitch and time," Acoust. Sci. & Tech. **25**, 433–438.

van de Vijver, F. (**1998**). "Towards a theory of bias and equivalence," in *Cross-Cultural Survey Equivalence*, edited by J. Harkness (ZUMA, Mannheim, Germany), pp. 41–65.

van deVijver, F., and Leung, K. (**1997**). *Methods and Data Analysis for Cross-Cultural Research* (Sage, Thousand Oaks, CA).

von Bismarck, G. (**1984**). "Sharpness as an attribute of timbre of steady sounds," Acta Acust. **30**, 159–172.

Vormann, M., Verhey, J. L., Mellert, V., and Schick, A. (**2000**). "Subjective ratings of tonal components in noise with an adaptive procedure," in *Contributions to Psychological Acoustics*, edited by A. Schick, M. Meis, and C. Reckhardt (BIS, Oldenburg, Germany), pp. 145–153.

Watkins, D. (**1989**). "The role of confirmatory factor analysis on cross-cultural research," Int. J. Psychol. **24**, 685–701.

Zimmer, K., Ellermeier, W., and Schmid, C. (**2004**). "Using probabilistic choice models to investigate auditory unpleasantness," Acta. Acust. Acust. **90**, 1019–1028.

Zwicker, E., and Fastl, H. (**1999**). *Psychoacoustics*, 2nd ed. (Springer, Berlin).

Zwicker, E., and Feldtkeller, R. (**1967**). *Das Ohr als Nachrichtenempfänger (The Ear as a Communication Receiver)*, 2nd ed. (Hirzel, Stuttgart, Germany).

Development of an analytical solution of modified Biot's equations for the optimization of lightweight acoustic protection

Jamil Kanfoud[a] and Mohamed Ali Hamdi
Laboratoire Roberval, Université de Technologie de Compiegne, Rue Personne de Roberval, Centre de Recherche CRD338, B.P. 20529, Compiegne Cedex 60205, France

François-Xavier Becot and Luc Jaouen
MATELYS - Acoustique and Vibrations, 20/24 rue Robert Desnos, F-69120 Vaulx-en-Velin, France

(Received 28 January 2008; revised 15 June 2008; accepted 21 July 2008)

During lift-off, space launchers are submitted to high-level of acoustic loads, which may damage sensitive equipments. A special acoustic absorber has been previously integrated inside the fairing of space launchers to protect the payload. A new research project has been launched to develop a low cost fairing acoustic protection system using optimized layers of porous materials covered by a thin layer of fabric. An analytical model is used for the analysis of acoustic wave propagation within the multilayer porous media. Results have been validated by impedance tube measurements. A parametric study has been conducted to determine optimal mechanical and acoustical properties of the acoustic protection under dimensional thickness constraints. The effect of the mounting conditions has been studied. Results reveal the importance of the lateral constraints on the absorption coefficient particularly in the low frequency range. A transmission study has been carried out, where the fairing structure has been simulated by a limp mass layer. The transmission loss and noise reduction factors have been computed using Biot's theory and the local acoustic impedance approximation to represent the porous layer effect. Comparisons between the two models show the frequency domains for which the local impedance model is valid.
 [DOI: 10.1121/1.2973197]

PACS number(s): 43.50.Gf, 43.20.Gp, 43.20.Hq, 43.20.El [KA] Pages: 863–872

I. INTRODUCTION

The purpose of this paper is to develop an analytical model permitting the solution of Biot's equations governing wave propagation for planar elastic porous layers. The study aims to design an optimized acoustic protection having a high acoustic absorption coefficient (AC) in the very low frequency band (the 63 Hz third octave band). As described in Refs. 1–5, original Biot's equations have been written in terms of solid and fluid displacements. In Ref. 6, Bonnet derived basic singular solutions of Biot's equations. By eliminating the fluid displacement, he established a system of four differential equations coupling the skeleton displacement and the acoustic pressure in the interstitial fluid. For finite element implementation purposes, a mixed displacement pressure integral formulation has been derived by Atalla *et al.*[7] Boundary conditions associated with the weak mixed displacement pressure formulation are discussed in Ref. 8. Previous papers assume that Biot's parameters are spatially constant. A modified system of four Biot's equations, valid for nonspatially constant Biot's parameters, was established by Hamdi *et al.* in Ref. 9. As described in Refs. 9 and 10, the weak formulation associated with modified Biot's equations has the great advantage of leading to natural boundary conditions at interfaces between adjacent layers.

To clarify the approach, the first section of the paper recalls the system of modified Biot's equations established in Ref. 9 in terms of the skeleton displacement vector and of the pressure inside the interstitial acoustic medium saturating the pores.

The second section derives analytical solutions of the modified Biot's equations propagating in planar and laterally infinite porous layers of finite thickness. Boundary conditions associated with the system of four differential equations are specified at interfaces between adjacent porous layers and at the interfaces between the porous layers and the fluid gaps. The case of a heavy septum covering porous layers is also studied.

The global analytical solution is derived for the case of an incident acoustic plane wave.

The third section is dedicated to the calculation and optimization of the acoustic AC with respect to Biot's parameters and the thickness of porous layers. A new absorber composed of a foam layer covered by a thin fabric has been optimized using the proposed analytical model. The results of the analytical study are validated using impedance tube measurements showing the effects of mounting conditions.

The fourth section of the paper is dedicated to calculations of the transmission loss (TL) factor of an impervious limp mass layer covered by a porous layer and of the noise reduction (NR) factor corresponding to two limp mass layers covered by the optimized porous layers and coupled by an air gap. Interesting new results, showing a comparison of TL

[a]Electronic mail: jkanfoud@utc.fr

and NR curves obtained with the classical local acoustic impedance (LAI) and Biot's models, are presented at the end of this section. Finally, the last section concludes the paper and gives a perspective of the present research work.

II. THE MODIFIED BIOT EQUATIONS

Standard Biot equations [Eqs. (1a), (1b), (2a), and (2b)] written in terms of the skeleton displacement vector (U^s) and the fluid displacement vector (U^f) are the following:

$$(\hat{\rho}_s \ddot{U}^s + \hat{\rho}_{sf} \ddot{U}^f) = \nabla \cdot \boldsymbol{\sigma}^s - b(\dot{U}^s - \dot{U}^f), \tag{1a}$$

$$(\hat{\rho}_{sf} \ddot{U}^s + \hat{\rho}_f \ddot{U}^f) = \nabla \cdot \boldsymbol{\sigma}^f - b(\dot{U}^f - \dot{U}^s), \tag{1b}$$

where $\dot{U}$ and $\ddot{U}$ correspond to the velocity and acceleration vectors of the solid (s) and fluid (f) phases and σ^s and σ^f are, respectively, the solid and fluid stress tensors given in Ref. 2 by

$$\sigma_{ij}^s = \left[\left(P - \frac{2\mu}{3}\right)\nabla \cdot U^s + Q \nabla \cdot U^s\right]\delta_{ij} + \mu\left[\frac{\partial U_i^s}{\partial x_j} + \frac{\partial U_j^s}{\partial x_i}\right], \tag{2a}$$

$$\sigma_{ij}^f = -\phi p \delta_{ij} = [Q \nabla \cdot U^s + R \nabla \cdot U^s]\delta_{ij}, \tag{2b}$$

where μ is the shear modulus of the skeleton elastic material, and P, Q, and R are the bulk moduli of the porous media. According to Ref. 2, they are related to the bulk modulus K_s of the skeleton elastic material, to the bulk modulus K_f of the interstitial fluid, to the bulk modulus K_b of the porous frame at constant pressure in the air, and to the porosity ϕ by the following expressions:

$$P = \frac{(1-\phi)(1-\phi-K_b/K_s)K_s + \phi K_s K_b/K_f}{1-\phi-K_b/K_s+\phi K_s/K_f}, \tag{3a}$$

$$Q = \frac{(1-\phi-K_b/K_s)\phi K_s}{1-\phi-K_b/K_s+\phi K_s/K_f}, \tag{3b}$$

$$R = \frac{\phi^2 K_s}{1-\phi-K_b/K_s+\phi K_s/K_f}. \tag{3c}$$

The coefficient b appearing on the right hand side of Eqs. (1a), (1b), (2a), and (2b) corresponds to the viscous coupling factor, and the other inertial coefficients appearing on the left side of the above equations are given by the following formulas:

$$\hat{\rho}_{sf} = (1-\alpha_\infty)\phi\rho_f, \tag{4a}$$

$$\hat{\rho}_f = \phi\rho_f - \hat{\rho}_{sf} = \alpha_\infty \phi \rho_f, \tag{4b}$$

$$\hat{\rho}_s = (1-\phi)\rho_s - \hat{\rho}_{sf}, \tag{4c}$$

where ρ_s and ρ_f are the mass densities of the skeleton and of the interstitial fluid, and $\alpha_\infty \geq 1$ is the high frequency limit of the dynamic tortuosity of the considered porous media, which is a dimensionless parameter.[2]

As shown in Ref. 9, for $e^{-i\omega t}$ harmonic time dependency, the modified Biot equations can be easily derived by eliminating the fluid phase displacement vector U^f from the system of Eqs. (1a), (1b), (2a), and (2b) in terms of the skeleton displacement, which for simplicity is denoted as U (without the prefix s) in the rest of the paper, and the interstitial fluid pressure p:

$$\tilde{\rho}_s \omega^2 U + \nabla \cdot (\boldsymbol{\sigma}^s - \alpha\phi p \boldsymbol{\delta}) + \beta \nabla(\phi p) = 0, \tag{5a}$$

$$\nabla \cdot \left(\frac{1}{\tilde{\rho}_f \omega^2}\nabla(\phi p) - \beta \mathbf{U}\right) + \frac{\phi p}{R} + \alpha \nabla \cdot U = 0. \tag{5b}$$

The interstitial fluid displacement is related to the gradient of the acoustic pressure and the skeleton displacement by the following formula:

$$U^f = \frac{1}{\omega^2 \tilde{\rho}_f}\nabla(\phi p) - \frac{\tilde{\rho}_{sf}}{\tilde{\rho}_f} U. \tag{5c}$$

Equations (5a) and (5b) couple the skeleton displacement U to the interstitial fluid pressure p.

Effective masses $\tilde{\rho}_s$, $\tilde{\rho}_f$, and $\tilde{\rho}_{sf}$ appearing in Eqs. (5a)–(5c) are related to the structure and fluid mass densities ρ_s and ρ_f, to dimensionless parameters ϕ (porosity) and α_∞ (tortuosity), and to the viscous coupling factor b of the porous media,

$$\tilde{\rho}_{sf} = (1-\tau)\phi\rho_f + \frac{b}{i\omega}, \tag{6a}$$

$$\tilde{\rho}_f = \tau\phi\rho_f - \frac{b}{i\omega}, \tag{6b}$$

$$\tilde{\rho}_s = (1-\phi)\rho_s - (1-\tau)\phi\rho_f - \frac{(1-\tau)^2\phi^2\rho_f^2}{\tilde{\rho}_f} - \frac{b}{i\omega}. \tag{6c}$$

According to Ref. 2, the viscous coupling factor b is given by the following formula:

$$b = \sqrt{1 - i\omega c^2 \frac{\alpha_\infty \rho_f}{2\phi\sigma}}, \tag{7a}$$

$$c = \frac{1}{\Lambda}\sqrt{\frac{8\alpha_\infty \eta}{\phi\sigma}}, \tag{7b}$$

where σ is the flow resistivity η to the viscosity of interstitial fluid and Λ is the viscous characteristic length.

Coefficients α and β couple the skeleton displacement U to the fluid pressure p in Eqs. (4a) and (4b). They correspond to the dimensionless stiffness and inertial coupling factors. They are given by

$$\alpha = 1 + \frac{Q}{R} = \frac{(1-K_b/K_s)}{\phi}, \tag{8a}$$

$$\beta = 1 + \frac{\tilde{\rho}_{sf}}{\tilde{\rho}_f} = \frac{\rho_f}{\tilde{\rho}_f}. \tag{8b}$$

In general the bulk modulus K_b is very small compared to the bulk modulus K_s of the skeleton material ($K_b \ll K_s$), and for-

mula (8a) shows that the coefficient $\alpha\phi$ can be approximated by unity ($\phi\alpha \cong 1$).

The total stress tensor $\boldsymbol{\sigma}^{\text{tot}}$ applied to an elementary infinitesimal volume of the porous material is given by

$$\boldsymbol{\sigma}^{\text{tot}} = \boldsymbol{\sigma}^{s} - \phi\alpha p\boldsymbol{\delta}, \tag{9a}$$

where $\boldsymbol{\sigma}^{s}$ is the stress tensor inside the skeleton elastic material, which is given by

$$\tilde{\boldsymbol{\sigma}} = (\lambda(\nabla \cdot U)\boldsymbol{\delta} + \mu(\nabla \cdot U + (\nabla \cdot U)^{T})), \tag{9b}$$

where λ and μ are the Lame coefficients.

III. WAVE PROPAGATION IN POROUS MEDIA

The problem investigated in this paper involves the propagation of elastic and acoustic waves inside planar layers of porous material of finite thickness and infinite lateral dimensions. The explicit expressions of the pressure, the skeleton, and fluid displacements and stresses can be expressed in terms of three waves: two longitudinal waves and a shear wave. The global solution is obtained by superposition of the forward and backward traveling waves. For each porous layer there are six waves. Their complex amplitudes are determined by applying the appropriate boundary conditions.

The porous medium is supposed isotropic homogeneous and submitted to an incident unitary plane wave propagating in the x y plane with an angle of incidence θ with respect to the x axis. In such configuration the incident plane wave has the following expression:

$$p_{\text{inc}}(x,y) = e^{ik_0(x\cos\theta + y\sin\theta)}, \tag{10}$$

where $k_0 = \omega/c_0$ is the acoustic wave number ω is the circular frequency, and c_0 is the speed of sound propagating in the surrounding acoustic medium at rest. In this case all field variables inside each porous layer are z independent and consequently have the following form:

$$p(x,y) = p(x)e^{ik_0 \sin\theta.y},$$

$$U(x,y) = U(x)e^{ik_0 \sin\theta.y}.$$

All equations can be written in the (x,y) plane, and since all variables have the same exponential dependency $e^{ik_0 \sin\theta.y}$ in the lateral y direction, the problem reduces to a one-dimensional x dependency.

A. Wave equations

The skeleton displacement U has only two non-null components since the third component in the lateral z direction is identically null. The displacement U could be expressed in the following form:

$$\vec{U}(x,y) = \vec{\nabla}(\Phi(x)e^{ik_0 \sin\theta.y}) + \vec{\nabla}\Lambda(\Psi(x)e^{ik_0 \sin\theta.y}\vec{K}), \tag{11a}$$

where $\vec{K}$ is the unitary vector collinear to the z axis orthogonal to the xy plane of Fig. 1. Substitution of Eq. (11a) in the system of modified Biot's equations [Eqs. (3a)–(3c) and (4a)–(4c)] leads to the following system of three equations:

$$\tilde{\rho}_s\omega^2\Psi + \mu\,\Delta\Psi = 0, \tag{11b}$$

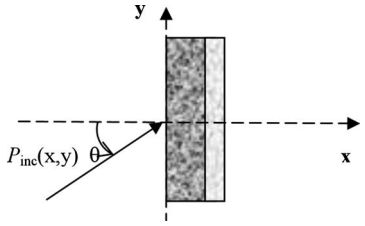


FIG. 1. Incident plane wave exciting porous layers.

$$\tilde{\rho}_s\omega^2\Phi + (\lambda + 2\mu)\Delta\Phi + \gamma\phi p = 0, \tag{11c}$$

$$\frac{\phi\Delta p}{\tilde{\rho}_f\omega^2} + \frac{\phi p}{R} - \gamma\Delta\Phi = 0, \tag{11d}$$

where

$$\gamma = (\beta - \alpha), \tag{12a}$$

$$\Delta = \frac{d^2}{dx^2} - k_0^2 \sin^2\theta. \tag{12b}$$

Substitution of ϕp from Eq. (11b) to Eq. (11c) leads to the following equation:

$$\left\{\Delta^2 + \omega^2\frac{R}{\tilde{\rho}_s}\left(\frac{\tilde{\rho}_s}{(\lambda+2\mu)} + \frac{\tilde{\rho}_f}{R} + \gamma^2\frac{\tilde{\rho}_f}{(\lambda+2\mu)}\right)\Delta + \frac{\tilde{\rho}_s\tilde{\rho}_f}{(\lambda+2\mu)}\omega^4\right\}\Phi = 0. \tag{13}$$

Equation (13) can be written in the following form:

$$\{\Delta^2 + \omega^2 S\Delta + \omega^4 P\}\Phi = 0, \tag{14a}$$

where

$$S = \frac{R}{\tilde{\rho}_s}\left(\frac{\tilde{\rho}_s}{(\lambda+2\mu)} + \frac{\tilde{\rho}_f}{R} + \gamma^2\frac{\tilde{\rho}_f}{(\lambda+2\mu)}\right), \tag{14b}$$

$$P = \frac{\tilde{\rho}_s\tilde{\rho}_f}{(\lambda+2\mu)}. \tag{14c}$$

Equation (14a) can be factorized in the following form

$$\{(\Delta + k_1^2)(\Delta + k_2^2)\}\Phi = 0, \tag{15a}$$

where k_1 and k_2 are the wave numbers corresponding to longitudinal waves given by

$$k_1^2 = \frac{\omega^2}{c_1^2}, \tag{15b}$$

$$k_2^2 = \frac{\omega^2}{c_2^2}, \tag{15c}$$

where

$$c_1^2 = \frac{2}{(S + \sqrt{S^2 - 4P})}, \tag{15d}$$

$$c_2^2 = \frac{2}{(S - \sqrt{S^2 - 4P})}. \tag{15e}$$

c_1 and c_2 correspond to the speeds of slow and fast Biot's longitudinal waves.

The wave Eq. (11a) can be written as follows:

$$\frac{d^2\Psi(x)}{dx^2} + (k_s^2 - k_0^2 \sin^2\theta)\Psi(x) = 0, \tag{15f}$$

where

$$k_s^2 = \omega^2 \frac{\mu}{\tilde{\rho}_s} \tag{16a}$$

corresponds to the square of the shear wave number k_s. Equations (16a) and (16b) have two fundamental solutions,

$$\Psi^+(x) = e^{+i\gamma_s^+ x} \quad \text{and} \quad \Psi^-(x) = e^{-i\gamma_s^- x},$$

where γ_s^+ and γ_s^- are the two roots of the dispersion equation,

$$\gamma_s^2 + k_s^2 - k_0^2 \sin^2\theta = 0. \tag{16b}$$

Equation (15a) shows that Eq. (13) has two solutions $\Phi_j(x)$ $(j=1,2)$ satisfying

$$\left\{\frac{d^2}{dx^2} + (k_j^2 - k_0^2 \sin^2\theta)\right\}\Phi_j(x) = 0 \quad (j=1,2). \tag{17a}$$

Equation (17a) also has two fundamental solutions,

$$\Phi_j^+(x) = e^{+i\gamma_j^+ x} \quad \text{and} \quad \Phi_j^-(x) = e^{-i\gamma_j^- x},$$

where γ_j^+ and γ_j^- are the two roots of the dispersion equation,

$$\gamma_j^2 + k_j^2 - k_0^2 \sin^2\theta = 0. \tag{17b}$$

Once the shear wave numbers k_s and the longitudinal wave numbers k_1 and k_2 are computed, there are six independent fundamental solutions corresponding to the roots of Eqs. (16b) and (17b).

The displacement and the pressure fields relative to each fundamental solution can be computed using Eqs. (11a)–(11d). The global solution can be expressed in the following form:

$$\begin{Bmatrix} U \\ p \end{Bmatrix} = \sum_{j=1}^{4} A_j \begin{Bmatrix} U \\ p \end{Bmatrix}_{\Phi_j} + \sum_{l=1}^{2} B_l \begin{Bmatrix} U \\ 0 \end{Bmatrix}_{\Psi_l}. \tag{18}$$

The six complex amplitudes can be determined using the boundary conditions at the layer's interfaces.

B. Boundary conditions

In this section boundary conditions are written for various interfaces.

Air/hard wall interface. The normal component of the acoustic (prefix 0) displacement is null,

$$U_z^0 = \frac{1}{\omega^2}\frac{dp^0}{dx} = 0.$$

Porous/hard wall interface. The porous medium can be considered to be sliding at the hard wall interface,

$$U_x^1 = U_x^{f1} = 0, \quad \tau_{xy}^1 = 0.$$

The porous medium can be fixed at the hard wall interface

$$U_x^1 = U_x^{f1} = 0; \quad U_y^1 = 0.$$

Air/porous interface. The boundary conditions depend on the nature of the interface, between the air and the porous medium. Since the interface could be impervious (closed cells) or perforated (open cells), the fluid domain is characterized by prefix (0) and the porous medium is characterized by prefix (1).

- perforated interface 0/1: 4 conditions:

$$(U_x^0 - U_x^1) = \phi^1(U_x^{f1} - U_x^1),$$

$$p^0 = p^1,$$

$$-p^0 = \sigma_{xx}^1,$$

$$\tau_{xy}^1 = 0.$$

- impervious interface 0/1: 4 conditions:

$$U_x^0 = U_x^1,$$

$$(U_x^{f1} - U_x^1) = 0,$$

$$-p^0 = \sigma_z^1,$$

$$\tau_{xz}^1 = 0.$$

If the fluid is in direct contact with the porous medium (no facing film), the first equation corresponds to the continuity of the relative fluid flow through the interface. Hence the second equation expresses the continuity of the pressure since the interface is perforated. The third and fourth equations express the continuity of stresses.

If a weightless facing film separates the fluid and porous media domains, there are two cinematic conditions: the first equation expresses the continuity of the skeleton and external fluid normal components of displacements. The second equation determines that the air flow is null through the closed facing film, which is henceforth impervious. The third and fourth equations translate the continuity of stresses.

The facing mass could be taken into account by modifying the last two equations,

$$-p^0 = \sigma_x^1 - m\omega^2 U_x^1 \quad \text{and} \quad \tau_{xy}^1 - m\omega^2 U_y^1 = 0.$$

Porous/porous interface. The boundary conditions depend on the nature of the interface, between the two porous layers since the interface could be impervious (closed cells) or perforated (open cells). The left porous domain is characterized by prefix (1), and the right porous domain is characterized by prefix (2):

Coupling two porous layers requires six boundary conditions.

- perforated interface 1/2:

$$U_x^1 = U_x^2,$$

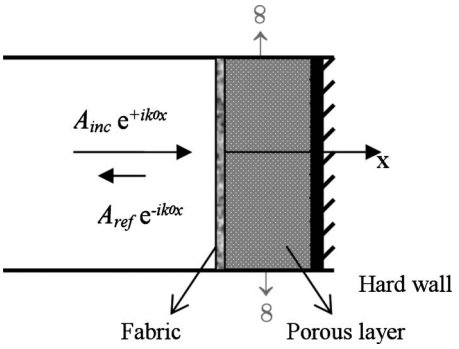


FIG. 2. Impedance tube configuration.

$$U_y^1 = U_y^2,$$

$$\sigma_x^1 = \sigma_x^2,$$

$$\tau_{xy}^1 = \tau_{xy}^2,$$

$$p^1 = p^2,$$

$$\phi^1(U_x^{f1} - U_x^1) = \phi^2(U_x^{f2} - U_x^2).$$

- impervious interface 1/2:

$$U_x^1 = U_x^2,$$

$$U_y^1 = U_y^2,$$

$$\sigma_x^1 = \sigma_x^2,$$

$$\tau_{xy}^1 = \tau_{xy}^2,$$

$$(U_x^{f1} - U_x^1) = 0,$$

$$(U_x^{f2} - U_x^2) = 0.$$

The first four equations are identical for the direct and impervious weightless film interface. These equations are similar to the coupling between elastic layers. The first and second equations express the cinematic coupling of the skeleton displacements. The third and fourth equations express the continuity of stresses.

The two latter equations depend on the configuration considered. For direct coupling, we have pressure and air flow conservation across the interface. For coupling through an impervious interface, the air flow vanishes for both layers.

When the mass of the facing is taken into account, the last two equations must be modified as follows:

$$\sigma_x^1 = \sigma_x^2 - m\omega^2 U_x \quad \text{and} \quad \tau_{xz}^1 = \tau_{xz}^2 - m\omega^2 U_y.$$

IV. ABSORPTION COEFFICIENT

Two standing wave impedance tubes have been used to measure the AC of porous material samples (a small tube of 46 mm diameter and a large tube of square cross section of 600×600 mm^2).

As shown in Fig. 2, numerical simulations have been restricted to the normal incidence case, and results have been obtained using the material properties summarized in Table I. The acoustic AC α_{abs} is defined by the following expression:

$$\alpha_{abs} = 1 - \frac{A_{\text{ref}}^2}{A_{\text{inc}}^2}, \tag{19}$$

where A_{inc} and A_{ref} are, respectively, the amplitudes of the incident and reflected waves.

TABLE I. Material properties.

Parameters	Foam	Fabric
Porosity ϕ	99.4%	13%[a]
Flow resistivity σ (N s m^{-4})	9045	66 639
Viscous length Λ (μm)	103	7.3[a]
Thermal length Λ' (μm)	197	12[a]
Tortuosity α_∞	1.02	1[a]
Skeleton density ρ_s (kg/m^3)	8.43	566.67
Young modulus E (kPa)	194.9	50[a]
Poisson ratio ν	0.42	0.3[a]
Damping factor η	5%	0[a]
Thickness (cm)	10	0.03
Fluid density ρ_0 (kg/m^3)	1.213	
Fluid celerity c_0 (m/s)	342.2	

[a]Parameters identified using analytical simulation and correlation with test data.

Measurements of the AC were made for bare foam and for foam covered by a fabric. Results were obtained in the frequency band ranging from 40 to 250 Hz for the large square tube and from 100 to 4000 Hz for the small circular tube. Two configurations were considered. The first one designated by (B) corresponds to the foam bonded to the termination of the tube. It was simulated by imposing the zero displacements of the skeleton and of the fluid. The second unbonded configuration designated by the symbol (U) is simulated by adding a thin (1 mm) air-gap between the foam and the tube termination.

Figure 3 shows good agreement between the experimental and the numerical results for a bare foam sample of 50 mm thick using both bonded and unbonded boundary conditions. However, to achieve such good agreement, it was necessary to tune the mechanical properties of the skeleton in

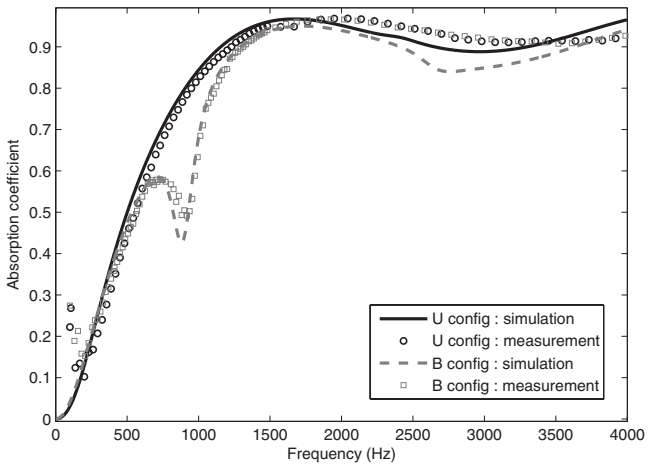


FIG. 3. AC of a bare 5 cm foam layer (U and B configurations).

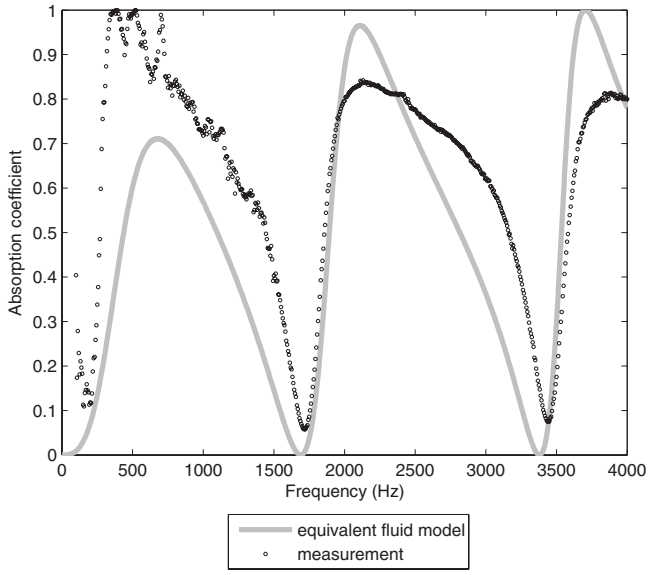


FIG. 4. AC measurement and simulation of a clamped fabric backed by a 10 cm cavity.

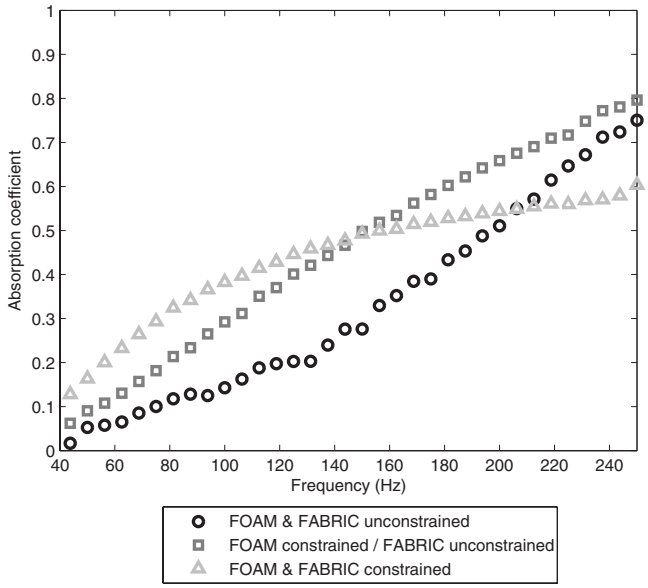


FIG. 5. Measurement of the effect of lateral mounting conditions on a 10 cm foam layer covered by a thin fabric.

order to adjust the resonance frequency of the porous sample since the skeleton resonance for the bonded case is controlled for one-dimensional propagation by the parameter $(\lambda+2\mu)$ appearing in Eq. (11c). This parameter can be modified by either changing Young's modulus or Poisson's ratio. Numerical results have been obtained using Young's modulus of Table I and a modified Poisson's ratio of 0.3 instead of 0.42. The resonance of the bonded foam occurs around 1 kHz, and there is no resonance for the unbonded configuration.

Three properties of the fabric were directly measured with acceptable precision: the flow resistivity, the mass density, and the thickness. The additional Biot parameters $(\phi, \alpha_\infty, \Lambda, \Lambda')$ and mechanical properties (E, ν) have been numerically identified using the best fit between the calculated and measured reflection coefficients for two configurations: (i) the fabric clamped in the small impedance tube and backed by an air gap of 100 mm depth and (ii) the foam covered by the fabric bonded to the termination of the small impedance tube.

Figure 4 shows the comparison between the measured and calculated ACs using the identified fabric parameters summarized in Table I. The frequencies corresponding to minimum and maximum values of the AC are well predicted, but the analytical model underestimates the first maximum level occurring at 500 Hz and overestimates the following maximum levels occurring at 2100 and 3750 Hz.

Here, the influence of lateral boundary conditions[11–13] has been studied experimentally from the 40 to 250 Hz range using the large square tube. Measurements were made on a 100 mm thick foam covered by fabric for different lateral boundary conditions. Figure 5 shows the influence of mounting conditions on the AC. When the fabric and the foam are simultaneously constrained in the impedance tube, the AC is higher in the low frequency band typically below 150 Hz and decreases above 150 Hz in comparison to the two other configurations, where the fabric is not constrained. Below 210 Hz, the third configuration corresponding to the unconstrained foam and fabric leads to the lowest level of the AC. Hence foam and fabric lateral constraints have similar effects. Experimental results show the strong influence of lateral boundary conditions on the AC for the low frequency range.

During the measurement of the AC of the fabric covered foam, very close attention has been paid to minimize the influence of lateral mounting conditions. Figure 6 shows very good agreement between measurements and calculations for the 100 mm thick bare foam and for the same foam sample covered by a fabric. However, to obtain such good agreement the structural loss damping of the foam has been increased to 18% in the numerical simulation. The fabric shifts the absolute maximum of the AC down from 750 to 500 Hz, leading to higher absorption below 500 Hz and lower absorption above 500 Hz. Figure 7(a) shows that the boundary condition between the foam and the tube termination has a significant influence only in the low frequency

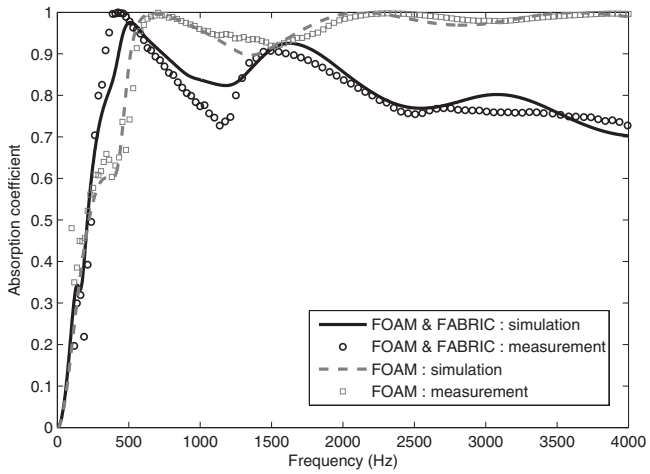
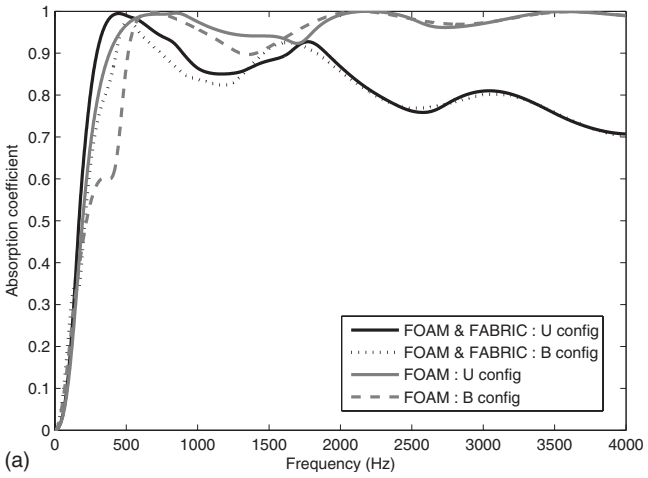


FIG. 6. AC of a bare 10 cm foam layer compared to the foam covered by a fabric.



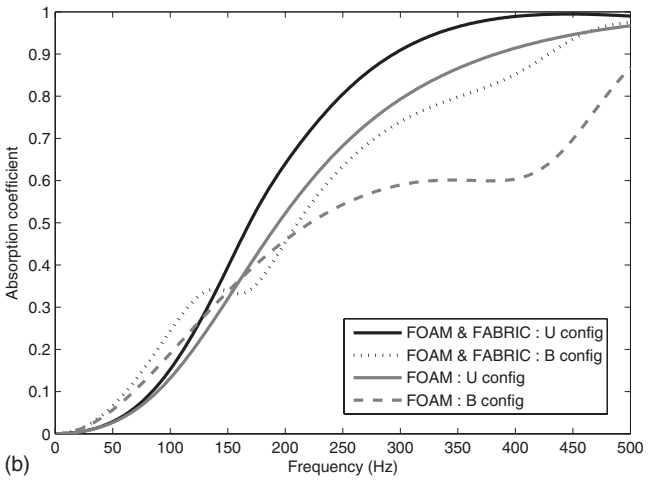


FIG. 7. (a) Simulation of the background mounting conditions on bare and covered 10 cm layers of foam (high frequencies). (b) Simulation of the background mounting conditions on bare and covered 10 cm layers of foam (low frequencies).

band below 750 Hz for the bare foam and 500 Hz for the foam covered by the fabric. The resonance of the skeleton is more pronounced for the bonded configuration. Figure 7(b) corresponds to a zoom below 500 Hz, showing that the unbonded configuration leads to higher absorption from 150 to 500 Hz. On the contrary, below 150 Hz, the bonded configuration leads to higher absorption. Globally, below 500 Hz the fabric increases significantly the level of AC compared to the bare foam.

A. The foam sensitivity to resistivity and Young modulus

In order to maximize the AC of the foam covered by the fabric in the low frequency band, a parametric study was conducted on foam material properties. It is well known that the parameters with most influence are the static air flow resistivity (σ) and the Young modulus (E) of the skeleton.

As shown in Fig. 8, the parametric study highlighted the existence of an optimal value of the resistivity, leading to an absolute maximum of the AC of the acoustic protection consisting of a fabric covered foam sample. The parametric

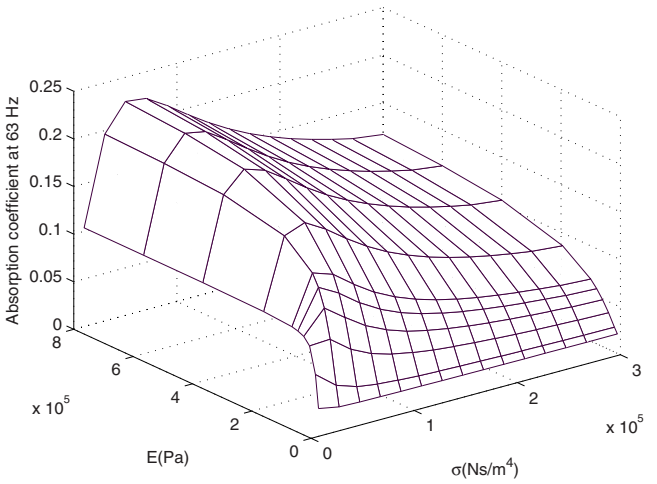


FIG. 8. Sensitivity of the absorption to the resistivity and the Young modulus at 63 Hz for the B configuration.

study was conducted at a particular frequency of 63 Hz for the bonded configuration. Figure 8 shows that a stiffer material leads to higher absorption.

Figure 9 shows variation with respect to the resistivity of the mean value of the AC averaged at the 63 Hz 1/3 octave band for two values of Young's modulus. The nominal value of Young's modulus of the foam and a higher value (ten times the nominal value) have been used. Both bonded and unbonded configurations have been studied, highlighting the influence of boundary conditions between the foam and the tube termination.

For the bonded configuration, the absolute maximum of the AC is obtained for optimal values of resistivity of 40 kN s m^{-4} for the nominal value of the Young's modulus and 50 kN s m^{-4} for the stiffer foam (ten times). The unbonded configuration requires very high resistive foams. In this case the optimal value of resistivity is located above

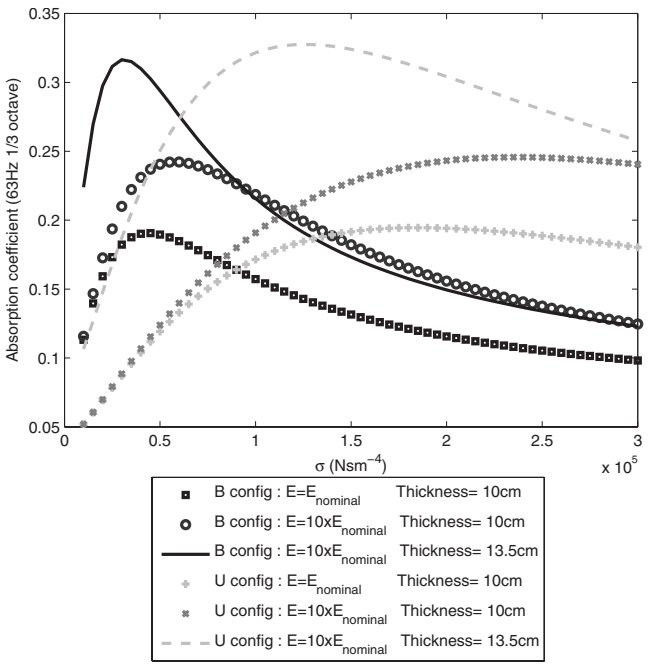


FIG. 9. Sensitivity of the AC to the resistivity at the 1/3 octave of 63 Hz.

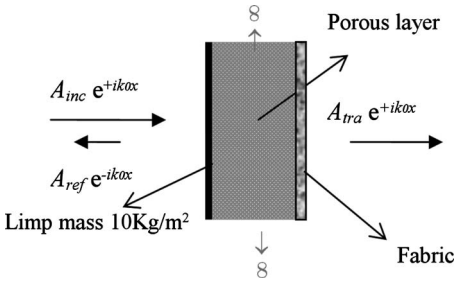


FIG. 10. Transmission loss (TL) configuration.

150 kN s m^{-4}. As expected, Fig. 9 shows that increasing the foam thickness from 10 to 13.5 cm leads to further improvement of the AC.

V. TRANSMISSION LOSS AND NOISE REDUCTION FACTORS

Many sound transmission studies have been conducted on structures lined with porous materials. Bolton *et al.*,[3] Bolton and Shiau,[4,5] Bolton and Green,[14] and Becot and Sgard[15] investigated an analytical model to predict sound transmission through double structure panels lined with porous layers simulating a simplified aircraft fuselage. Other researchers developed various numerical methods to analyze sound radiation and transmission by elastic structures covered by porous elastic layers.[16,17]

Here, the proposed analytical model is applied to calculate the TL and the NR factors of infinite structure panels lined by a foam layer covered by a fabric. The structure panels are represented by a limp mass.

The simple configurations considered try to derive rules for the design of lightweight acoustic protection to be integrated in the fairing of space launchers. In this context, the noise TL coefficient is calculated (Fig. 10) by considering a laterally infinite impervious limp mass of 10 kg/m^2, covered by a porous layer as shown below. The porous layer is 10 cm thick. Nominal parameters given in Table I are used. A Poisson's ratio of ν=0.3 has been used since it led to a better correlation with measurements.

The TL is defined by the following formula:

$$\mathrm{TL} = 10 \log \frac{A_{\mathrm{inc}}^2}{A_{\mathrm{tra}}^2}, \tag{20}$$

where A_{inc} and A_{tra} are the amplitudes of incident and transmitted acoustic waves.

The NR factor simulates the noise level inside a laterally infinite cavity of finite width (Fig. 11) limited by two panels

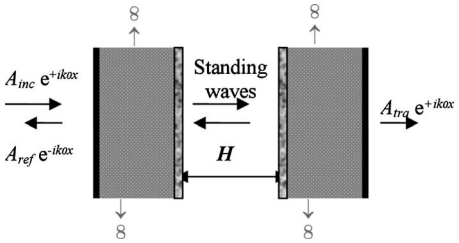


FIG. 11. Noise reduction (NR) configuration.

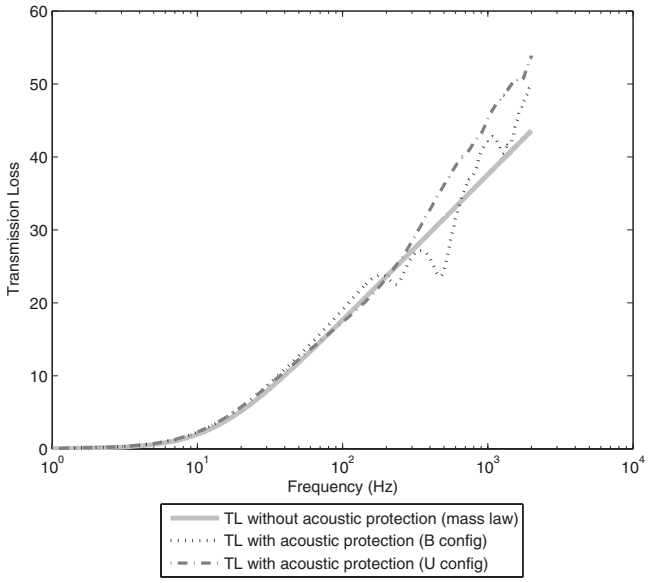


FIG. 12. TL of a limp mass lined with an absorbing complex.

represented by limp masses covered by the optimized acoustic protection. The NR is defined by the following formula:

$$\mathrm{NR} = 10 \log \frac{A_{\mathrm{inc}}^2}{P_{\mathrm{int}}^2}, \tag{21}$$

where P_{int}^2 is given by

$$P_{\mathrm{int}}^2 = \frac{1}{H} \int_0^H p_{\mathrm{int}}(x) p_{\mathrm{int}}^*(x) dx, \tag{22}$$

where $P^*(x)$ is the complex conjugate of the internal pressure $P(x)$.

Figure 12 shows that below 100 Hz, the foam has very little effect on the TL since the curve coincides with the mass law. Above 100 Hz, the porous complex enhances the TL in the case of the unbonded configuration. For the bonded configuration, local frequency drops of the TL curve are observed at porous layer resonance frequencies. A parametric study has been carried out for fixed mechanical properties of the skeleton. It shows that the Biot's parameters related to the propagation of the airborne wave in the porous medium have very little effect on the TL curve. The structural resonance frequencies vary mainly with mechanical properties of the skeleton of the porous layer.

Figure 11 shows an air-filled cavity of width H, limited at each side by a limp mass layer covered by a porous layer simulating the acoustic protection. Figure 13 superimposes TL curves ($H \to \infty$) to the NR curves obtained for a cavity of finite width (H=1 m).

The NR curves obtained with and without the porous layer oscillate around the corresponding TL curves. Without the porous layer, sharp peaks appear at the resonances of the acoustic cavity. The amplitudes of these picks are strongly attenuated when limp mass layers are covered by the porous layers simulating the acoustic protection. When the limp mass is covered by the porous material layers, the NR is very close to the TL curve above 500 Hz. Below 500 Hz the

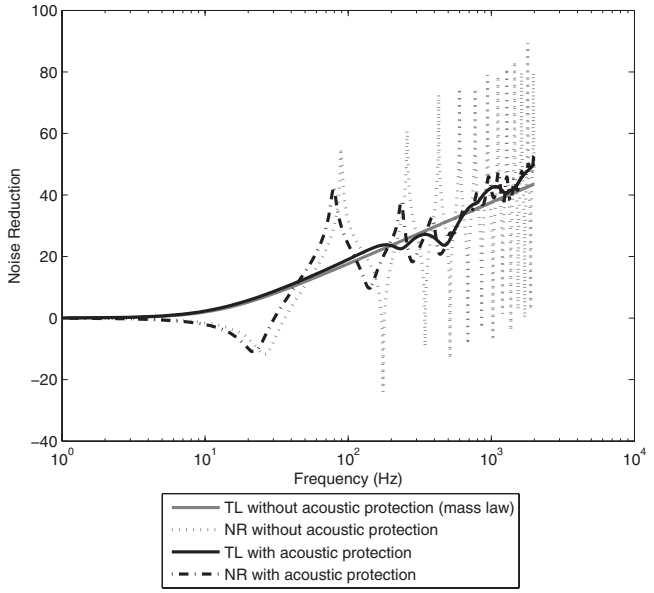


FIG. 13. NR of a cavity of width $H=1$ m.

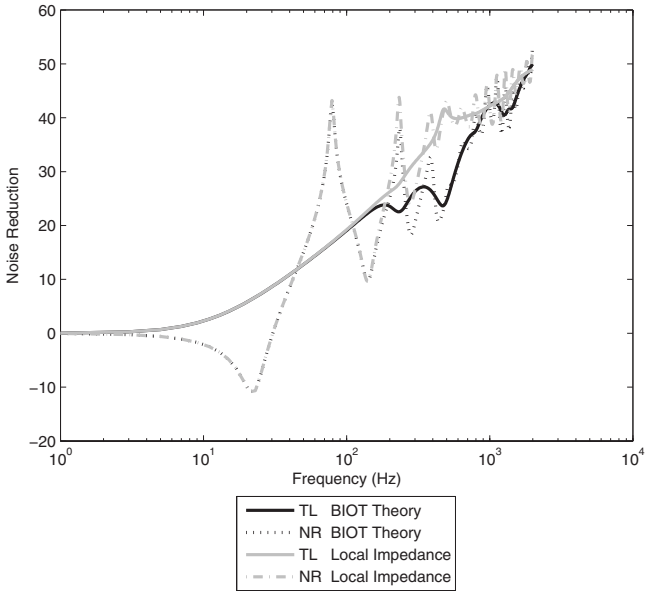


FIG. 14. Comparison of local impedance and Biot models.

optimized porous layer is still very efficient. It attenuates the first acoustic mode by 10 dB, the second by 30 dB, and the third by more than 40 dB.

In the current state of the art, NR factors of space structures[18] are predicted using the LAI measured in an impedance tube. The value of the local impedance is determined by placing the porous layer at the termination of an impedance tube. The associated boundary conditions at the interface of the acoustic cavity are written as follows:

$$\frac{1}{\rho_0}\frac{dp_{\text{int}}}{dx} - \frac{i\omega}{Z}p_{\text{int}} = -\omega^2 U^s, \tag{23a}$$

$$p_{\text{ext}} - p_{\text{int}} = -\omega^2 m U^s, \tag{23b}$$

where ρ_0 is the mass density of air, Z is the LAI of the porous layer bonded at the termination of the impedance tube, p_{int} is the internal acoustic pressure, p_{ext} is the external pressure (sum of incident and reflected waves), and U^s is the displacement of the limp mass m per unit area.

Results corresponding to the LAI model by applying boundary conditions (23a) and (23b) are compared to those obtained using the present modified Biot model. Figure 14 superimposes TL and NR curves corresponding to LAI and present Biot models.

Below 200 Hz, the curves corresponding to the two models coincide. The two models diverge between 200 and 2000 Hz because the LAI model does not predict the resonances of the porous layer and consequently overestimates the NR and the TL. Above 2000 Hz the two models start to converge again.

VI. CONCLUSION

A modified system of Biot equations has been solved analytically for infinite planar layers of porous media. Various acoustic indicators such as the acoustic absorption coefficient (AC), TL, and NR factors have been computed for different configurations. The sensitivity of the acoustic AC to the boundary conditions, to the variation of Biot's acoustical parameters, and to the mechanical properties of the skeleton has been thoroughly studied. The analytical results have been validated by impedance tube measurements, showing good agreement between experimental and numerical results.

This study demonstrated that the mounting boundary conditions (lateral and longitudinal) influence the AC in the low frequency band. The analytical model developed permitted a rapid parametric study to be made of the influence of the two most significant parameters, which are the static air flow resistivity and the Young modulus of the porous material. For a given low frequency band (1/3 octave), an optimal value of the flow resistivity leads to a maximum value of the AC. Stiff materials (high Young's modulus) lead to higher absorption at the low frequency range. For high frequency bands, the increase in resistivity leads to an enhancement in the AC. In low frequency bands, the absolute value of the maximum of the AC is very sensitive to the mechanical properties of the skeleton and the boundary conditions.

The developed method allowed the calculation of the TL factor of limp mass panels covered by the optimized acoustic protection composed by two layers (foam and fabric) of porous materials. It is demonstrated that the TL obtained with the acoustic protection is always higher than the TL curve of the classical mass law, except at the first resonances of the foam layer, where the TL curve drops locally.

The proposed method also permits the calculation of the NR factor of a simplified structure composed of two limp mass layers covered by the optimized acoustic protection and containing an acoustic cavity of finite width. It has been demonstrated that the optimized acoustic protection is very efficient. It allows strong damping of internal acoustic resonances.

In addition the developed method allowed the analysis of results obtained using the approximate LAI model, which gives good results in low and high frequency limits. In the medium frequency range, the LAI model is less accurate than

the present modified BIOT model since it overestimates the NR factor near the resonances of the foam layer.

ACKNOWLEDGMENTS

The authors would like to thank the CNES (Evry, France) for the financial support of the research program leading to the publication of the present paper and to highlight the valuable cooperation with the ENTPE (Lyon, France), which made the measurements for the characterization of the foam and fabric physical properties.

[1]M. A. Biot, "Theory of propagation of elastic waves in a fluid-saturated porous solid I: Low-frequency range II: Higher frequency range," J. Acoust. Soc. Am. **28**, 168–191 (1956).
[2]J. F. Allard, *Propagation of Sound in Porous Media: Modelling Sound Absorbing Materials* (Elsevier, New York, 1993).
[3]J. S. Bolton, N. M. Shiau, and Y. J. Kang, "Sound transmission through multi-panel structures lined with elastic porous materials," J. Sound Vib. **191**, 317–347 (1996).
[4]J. S. Bolton and N. M. Shiau, "Oblique incidence sound transmission through multi-panel structures lined with elastic porous material," in Proceedings of the 11th AIAA Aeroacoustics Conference, Sunnyvale, CA (October 1987), Paper No. AIAA-87-2660, pp. 19–21.
[5]J. S. Bolton and N. M. Shiau, "Random incidence sound transmission through multi-panel structures lined with elastic porous material," in Proceedings of the 12th AIAA Aeroacoustics Conference, San Antonio, TX (October 1989), Paper No. AIAA-89-1048, pp. 10–12.
[6]G. Bonnet, "Basic singular solutions for a poroelastic medium in the dynamic range," J. Acoust. Soc. Am. **82**, 1758–1762 (1987).
[7]N. Atalla, R. Panneton, and P. Debergue, "A mixed displacement-pressure formulation for poroelastic materials," J. Acoust. Soc. Am. **104**, 1444–1452 (1998).
[8]P. Debergue, R. Panneton, and N. Atalla, "Boundary conditions for the weak formulation of the mixed (u,p) poroelasticity problem," J. Acoust. Soc. Am. **106**, 2383–2390 (1999).
[9]M. A. Hamdi, L. Mebarek, A. Omrani, and N. Atalla, "An efficient formulation for the analysis of acoustic and elastic waves propagation in porous-elastic materials," in ISMA 25, International Conference on Noise and Vibration Engineering, Katholieke Universiteit Leuven, Belgium (13–15 September 2000).
[10]N. Atalla, M. Hamdi, and R. Panneton, "Enhanced weak integral formulation for the mixed (U,p) porous-elastic equations," J. Acoust. Soc. Am. **109**, 3065–3068 (2001).
[11]D. Pilon and R. Panneton, "Behavioral criterion quantifying the effect of circumferential air gaps on porous materials in the standing wave tube," J. Acoust. Soc. Am. **116**, 344–356 (2004).
[12]B. H. Song, J. S. Bolton, and Y. J. Kang, "Effect of circumferential edge constraint on the acoustical properties of glass fiber materials," J. Acoust. Soc. Am. **110**, 2902–2916 (2001).
[13]A. Cummings, "Impedance tube measurements on porous media: The effect of air-gaps around the sample," J. Sound Vib. **151**, 63–75 (1991).
[14]J. S. Bolton and E. R. Green, "Normal incidence sound transmission through double-panel systems lined with relatively stiff, partially reticulated polyurethane," Appl. Acoust. **39**, 23–51 (1993).
[15]F. X. Becot and F. Sgard, "On the use of poroelastic materials for the control of the sound radiated by a cavity backed plate," J. Acoust. Soc. Am. **120**, 2055–2066 (2006).
[16]N. Atalla, F. Sgard, and C. K. Amedin, "On the modelling of sound radiation from poroelastic materials," J. Acoust. Soc. Am. **120**, 1990–1995 (2006).
[17]R. Panneton and N. Atalla, "Numerical prediction of sound transmission through finite multilayer systems with poroelastic materials," J. Acoust. Soc. Am. **100**, 346–354 (1996).
[18]H. Defosse and M. A. Hamdi, "Vibro-Acoustic study of Ariane V launcher during lift off," in Proceeding of the 29th International Congress on Noise Control Engineering, Inter-Noise 2000, Nice, France (27–30 August 2000).

Demonstration of a wireless, self-powered, electroacoustic liner system[a)]

Alex Phipps
Department of Electrical and Computer Engineering, University of Florida, Gainesville, Florida 32611-6130

Fei Liu, Louis Cattafesta, and Mark Sheplak
Department of Mechanical and Aerospace Engineering, University of Florida, Gainesville, Florida 32611-6250

Toshikazu Nishida[b)]
Department of Electrical and Computer Engineering, University of Florida, Gainesville, Florida 32611-6130

(Received 19 February 2008; revised 24 October 2008; accepted 21 November 2008)

This paper demonstrates the system operation of a self-powered active liner for the suppression of aircraft engine noise. The fundamental element of the active liner system is an electromechanical Helmholtz resonator (EMHR), which consists of a Helmholtz resonator with one of its rigid walls replaced with a circular piezoceramic composite plate. For this system demonstration, two EMHR elements are used, one for acoustic impedance tuning and one for energy harvesting. The EMHR used for acoustic impedance tuning is shunted with a variable resistive load, while the EMHR used for energy harvesting is shunted to a flyback power converter and storage element. The desired acoustic impedance conditions are determined externally, and wirelessly transmitted to the liner system. The power for the receiver and the impedance tuning circuitry in the liner are supplied by the harvested energy. Tuning of the active liner is demonstrated at three different sound pressure levels (148, 151, and 153 dB) in order to show the robustness of the energy harvesting and storage system. An acoustic tuning range of approximately 200 Hz is demonstrated for each of the three available power levels. © *2009 Acoustical Society of America.* [DOI: 10.1121/1.3050287]

PACS number(s): 43.50.Gf, 43.38.Fx, 43.55.Ev [AJZ] Pages: 873–881

I. INTRODUCTION

Noise suppression provided by acoustic liners in aircraft engine ducts is necessary for modern commercial aircraft to meet Federal Aviation Administration flyover noise certifications.[1,2] There are two categories of acoustic liners used to reduce engine noise: passive and active. Passive acoustic liners usually employ a single layer system consisting of a honeycomb septum sandwiched between a solid backplate and a perforated facesheet that provides complex impedance boundary conditions for the noise propagation along the engine duct.[1] The advantage of passive liners is their simplicity and structural integrity, which leads to relatively low-cost fabrication and implementation. However, such systems are only capable of operating over a relatively narrow frequency range. For instance, the useful bandwidth of a single degree-of-freedom liner is approximately one octave.[1] Although it is possible to extend the frequency range of noise reduction of passive liners by adopting a multilayer structure, the weight and size requirements are often prohibitive. Moreover, the performance characteristics of a passive acoustic liner are essentially fixed, except for flow and acoustic level dependence.[1]

Active or adaptive liners, on the other hand, offer the promise of *in situ* adjustment of liner performance over a broad noise frequency range. Generally, active or adaptive liners are systems that employ sensors, mechanical actuators, and controllers to modify the geometry of the liner or use bias flow through the perforated sheet.[3–6] However, the sensors and actuators add size and complexity to the liner system, making design and implementation difficult.

In order to achieve *in situ* adjustment of the liner without the use of mechanical actuators, a novel method using electromechanical Helmholtz resonators (EMHRs) has been reported by Horowitz *et al.*[7] and Liu *et al.*[8] An EMHR, shown in Fig. 1, is comprised of an orifice or neck, backing cavity, and a compliant, composite piezoelectric backplate that is attached to an electrical shunt network. By modifying the shunt network, the acoustic impedance of the EMHR and its resonant frequencies can be tuned *in situ*.[7,8]

Similar to all active tuning methods, control signals and power must be provided for the tuning circuitry of the EMHR. The control signals can be transmitted wirelessly to EMHR tuning networks, which eliminates electrical wiring but does not alleviate the power requirements. An alternative to providing power from an external source is to configure

[a)]Preliminary portions of this work were presented in "A self-powered wireless active acoustic liner," 12th AIAA/CEAS Aeroacoustics Conference, Cambridge, MA, May 2006, AIAA Paper No. 2006-2400 and "Technology development for electromechanical acoustic liners," Active 04, Williamsburg, VA, May 2004.

[b)]Electronic mail: nishida@ufl.edu

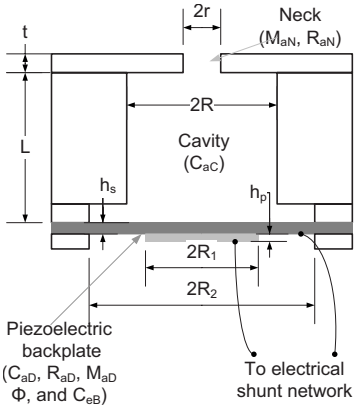


FIG. 1. Cross section of an EMHR.

the system to be self-powered. A self-powered system is one in which all of the required operating energy can be harvested from the environment.[9–11] Recently, it was shown by Liu *et al.*[8] that energy harvesting in an acoustic environment can be achieved with the same EMHR geometry used for acoustic impedance tuning.[12] Thus, a wireless self-powered active liner system can be realized by employing an array of EMHR elements with a common geometry. Figure 2 shows one potential configuration in which EMHR elements can be used to either adjust the local acoustic impedance boundary condition or to harvest the energy needed to operate the liner.

This paper presents the integration and experimental verification of an acoustically self-powered active liner *system*, expanding on the previous preliminary work reported by Kadirvel *et al.*[13] In the previous conference paper, results on the operation of each *isolated* component were presented. In this paper, the frequency tuning of the acoustically self-powered active liner is characterized in detail as a whole, verifying interoperability of the entire system. Three different acoustic energy harvesting scenarios are employed to demonstrate system robustness and sustainability with different amounts of available acoustic power.

The paper is organized as follows. Section II examines the operation of the active liner system. The complete integrated system is presented first, followed by a detailed discussion of each system block. Section III describes the specific components used for implementation of the prototype

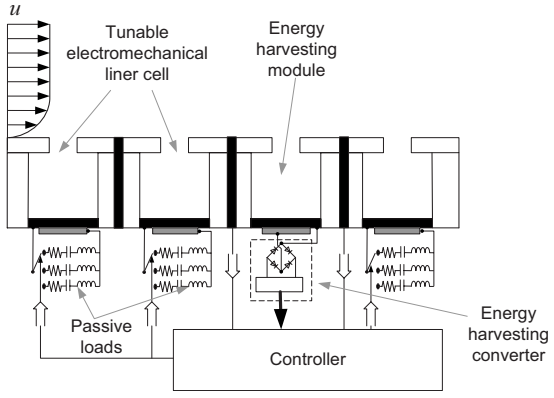


FIG. 2. Active liner composed of an array of two types of EMHR elements, one to tune the acoustic impedance of the liner and the other to harvest acoustic energy.

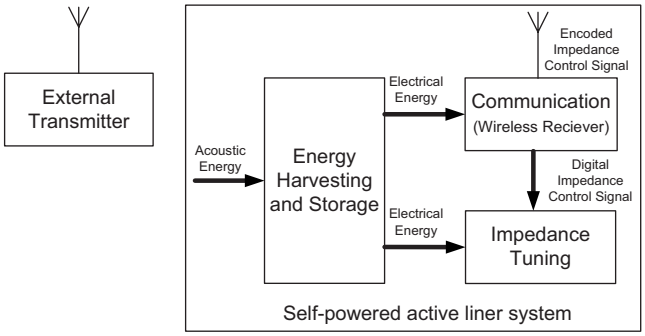


FIG. 3. Active liner system with external transmitter and various blocks for impedance tuning, energy harvesting, and communication.

and the methods employed for experimental verification. Experimental results are presented in Sec. IV, and conclusions and future work are provided in Sec. V.

II. EMHR-ARRAY ACTIVE LINER SYSTEM

The active liner system in this paper is operated in an open-loop configuration to reduce complexity and minimize the energy requirements of the liner. Closed-loop systems require output detection and feedback circuitry, both of which increase the system complexity and size, and possibly the power requirements.

The self-powered active liner system presented in this work can be divided into three functional blocks: impedance tuning, energy harvesting, and communication. A system diagram, shown in Fig. 3, illustrates how the blocks interact to tune the impedance boundary conditions inside the duct. The impedance tuning block is comprised of components that physically adjust the local impedance boundary condition of the duct. In this open-loop implementation, the control signal to set the desired boundary conditions is wirelessly transmitted from an external transmitter in the form of the impedance control signal (ICS). The ICS is received by the communication block where it is decoded, and is used by the impedance tuning block to set the desired acoustic impedance. Incident acoustic energy is converted into electrical energy within the energy harvesting block, where it is used to power the other liner blocks or is stored for future use.

A. Impedance tuning block

The EMHR element is the fundamental structure within the impedance tuning block for adjusting the acoustic boundary conditions for the noise propagation along the duct. The EMHR is a multiple energy domain device, which couples the acoustic and electrical domains via the piezoelectric composite backplate. A lumped element model (LEM) of the EMHR was developed by Liu *et al.*[8] The LEM, shown in Fig. 4, models the parameters of the EMHR as discrete cir-

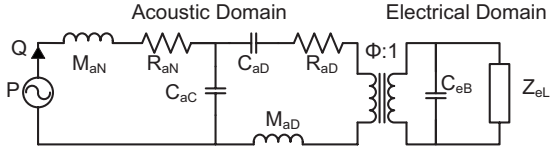


FIG. 4. LEM of the EMHR.

TABLE I. LEM parameters for a single EMHR element.

LEM parameter	Definition
P	Incident acoustic pressure
Q	Incident volumetric flow
M_{aN}	Neck acoustic mass
R_{aN}	Neck acoustic damping loss
C_{aC}	Cavity acoustic compliance
C_{aD}	Diaphragm acoustic compliance
M_{aD}	Diaphragm acoustic mass
R_{aD}	Diaphragm acoustic damping loss
ϕ	Piezoelectric transduction ratio
C_{eB}	Blocked electrical capacitance

cuit elements connected to a passive electrical load, Z_{eL}. The definition of each element in the LEM is provided in Table I. Unlike a standard Helmholtz resonator, whose acoustic impedance is only a function of the device geometry, the acoustic impedance of the EMHR is also a function of its electrical load. The acoustic input impedance of an EMHR is given by[8]

$$Z_{aIN} = R_{aN} + sM_{aN} + \left. \frac{\frac{1}{sC_{aC}}\left(\frac{1}{sC_{aD}} + R_{aD} + sM_{aD} + \frac{\phi^2 Z_{eL}}{1 + sC_{eB}Z_{eL}}\right)}{\frac{1}{sC_{aC}} + \frac{1}{sC_{aD}} + R_{aD} + sM_{aD} + \frac{\phi^2 Z_{eL}}{1 + sC_{eB}Z_{eL}}} \right|_{s=j\omega}, \quad (1)$$

where $s=j\omega$ and ω is the angular frequency. The effects of using resistive, capacitive, or inductive electrical loads to tune the acoustic impedance were demonstrated by Liu *et al.*[8] For illustration purposes, only purely resistive electrical loads are considered in this paper. The EMHR with a resistive load is analogous to a two degree-of-freedom system, and has two resonant frequencies [i.e., f_1 and f_2 at $\mathrm{Im}(Z_{aIN})=0$]. Both resonant frequencies of the EMHR can be tuned simply by varying the resistance value. The tuning range of the resonant frequencies extend between the short-circuit ($Z_{eL}=0$) and open-circuit cases ($Z_{eL}=\infty$).[8]

The other element of the impedance tuning block is the switching array, which is used to vary the shunt electrical impedance across the EMHR. The switching array is comprised of N switch-load pairs connected in parallel across the EMHR. The case of purely resistive loads is shown in Fig. 5. After the ICS is received and decoded in the communication block, an N-bit digital signal is sent to the switching array. Each of the N bits has a logic value of either 0 or 1, and is used to control a specific switch. Depending on the logic value, the switch is either opened (i.e., 0) or closed (i.e., 1) to set the shunt impedance of the EMHR. While the ability to have multiple switches closed at the same time increases the number of possible impedance values, this work will only consider the case of any one of the N single closed switches for simplicity.

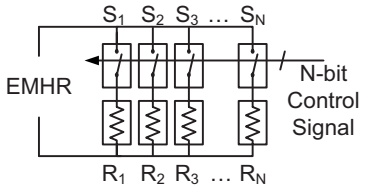


FIG. 5. Switching array in the impedance tuning block using purely resistive loads.

B. Energy harvesting block

The energy harvesting block is designed to provide electrical power to the active liner system. It is comprised of an EMHR element, a power converter, and a storage device. The EMHR can gather energy from the high intensity acoustic field present in the engine nacelle, which can reach levels up to 160 dB (ref. 20 μPa).[14] As stated previously, the EMHR element used in this block has the same geometry as in the impedance tuning block but functions as a transducer to convert acoustic energy into electrical energy. Electrical energy harvested by the EMHR is conditioned by the power converter and is either delivered to the liner system or stored for later use. Figure 6 shows the general flow of power in the energy harvesting block. The behavior of the storage element is determined by the amount of conditioned power available from the EMHR and power converter. When the amount of conditioned power is larger than what is required by the liner electronics, excess power flows into the storage element. When the two are equal, the net power flow into the storage element is zero. For the case where the required power is larger than the available conditioned power, power flows from the storage element for sustained operation.

The role of the power converter in the energy harvesting block is to maximize the amount of electrical power delivered from the EMHR energy harvester to the load. The circuitry used in the power converter, shown in Fig. 7, is comprised of a full-bridge rectifier-capacitor circuit and a flyback converter. Operation of this circuit for acoustic energy harvesting was examined in detail by Liu *et al.*,[12] and is only briefly reviewed here. The active liner electronics requires a stable dc voltage, and the rectifier is used to condition the time varying ac signal produced by the EMHR. It was shown that for a rectifier-capacitor circuit driven by a sinusoidal source, an optimal resistive load for the rectifier exists for

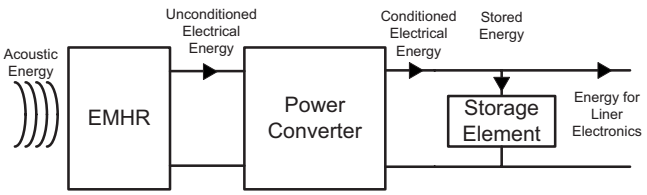


FIG. 6. Power flow diagram of the energy harvesting block.

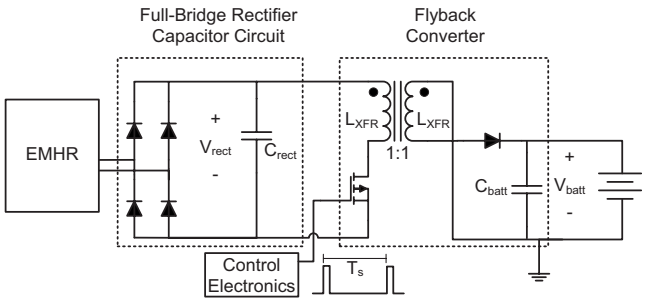


FIG. 7. Detailed diagram of the circuitry within the energy harvesting block.

which the power delivered from the EMHR to the load will be maximized.[12] The flyback converter is used to emulate this optimal loading condition for the rectifier. The input impedance of the flyback converter, operated in discontinuous conduction mode, is purely resistive and is given by[15]

$$R_{\text{in}} = \frac{2f_s L_{XFR}}{d^2}, \tag{2}$$

where L_{XFR} is the inductance of the flyback transformer, f_s is the switching frequency of the converter, and d is the duty cycle. Adjustment of the input impedance can be controlled electronically by varying d. Since the input impedance of the flyback is independent of its load, it is placed between the EMHR energy harvester and the liner electronics to decouple the two and provide the optimal resistance.

A rechargeable battery is chosen as the storage element in the energy harvesting block because its voltage remains approximately constant as it is charged and discharged. The constant voltage of the battery can be used to provide a stable voltage source for the liner electronics and eliminates the need for voltage regulation circuitry.

C. Wireless communication block and system integration

The desired acoustic impedance is determined externally to the active liner, as a function of engine operating conditions (e.g., take-off, cut back, and approach), and the appropriate ICS is wirelessly transmitted to the system. The communication block, comprised of the self-powered receiver and associated electronics, receives the ICS. A detailed block diagram, presented in Fig. 8, illustrates how the communication block interacts with the other system blocks. The wireless receiver is integrated directly into the liner system and receives its power from the energy harvesting block. The wireless transmitter is external to the active liner system and receives power and input signals externally.

Transmission of the ICS is accomplished by a Holtek HT-12E encoder and a Rayming TX-99 transmission module, shown in Fig. 9(a). The HT-12E encodes 12 bits of digital data for transmission: 4 bits for the ICS and 8 address bits. Each of the four ICS bits corresponds to a specific switch in the impedance tuning switching array and determines whether it will be opened or closed. The address bits allow the transmitted signal to target specific receiver nodes.

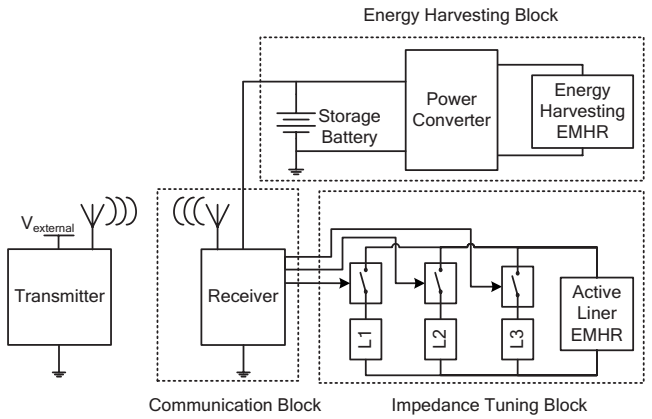


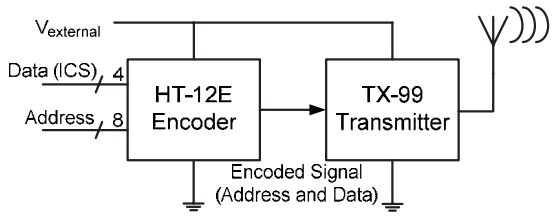
FIG. 8. Detailed block diagram of the integrated active liner system with three different resistive loads L1, L2, and L3.

Once encoded, the signal is transmitted by the TX-99 using an amplitude modulation scheme at 300 MHz with a loop antenna.

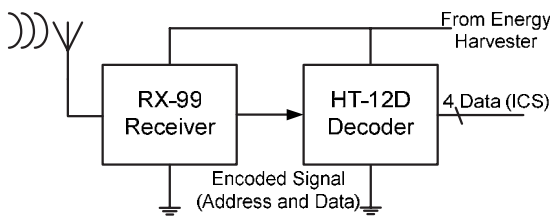
The data transmitted by the TX-99 are received at the active liner by its complementary receiver, the Rayming RX-99. The Holtek HT-12D decoder separates the 12 data bits to recover the ICS and address, as shown in Fig. 9(b). If the transmitted address bits match the preset address of the receiver, the ICS is latched in and sent to the impedance tuning block to select the appropriate load.

III. EXPERIMENTAL SETUP

Previous studies have considered the operation of the impedance tuning and energy harvesting subsystems individually.[8,12] For the self-powered active liner system presented in this work, operation of the entire integrated system is demonstrated. In practice, a complete active liner system would be comprised of an array of EMHR elements. For proof of concept, the active liner system in this work consists of only two EMHR elements—one for noise suppression and



(a)



(b)

FIG. 9. Schematic of (a) the external transmitter and (b) the receiver circuitry used in the communication block.

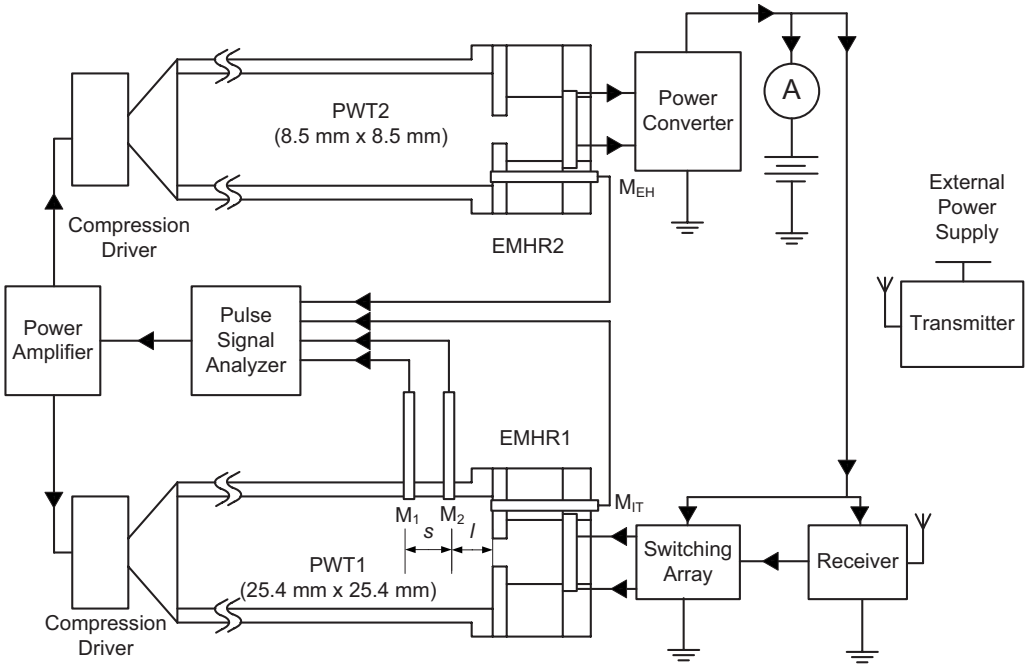


FIG. 10. Experimental setup used for self-powered active liner system testing.

one for energy harvesting—and a single receiver module. To characterize the acoustic impedance of the EMHR, the two microphone method (TMM) is used.[16,17] The TMM is a standard technique for the determination of the normal specific acoustic impedance of acoustic samples.[16] The TMM employs a plane wave tube (PWT) and uses two flush wall-mounted microphones to simultaneously measure acoustic pressure at two known positions in the tube. The acoustic sample is mounted at one end of the PWT and the sound source at the other end. The frequency response function (FRF) is computed between the two microphones using the measured pressure signals. The complex acoustic reflection coefficient is then calculated by

$$\Re(\omega) = \frac{\hat{H}(\omega) - e^{-jks}}{e^{jks} - \hat{H}(\omega)} e^{j2k(l+s)}, \tag{3}$$

where $\hat{H}(\omega)$ is the estimated FRF, $k = \omega / c_0$ is the wavenumber, c_0 is the isentropic speed of sound, s is the spacing between the microphone locations, and l is the distance from the sample to the nearest microphone location. The normalized specific acoustic impedance is then calculated by

$$\zeta(\omega) = \theta(\omega) + j\chi(\omega) = \frac{1 + \Re(\omega)}{1 - \Re(\omega)}, \tag{4}$$

where $\theta(\omega)$ is the normalized resistance and $\chi(\omega)$ is the normalized reactance. Once the normalized specific acoustic impedance of the EMHR is known, the power into the EMHR is then computed by

$$W = \frac{|P|^2 \cos \vartheta}{2|\zeta(\omega)/A_D|}, \tag{5}$$

where P is the input pressure at the entrance of the EMHR, ϑ is the phase angle of ζ, and A_D is the cross-sectional area of the PWT. Note that the power into the EMHR is less than the acoustic power of the incident plane wave since a portion of the incident acoustic power is reflected by the EMHR. The incident acoustic power is given by

$$W_{\text{in}} = \frac{W}{1 - |\Re(\omega)|^2}. \tag{6}$$

The experimental setup used to demonstrate operation of the integrated active liner system is shown in Fig. 10. Two PWTs, PWT_1 and PWT_2, are used to demonstrate the impedance tuning and energy harvesting performance of the EMHRs, respectively. The signal used in each of the PWTs is generated using the Bruel & Kjær (B&K) pulse analyzer system and amplified with a Techron 7540 power amplifier. A BMS 4590 compression driver in each tube uses the amplified signal to create the acoustic field. The EMHR backplate is comprised of an APC 850 piezoelectric disk on a brass shim. The dimensions of the PWTs are shown in Fig. 10 and the dimensions of the EMHR element used in each tube are given in Table II. The material properties for the piezoelectric backplates are provided in Table III. A total of four B&K 4138-A 1/8 in. microphones are used to simultaneously measure acoustic pressure in the PWTs. Two microphones, M_{IT} (impedance tuning) and M_{EH} (energy harvesting), monitor the acoustic pressure at the faces of the EMHR in PWT_1 and PWT_2, respectively. Two additional microphones, M_1 and M_2, are used in PWT_1 to measure the normal specific acoustic impedance of the EMHR using the standard TMM.

The flyback converter in the energy harvesting block uses the standard topology presented in Fig. 7. Both capacitors, C_{rect} and C_{batt}, are 330 μF, and the inductance of the flyback transformer, L_{XRF}, is 6.8 mH. A complete discussion of the control electronics can be found in Ref. 18. The storage element is a 5 V battery comprised of four rechargeable AA Energizer cells in series. An ammeter placed in series with the battery monitors the magnitude and direction of

TABLE II. EMHR dimensions.

Dimensions of EMHR1	
Neck	Dimension (mm)
Radius r	2.42
Length t	3.16
Cavity	Dimension (mm)
Radius R	6.34
Depth L	16.4
Piezoelectric backplate	Dimension (mm)
Radius of the piezoelectric layer R_1	9.06
Thickness of the piezoelectric layer h_p	0.11
Radius of shim R_2	13.5
Thickness of shim h_s	0.22
Dimensions of EMHR2	
Neck	Dimension (mm)
Radius r	2.42
Length t	3.16
Cavity	Dimension (mm)
Radius R	6.34
Depth L	16.4
Piezoelectric backplate	Dimension (mm)
Radius of the piezoelectric layer R_1	11.75
Thickness of the piezoelectric layer h_p	0.14
Radius of shim R_2	17.25
Thickness of shim h_s	0.33

power flow of the storage element. The power for the transmitter is provided externally with a 9 V voltage source.

The switching array of the impedance tuning block is configured with four different electrical loads; open circuit, short circuit, 5 kΩ, and 500 Ω, where each electrical load provides a different acoustic impedance in accordance with

TABLE III. Materials properties of the piezoelectric backplate.

Piezoceramic APC850	
Young's modulus (GPa)	63
Poisson's ratio	0.31
Density (kg/m^3)	7700
Relative dielectric constant	1750
Piezoelectric strain constant d_{31} (pC/N)	−175
Shim (260 half hard brass)	
Young's modulus (GPa)	110
Poisson's ratio	0.38
Density (kg/m^3)	8530

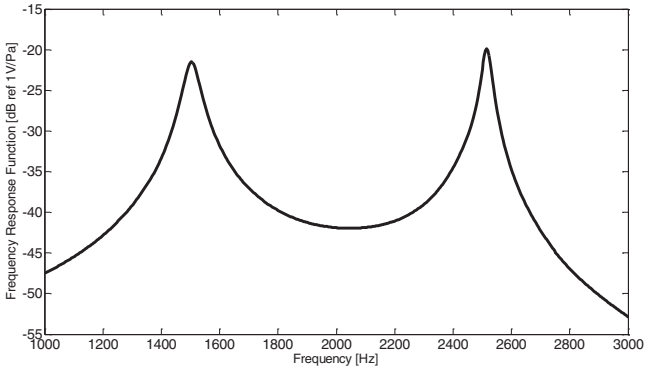


FIG. 11. Measured FRF (output voltage vs input acoustic pressure) of the EMHR in the energy harvesting block.

Eq. (1). These specific resistive loads were chosen to demonstrate the tuning range of the acoustically self-powered active liner. MAX 4544 single-pole, double-throw switches were used to connect the resistive loads to the EMHR when a logic value of 1 was sent by the ICS.

IV. EXPERIMENTAL VERIFICATION AND RESULTS

The first step in testing the active liner system was to determine the optimum acoustic signal to be used for energy harvesting. Turbomachinery noise in aircraft nacelles exhibit both tonal and broadband characteristics.[19] For this demonstration, a single frequency excitation was chosen to emulate the tonal blade passage frequency component. For the EMHR in the energy harvesting block, the FRF between the output voltage and the input acoustic pressure, shown in Fig. 11, exhibited two peaks, which coincided with the two resonant frequencies of the EMHR.[8] Due to the weak coupling between the piezoelectric backplate and the solid-walled Helmholtz resonator,[8] the first peak is associated with the solid-walled Helmholtz resonator and occurs at 1.504 kHz. The second peak is associated with the resonant frequency of the piezoelectric backplate of the EMHR and occurs at 2.516 kHz. The excitation signal was then set to 2.516 kHz because it provided a larger electrical response for a given acoustic input. This maximizes the amount of energy harvested by the active liner.

As stated previously, for an EMHR element connected to a rectifier-capacitor circuit, a maximum amount of power will be harvested when the circuit is loaded with an optimal resistance value. The flyback converter is used to emulate a resistive load, and can be electronically tuned to the optimal resistance. For this experiment, the optimal resistance was found experimentally by replacing the flyback converter with a range of resistor values. The output voltage across each resistance was measured and the power was calculated using

$$P = \frac{V^2}{R}, \tag{7}$$

where V is the measured output voltage across the resistance R. The results of this test, shown in Fig. 12, indicate that the optimal resistance is approximately 20 kΩ. The flyback converter was then replaced, and tuned using Eq. (2). With a

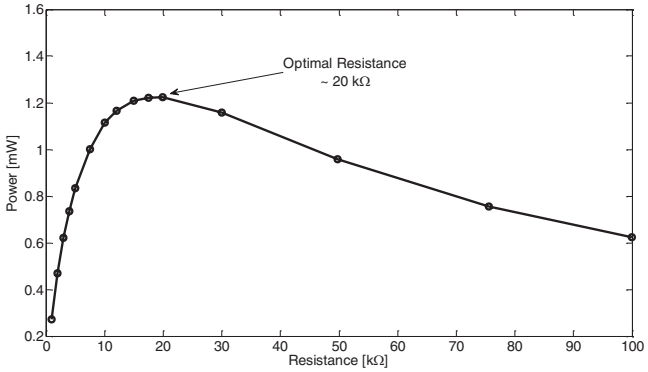


FIG. 12. Output power vs resistance for an EMHR element showing an optimal resistive load of 20 kΩ.

6.8 mH inductor and a switching frequency of 5 kHz, the duty cycle of the pulse width modulated (PWM) signal was set to 5.8% to achieve the optimal resistance of 20 kΩ.

Next, in order to determine if the self-powered active liner system was properly tuning the acoustic impedance of the duct, a control experiment was performed. Discrete loads, corresponding to the four loading conditions of the active liner (open circuit, short circuit, 500 Ω, and 5 kΩ), were directly shunted to the EMHR in the impedance tuning block instead of the switching array. The pulse system was used to provide a broadband excitation in PWT_1 with a pseudorandom bandpass filtered signal from 300 Hz to 6.7 kHz. The acoustic signals needed for the TMM were recorded with M_1 and M_2. A fast Fourier transform (FFT) was performed (1600 spectral lines, 3.5 kHz center frequency, 6.4 kHz span, and 300 ensemble averages) on the acoustic signals.

The normal specific acoustic impedance of the EMHR for the control experiment with the four loading conditions is shown in Fig. 13(a). The resistive loads effectively change the acoustic impedance of the piezoelectric backplate and thus shift the resonant frequencies of the EMHR. However, as shown in Fig. 13(a), when the EMHR was attached with the resistive loads, the tuning ranges of the resonant frequencies are restricted between the short- and open-circuit cases. As the resistance is increased, the resonant frequencies shift toward the open-circuit case as expected. By varying the electrical resistance attached to the EMHR, the control experiment shows that the second resonant frequency of the EMHR can be tuned to 212 Hz, from 2324 Hz (short-circuit

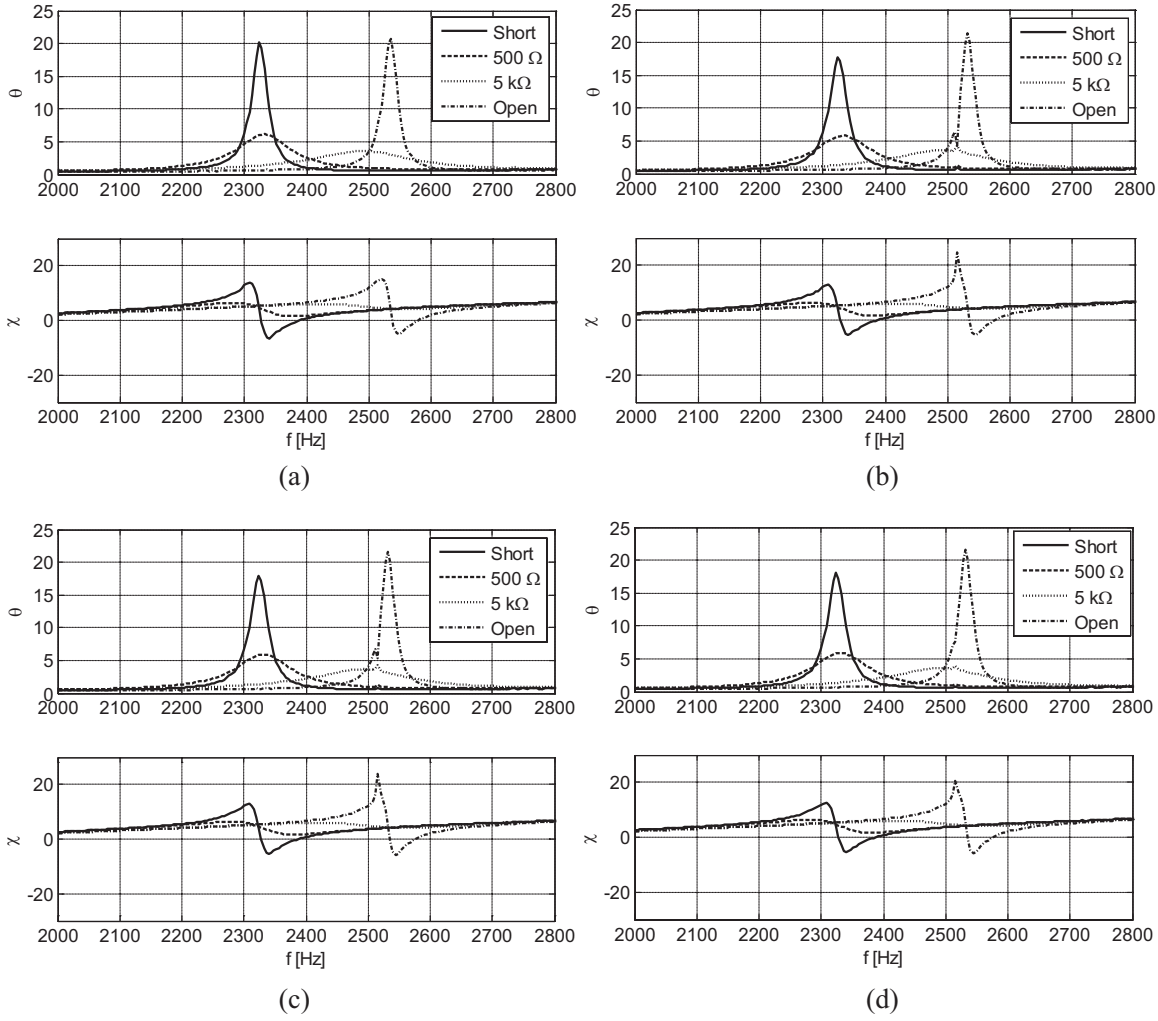


FIG. 13. Acoustic impedance of the EMHR for the (a) reference case, (b) positive battery current test case, (c) zero battery current test case, and (d) negative battery current test case.

TABLE IV. SPL and battery current from the active liner experiments.

	Case I	Case II	Case III
SPL (dB)	153.3	151.8	148.3
Current to the battery (μA)	680	10	−620

case) to 2536 Hz (open-circuit case). Moreover, resistive loads reduce the amplitude of the impedance peaks by increasing the system damping.

With the control experiment as a reference case, the acoustically powered switching array was used to verify the operation of the self-powered active liner system using wireless tuning. The complete liner system was tested under three different scenarios to demonstrate operation with different amounts of available energy. The three test cases were (I) harvested energy greater than required, and energy is stored, (II) harvested energy equal to required, and no energy is stored, and (III) harvested energy less than required, and energy must be provided by the storage element. Note that case III assumes that at some previous time energy was harvested and delivered to the storage element. For each of the three scenarios, the amount of available energy was regulated by adjusting the sound pressure level (SPL) of the single frequency tone inside PWT_2 and monitoring the ammeter in series with the battery. Positive current indicated that energy was flowing into the battery and that the harvested energy was greater than the requirements of the liner circuitry. Negative current indicated that the harvested energy was insufficient to power the liner, and stored energy must be used. Zero current occurred when the amounts of harvested energy and required energy were equal.

As with the reference case, the TMM was used to find the normal specific acoustic impedance of the EMHR for each of the test cases I–III, measured under the different resistive loading conditions (open circuit, short circuit, 500 Ω, and 5 kΩ). The same FFT settings were employed for consistency. The plots of the normal specific acoustic impedance of the EMHR for cases I–III are shown in Figs. 13(b)–13(d), respectively. The SPL needed to achieve each of these cases, as well as the battery current levels, is listed in Table IV.

Examination of the experimental results in Fig. 13 shows good agreement between the reference system and the three wirelessly tuned test cases. For all three test cases, the shift in the second resonant frequency of the EMHR was approximately 208 Hz from short- to open-circuit conditions, and is very close to the 212 Hz exhibited by the reference system. As shown in Fig. 13, the amplitude of the peak of the specific acoustic resistance of the EMHR with the short-circuit load was lower for these three test cases than for the reference. This discrepancy can be attributed to the finite resistance of the switches in the switching array of the impedance tuning block. Since the switch resistance is small (on the order of ohms), it was negligible for the other loads. On the other hand, the amplitude of the peak of the specific acoustic resistance of the open-circuited EMHR was systematically higher for the test cases, which employed the switching array circuitry. The exact reason for this phenomenon is not yet clear.

Examination of the specific acoustic reactance for all three of the test cases shows a small peak at 2.516 kHz, which is the frequency of the single tone used for energy harvesting in PWT_2. However, no peak is present in the reference case. One possible explanation for this peak is coupling of the acoustic signals between the PWTs. In order to power the liner electronics, the single tone signal used for energy harvesting required a relatively high SPL, on the order of 150 dB. It is possible that this high energy signal coupled with the broadband signals used for the TMM measurements in the neighboring tube and caused the peaks. This explanation seems reasonable since PWT_2 was off during the reference experiment, and the magnitude of the peaks decreased as the SPL in PWT_2 decreased for the test cases.

V. CONCLUSIONS

For the first time, an integrated, wireless, self-powered, active liner system has been demonstrated using EMHR elements for both acoustic impedance tuning and energy harvesting. Wireless operation of the integrated system was verified by demonstrating the ability to tune the liner using only acoustic inputs and wireless control signals.

The principles used to demonstrate the liner operation using only two EMHR elements can be expanded to encompass an array of elements. If more power is required or lower SPL is present, more than one EMHR can be used to power the liner. Each group of EMHR for power generation and tuning can be generalized as a module that is operated by a "master" controller. In this manner, the local liner impedance can be adjusted using an array of such modules. The independent operation of each module of the array improves the system robustness and reliability.

Finally, it should be emphasized that since this demonstration system was designed using mostly off-the-shelf components, future work should investigate the design of a custom receiver module and power converter to minimize both power demand and size.

ACKNOWLEDGMENTS

Financial support for this project is provided by the NASA Langley Research Center (Grant No. NAG-1-2261), monitored by Mr. Michael G. Jones. The authors gratefully acknowledge the contributions of Dr. Khai Ngo, Selvi Kadirvel, and Robert Taylor in helpful discussions and suggestions during this work.

[1]R. E. Motsinger and R. E. Kraft, "Design and performance of duct acoustic treatment," in *Aeroacoustics of Flight Vehicles: Theory and Practice Volume 2*, edited by H. H. Hubbard (Acoustical Society of America, New York, 1995), pp. 165–206.
[2]Federal Aviation Administration, *Noise Levels for U.S. Certified and Foreign Vehicles* (Federal Aviation Administration, Washington, DC, 2001).
[3]D. Guicking and E. Lorenz, "An active sound absorber with porous plate," Trans. ASME, J. Vib., Acoust., Stress, Reliab. Des. **106**, 389–392 (1984).
[4]S. Beyene and R. A. Burdisso, "A new hybrid passive/active noise absorption system," J. Acoust. Soc. Am. **101**, 1512–1515 (1997).
[5]J. M. deBedout, M. A. Franchek, R. J. Bernhard, and L. Mongeau,

"Adaptive-passive noise control with self-tuning Helmholtz resonators," J. Sound Vib. **202**, 109–123 (1997).
[6]X. D. Jing and X. F. Sun, "Experimental investigations of perforated liners with bias flow," J. Acoust. Soc. Am. **106**, 2436–2441 (1999).
[7]S. Horowitz, T. Nishida, L. Cattafesta, and M. Sheplak, "Characterization of compliant-backplate Helmholtz resonators for an electromechanical acoustic liner," Int. J. Aeroacoust. **1**, 183–205 (2002).
[8]F. Liu, S. Horowitz, T. Nishida, L. Cattafesta, and M. Sheplak, "A multiple degree of freedom electromechanical Helmholtz resonator," J. Acoust. Soc. Am. **122**, 291–301 (2007).
[9]N. E. duToit, B. L. Wardle, and S. G. Kim, "Design considerations for MEMS-scale piezoelectric mechanical vibration energy harvesters," Integr. Ferroelectr. **71**, 121–160 (2005).
[10]M. Ferrari, V. Ferrari, D. Marioli, and A. Taroni, "Modeling, fabrication and performance measurements of a piezoelectric energy converter for power harvesting in autonomous microsystems," IEEE Trans. Instrum. Meas. **55**, 2096–2101 (2006).
[11]J. Jun, B. Chou, J. Lin, A. Phipps, X. Shengwen, K. Ngo, D. Johnson, A. Kasyap, T. Nishida, H. T. Wang, B. S. Kang, F. Ren, L. C. Tien, P. W. Sadik, D. P. Norton, L. F. Voss, and S. J. Pearton, "A hydrogen leakage detection system using self-powered wireless hydrogen sensor nodes," Solid-State Electron. **51**, 1018–1022 (2007).
[12]F. Liu, A. Phipps, S. Horowitz, T. Nishida, L. Cattafesta, and M. Sheplak, "Acoustic energy harvesting using an electromechanical Helmholtz resonator," J. Acoust. Soc. Am. **123**, 1983–1990 (2008).
[13]S. Kadirvel, F. Liu, S. Horowitz, T. Nishida, K. Ngo, L. Cattafesta, and M. Sheplak, "A self-powered wireless active acoustic liner," presented at the 12th AIAA/CEAS Aeroacoustics Conference, Cambridge, MA, 2006.
[14]R. A. Mangiarotty, "Acoustic-lining concepts and materials for engine ducts," J. Acoust. Soc. Am. **48**, 783–794 (1970).
[15]J. Chen and K. D. T. Ngo, "Alternate forms of the PWM switch model in discontinuous conduction mode," IEEE Trans. Aerosp. Electron. Syst. **37**, 754–758 (2001).
[16]"Impedance and absorption of acoustical materials using a tube, two microphones, and a digital frequency analysis system," ASTM-E1050-98, ASTM International, 1998.
[17]T. Schultz, M. Sheplak, and L. N. Cattafesta, "Uncertainty analysis of the two-microphone method," J. Sound Vib. **304**, 91–109 (2007).
[18]R. J. Taylor, "Optimization of a discontinuous conduction mode flyback for acoustical energy harvesting," MS thesis, University of Florida, Gainesvile, FL (2004).
[19]M. J. T. Smith, *Aircraft Noise* (Cambridge University Press, Cambridge, 1989).

Active acoustical impedance using distributed electrodynamical transducers

M. Collet,[a] P. David, and M. Berthillier
FEMTO-ST, DMA, CNRS UMR 6604, University of Franche-Comté, 24 Chemin de l'Épitaphe, 25000 Besançon, France

(Received 5 January 2007; revised 30 May 2008; accepted 23 October 2008)

New miniaturization and integration capabilities available from emerging microelectromechanical system (MEMS) technology will allow silicon-based artificial skins involving thousands of elementary actuators to be developed in the near future. SMART structures combining large arrays of elementary motion pixels coated with macroscopic components are thus being studied so that fundamental properties such as shape, stiffness, and even reflectivity of light and sound could be dynamically adjusted. This paper investigates the acoustic impedance capabilities of a set of distributed transducers connected with a suitable controlling strategy. Research in this domain aims at designing integrated active interfaces with a desired acoustical impedance for reaching an appropriate global acoustical behavior. This generic problem is intrinsically connected with the control of multiphysical systems based on partial differential equations (PDEs) and with the notion of multiscaled physics when a dense array of electromechanical systems (or MEMS) is considered. By using specific techniques based on PDE control theory, a simple boundary control equation capable of annihilating the wave reflections has been built. The obtained strategy is also discretized as a low order time-space operator for experimental implementation by using a dense network of interlaced microphones and loudspeakers. The resulting quasicollocated architecture guarantees robustness and stability margins. This paper aims at showing how a well controlled semidistributed active skin can substantially modify the sound transmissibility or reflectivity of the corresponding homogeneous passive interface. In Sec. IV, numerical and experimental results demonstrate the capabilities of such a method for controlling sound propagation in ducts. Finally, in Sec. V, an energy-based comparison with a classical open-loop strategy underlines the system's efficiency.
 [DOI: 10.1121/1.3026329]

PACS number(s): 43.50.Ki [KAC] Pages: 882–894

I. INTRODUCTION

Generally, noise reduction techniques aim at modifying the acoustic properties of a medium, or of its interfaces, in order to reduce a sound field on the observer's side. This proves to be a very complex problem that is often dealt with using passive techniques, such as using sound absorbing and attenuating materials,[1,2] silencers,[3] vibration isolators,[4] damping treatments,[5] conventional mufflers[6] like the ones used on today's automobiles, or placing barriers between the noise source and observer. Such methods offer favorable results at middle and high frequencies but in the low frequency band, if the realization is even technically conceivable, they become cumbersome, space intrusive, and expensive.

The rapid development of modern digital computers in the 1970s and 1980s has enabled very large improvements of active noise control (ANC) and especially its practical implementations. ANC technologies still remain widely studied at universities and laboratories with hundreds of research papers published every year.[7–11] In contrast to passive systems, an ANC system may be much smaller and lighter even for low frequency applications but, at high frequencies, it tends to be less efficient and unstable. A suitable combination of both methods allows us to achieve an interesting compromise between dimensions, weight, and efficiency of the system over a wide frequency range.[12,13]

As the noise is often produced by vibrating structures, mechanical sources (shakers, piezoceramic patches, etc.)[14,15] can also be used as actuators rather than sound sources (speakers) when the noise is airborne. Noise control and vibration control are thus closely related but reducing vibration does not necessarily lead to a large reduction of the noise level.

Common ANC strategies, comprising *local* and *global* approaches generally using some kind of adaptive feedback or feedforward algorithm, were enriched in recent years by active acoustical impedance control.[16,17] While local control methods consist basically in pressure suppression at particular locations predefined by the placement of error microphones, global methods deal with the reduction of overall acoustic energy or radiated power of a primary source.[11] Both previous strategies act directly on the acoustical medium by using secondary sources while the impedance control methodology tends to adjust the acoustical properties of boundary interfaces for inducing suitable sound diffusion. However, the implementation of impedance control appears to be more complex due to the necessity of knowing not only pressure, but also the corresponding velocity.

[a]Electronic mail: manuel.collet@univ-fcomte.fr

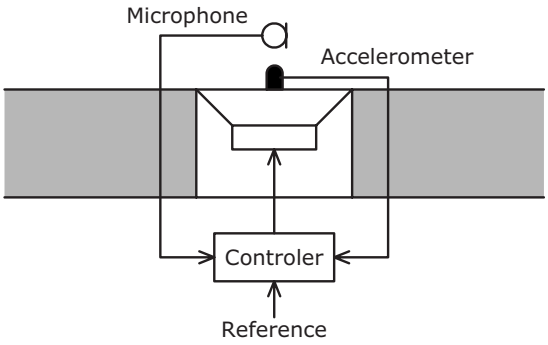


FIG. 1. Direct impedance control.

Aside from the direct control of the acoustic impedance[18–20] via the simultaneous processing of a loudspeaker's membrane velocity and acoustic pressure in its close vicinity[16] (Fig. 1), other methods use Darcy's law to induce a suitable impedance by controlling the acoustic pressure at the rear face of a well designed porous layer[16,21] (Fig. 2). This setup represents an active equivalent of a well known $\lambda/4$ resonance absorber to improve the low frequency absorption of a thin sample of porous material. The idea, first proposed by Olson and May,[22] is to maintain low impedance at the back face of the passive interface by using an active noise control system. This method was experimentally investigated in an impedance tube by Guicking and Lorentz[23] who also pointed out the advantages of the active control solution, compared to the conventional purely passive methods.

The classical approaches used by many authors studying the implementation of active acoustic impedance based on Guicking's work, such as Galland *et al.*[24] for example, are based on the knowledge of the so-called optimal complex impedance, which they tend to implement experimentally using hybrid parietal ANC systems. Their purpose is to locally control a boundary interface to follow the pre-established frequency-dependent complex impedance. Indeed, the total absorption of normal incident waves is described by the simple impedance value $Z=\rho_0 c_0$, where ρ_0 is the density of air and c_0 is the speed of sound in air. However, since the acoustic waves are not always normal incident, this simple absorption condition has to be modified. The proposed solutions for impedance optimization[2,24] lead to theoretical frequency-dependent impedances $Z(\mathrm{j}\omega)$ guaranteeing efficient sound absorption and depending on the disturbances, boundary conditions, speed of sound, frequency band, structural interfaces, and geometry. These dependencies prohibit broad frequency band applications and good efficiency is restricted to a specific configuration.

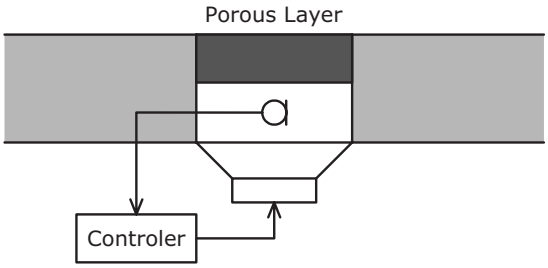


FIG. 2. Impedance control with a porous layer.

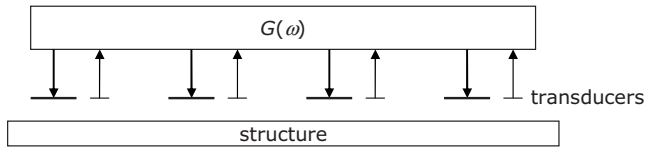


FIG. 3. Centralized control system.

The optimally designed impedance[2,24] appears as a frequency-dependent function $Z(\mathrm{j}\omega)$, linking acoustic pressure and velocity. While its experimental implementation allows to control these parietal quantities by using only the corresponding time operator, it appears very interesting to introduce also a wave number dependency so that $Z=Z(\mathrm{j}\omega,\mathrm{j}\boldsymbol{k})$ in order to improve efficiency and versatility.[25–27] The real-time implementation of such an optimal boundary control involves a complex partial differential operator with not only time but also space derivatives.[28] This approach requires a large amount of system resources (memory and processing speed) and therefore the use of a classical centralized controller (Fig. 3) is not suitable for practical applications.

New methods are now available to deal with the limitations of the centralized system. They allow active transducers and their driving electronics to be directly integrated into otherwise passive structures. The number of potential applications for these approaches is growing significantly in many industrial fields and the main research challenge today deals with the development of new multifunctional structures integrating electromechanical systems in order to optimize their intrinsic static or load-bearing mechanical behavior while also achieving goals specific to their dynamic response.[29,30]

In the quest for this new technology, let us consider the schematic representation of the systems in Figs. 4 and 5. They represent a fluid-structure interface covered by a network of sensors and actuators that are connected by an *ad hoc* electronic circuit in two different configurations. Figure 4 represents a fully distributed system where one sensor is used to control only one actuator and allows the implementation of a unique temporal operator, that is to say, a *standard acoustical impedance* $Z(\mathrm{j}\omega)$. Figure 5 describes a first order system where signals from two neighboring sensors are used to control one actuator. This configuration allows implementation of a complete *generalized acoustical impedance* $Z(\mathrm{j}\omega,\mathrm{j}\boldsymbol{k})$, but involving only first order derivative in space. These two examples underline the inherent dependency of the theoretical control operator and the technological architecture of the physical implementation. This duality

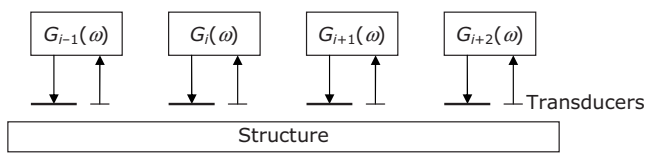


FIG. 4. Completely distributed system (order 0).

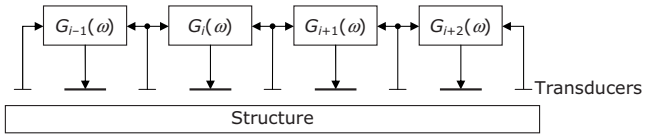


FIG. 5. Distributed quasicollocated system of the first order.

stands as a major constraint in the proposed methodology for designing such distributed active acoustical impedance.

Section II presents the generic methodology for computing a generalized acoustical impedance operator for a total absorption of two-dimensional (2D) acoustic waves along a straight line interface.[25] The obtained complex operator contains space and time pseudodifferential terms that prevent any simple experimental implementation. Section III is dedicated to the synthesis of the control law for limiting the acoustic wave propagation in a plane-wave tube by using a simplified version of the previous methodology in order to obtain a standard low order time-space impedance operator. After establishing the physical dimensions, in Sec. IV, a numerical study of the coupled system properties as well as some experimental results are also presented. Finally, in Sec. V, an energy-based comparison with an open-loop algorithm is performed.

II. TOTAL ABSORPTION OF ACOUSTIC WAVES

A. Introduction

The absorption boundary condition for the total acoustic energy corresponds to the nonreflection of all incident waves, whatever their orientations with respect to the active interface. From within the considered acoustic domain, all waves propagate as if the domain was infinite. Recently, significant efforts have been made in the characterization of this absorption condition; on one hand, to actually introduce this type of condition for accurately computing numerical solutions of the open field problem and, on the other hand, for controlling acoustic waves in a domain bounded by active interfaces. Instead of presenting the formal definition of such absorbing boundary operator (generalized impedance operator), we introduce this notion and consider the total absorption of acoustic energy contained in a half-plane by optimizing the boundary impedance. This example allows us to present a new methodology for designing interface control strategy. The total absorption of the acoustic waves by an active boundary in two dimensions does not represent a significant practical interest. It is appropriate to underline that the proposed method can be extrapolated on much more complex systems. Section IV presents, however, one realistic application of this concept.

B. The control problem

Let us consider the system represented in Fig. 6. The acoustic energy is located on the left half-plane $\Omega_L=\mathbb{R}_x^- \times \mathbb{R}_y$. A pressure $p(y,t)=p_\Gamma$ is imposed on the control surface $\Gamma=\delta(x)\times\mathbb{R}_y$. It is assumed that we measure the normal derivative of acoustic pressure $w(y,t)=\partial_x p_\Gamma$. The system is then described by the wave equation

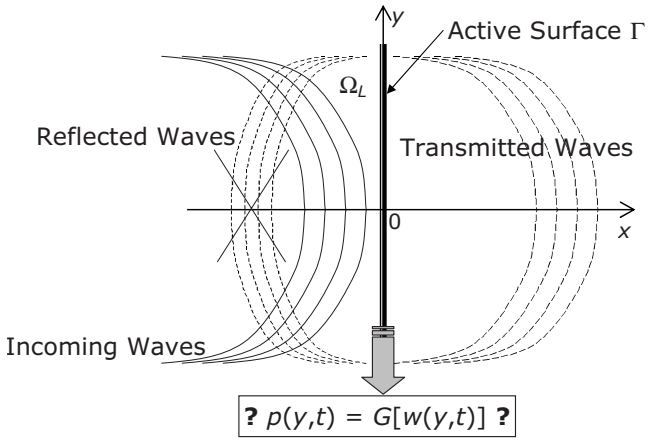


FIG. 6. Active surface in the system of acoustic waves.

$$\frac{1}{c_0^2}\frac{\partial^2 p}{\partial t^2}-\Delta p=0 \quad \text{on } \Omega_L\times\mathbb{R}_t^{*+},$$

$$p(0,y,t)=p_\Gamma,$$

$$w(y,t)=\frac{\partial p(0,y,t)}{\partial x}. \tag{1}$$

C. Definition of the control

We seek to define a relation $p=G(w)$ such that the acoustic waves on the boundary $x=0$ are totally absorbed by the active interface. This type of control is asymptotically stable because the energy inside the acoustic domain can only decrease. The controlled acoustical field is also a solution of the fully open 2D domain and is given by the following system:

$$\frac{1}{c_0^2}\frac{\partial^2 p}{\partial t^2}-\Delta p=0 \quad \text{on } \Omega_L\times\mathbb{R}_t^{*+}$$

$$\frac{1}{c_0^2}\frac{\partial^2 \theta}{\partial t^2}-\Delta \theta=0 \quad \text{on } \mathbb{R}_x^{*+}\times\mathbb{R}_y\times\mathbb{R}_t^{*+},$$

$$p(0,y,t)=\theta(0,y,t),$$

$$\frac{\partial\theta(0,y,t)}{\partial x}=\frac{\partial p(0,y,t)}{\partial x}, \tag{2}$$

with initial conditions $p(x,y,0)=0$ and $\partial_t p(x,y,0)=0$. We can interpret system (2) as the control equations. In this case the control function $p=G(w)$ is completely defined by the system

$$\frac{1}{c_0^2}\frac{\partial^2 \theta}{\partial t^2}-\Delta \theta=0 \quad \text{on } \mathbb{R}_x^{*+}\times\mathbb{R}_y\times\mathbb{R}_t^{*+},$$

$$\theta(0,y,t)=p_\Gamma,$$

$$w(y,t) = \frac{\partial \theta(0,y,t)}{\partial x}. \tag{3}$$

By a simple symmetrization, the preceding equation can be extended on the entire x-axis. We thus obtain the following expression of the control law G:

$$\frac{1}{c_0^2}\frac{\partial^2 \theta}{\partial t^2} - \Delta\theta = -2\delta(x)w(y,t) \quad \text{on } \mathbb{R}_x \times \mathbb{R}_y \times \mathbb{R}_t^{*+},$$

$$p_\Gamma = \theta(0,y,t). \tag{4}$$

By using the Fourier transforms of x, y, and t associated, respectively, with the variables ξ, η, τ, control function (4) can be formally rewritten as

$$\left(-\frac{4\pi^2}{c_0^2}\tau^2 + 4\pi^2\xi^2 + 4\pi^2\eta^2\right)\tilde{\theta} = -2\tilde{w},$$

$$\tilde{p}_\Gamma = \int \tilde{\theta} \mathrm{d}\xi. \tag{5}$$

Then, in the Fourier space, the expression of G becomes

$$\tilde{p}_\Gamma = G(\eta,\tau)\tilde{w} = -\frac{1}{2\pi}\frac{1}{\sqrt{\eta^2 - r^2/c_0^2}}\tilde{w}. \tag{6}$$

The pseudodifferential operator corresponding to control equation (6) is $-1/\sqrt{\partial_y^2 - \partial_t^2}$. The interface impedance relationship implies a total wave absorption. This operator is pseudodifferential, nonlocal in time and space, and also appears difficult to implement on a realistic system. Different mathematical techniques have been introduced for the numerical realization of such pseudodifferential operators as the diffusive representations presented in the work of Matignon *et al.*[27] However, the mathematical details will not be discussed here. A concrete realization of such a generalized impedance in real time through a suitable distributed active system is today unrealistic because of the number and the complexity of necessary calculations. A large part of the research in this domain deals with the implementation of diffusive representations and simplifications that could be carried out.[26,31]

III. TRANSMISSIBILITY CONTROL IN THE TUBE

A. Problem synthesis

The method presented in Sec. II can be directly applied for controlling the wave transmissibility in tube. Indeed, the control equations suggest a system where the tube becomes sound transparent (see Fig. 7 upper right scheme). The energy transmitted through the tube's cross section would also decrease with respect to the length of the tube as described in Fig. 7. The longer the absorbing surface, the weaker the acoustic energy transmitted through the end section of the tube. Another choice, presented in this section, is to construct an impedance condition so that all waves interacting with the active interface are reflected backwards. Figure 7 presents what would be the behavior of a tube equipped with such an active surface. Thus, the considered boundary condition appears as a unidirectional reflecting condition, which yields a greater acoustic transfer attenuation as described in Fig. 7.

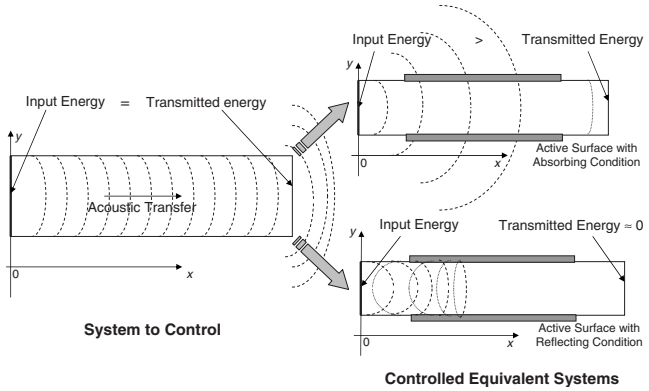


FIG. 7. Absorbing and reflecting boundary conditions.

To obtain such a condition, the methodology described in Sec. II can be directly applied. As the main aim of this work is to design a realistic distributed system that could be physically implemented, two remarks must be made in order to account for some basic physical constraints. Firstly, if one considers classical electrodynamic acoustic transducers, their operating mode does not directly generate parietal acoustic pressure but, in a given frequency band, a gradient of pressure; so it is important to adapt the method for using this type of boundary condition. The well known operating mode of such acoustic transducers is detailed in Sec. IV B. Moreover, it is judicious to conceive a strategy based on a low order differential operator that could be implemented by using distributed active cells in a configuration close to those described in Fig. 5. Pseudodifferential or high order differential operators prohibit any realistic experimental real-time realization.

First of all let us consider the problem in two dimensions described by the following equations representing the vibroacoustic system in Fig. 8, where $u(x,t)$ is the control parameter and $m(x,t)$ is the measured variable:

$$\frac{1}{c_0^2}\frac{\partial^2 p}{\partial t^2} - \Delta p = 0 \quad \text{on } \mathbb{R}_x \times \mathbb{R}_y^{-*} \times R_t^{*+},$$

$$\frac{\partial p(x,0,t)}{\partial y} = u(x,t),$$

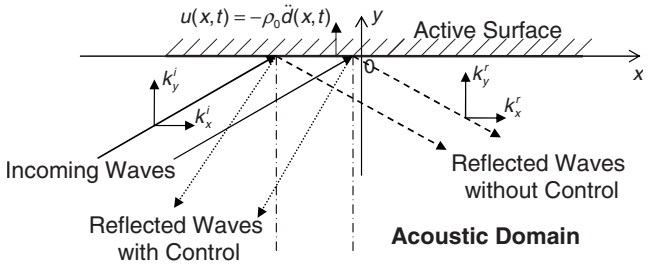


FIG. 8. Acoustic waves interacting with the active surface.

$$m(x,t) = p(x,0,t). \tag{7}$$

As indicated in Fig. 8, the control variable $u(x,t)$ is proportional to the normal acceleration $\ddot{d}(x,t)$ of the active surface. The objective is then to induce a reflection of incidental waves so that the projection of the reflected wave numbers k_x^i is negative. Thus, all acoustic energy intercepted by the active surface turns over to the direction of negative x in order to cancel the acoustic power flow in the direction of positive x. Such a parietal behavior is described in Fig. 8. This 2D condition can be directly extended for controlling acoustic power flow in a tube.

B. Control strategy

The control strategy is given by the following, which represents an advection equation (or a transport equation) directed to the negative x:

$$u(x,t) = -\left(\frac{1}{c_a}\frac{\partial p(x,0,t)}{\partial t} - \frac{\partial p(x,0,t)}{\partial x}\right), \tag{8}$$

where c_a represents the transportation celerity, which stands for the control parameter. This nontraditional approach can be related to the classical results concerning fluid/structure coupling and, in particular, the calculation of acoustic concordance frequencies.[32] We seek to impose a solid-state transport with only one direction of propagation. This direction is obviously contrary to that for which we want to minimize the transmittance. This coupling induces only evanescent waves propagating toward the positive x direction thereby avoiding the existence of any concordance frequency.

By combining Eqs. (7) and (8), the controlled behavior is governed by the following equations:

$$\begin{aligned}
&\frac{1}{c_0^2}\frac{\partial^2 p}{\partial t^2} - \Delta p = 0 \quad \text{on } \mathbb{R}_x \times \mathbb{R}_y^{-*} \times \mathbb{R}_t^{*+},\\
&\frac{\partial p(x,0,t)}{\partial y} = \frac{\partial m(x,t)}{\partial x} - \frac{1}{c_a}\frac{\partial m(x,t)}{\partial t},\\
&m(x,t) = p(x,0,t),\\
&+ \text{initial conditions}.
\end{aligned} \tag{9}$$

By symmetrization of problem (9), we obtain equivalent system (10) defined over $\mathbb{R}^2 \times \mathbb{R}_t^{*+}$ as follows:

$$\begin{aligned}
&\frac{1}{c_0^2}\frac{\partial^2 p}{\partial t^2} - \Delta p = 2\delta(y)\left(\frac{1}{c_a}\frac{\partial m(x,t)}{\partial t} - \frac{\partial m(x,t)}{\partial x}\right)\\
&\qquad \text{on } \mathbb{R}_x \times \mathbb{R}_y \times \mathbb{R}_t^{*+},\\
&m(x,t) = p(x,0,t),\\
&+ \text{initial conditions}.
\end{aligned} \tag{10}$$

Using the Fourier transforms into x, y, and t, respectively, associated with Fourier variables ξ, η, and τ, the controlled pressure, solution of system (10), can be formally rewritten as

$$\begin{aligned}
&\left(-\frac{4\pi^2}{c_0^2}\tau^2 + 4\pi^2\xi^2 + 4\pi^2\eta^2\right)\tilde{P}(\xi,\eta,\tau)\\
&\quad = 2\left(\frac{2\pi \mathrm{j}}{c_a}\tau - 2\pi \mathrm{j}\xi\right)\tilde{m}(\xi,\tau),
\end{aligned} \tag{11}$$

where $\tilde{m}(\xi,\tau)$ is the Fourier transform of $p(x,0,t)\delta(y)$, so that

$$\tilde{m}(\xi,\tau) = \int_{-\infty}^{\infty} \tilde{P}(\xi,\eta,\tau)\mathrm{d}\eta. \tag{12}$$

The second part of Eq. (11) is a function only of variables ξ and τ, so by using Eqs. (12) and (11), we obtain

$$\begin{aligned}
\tilde{m}(\xi,\tau) &= \tilde{m}(\xi,\tau)\frac{1}{\pi}\int_{-\infty}^{\infty} \frac{\mathrm{j}\tau/c_a - \mathrm{j}\xi}{-\tau^2/c_0^2 + \xi^2 + \eta^2}\mathrm{d}\eta\\
&= \tilde{m}(\xi,\tau)\frac{\mathrm{j}(\tau/c_a - \xi)}{\sqrt{\xi^2 - \tau^2/c_0^2}}.
\end{aligned} \tag{13}$$

The function $\tilde{m}(\xi,\tau)$, the Fourier transform of the acoustic pressure restriction $p(x,0,t)$ on x-axis, is nonzero if and only if

$$\mathrm{j}\left(\frac{\tau}{c_a} - \xi\right) = \sqrt{\xi^2 - \frac{\tau^2}{c_0^2}}. \tag{14}$$

Equation (14) can be immediately reduced to the sufficient condition

$$2\xi^2 - \frac{2\xi\tau}{c_a} + \left(\frac{1}{c_a^2} - \frac{1}{c_0^2}\right)\tau^2 = 0. \tag{15}$$

The values of ξ are the roots of the second order polynomial (15). Its determinant is $\Delta = 4\tau^2(2/c_0^2 - 1/c_a^2)$, so three cases exist.

- *Case 1*. The determinant $\Delta > 0$ so $c_a^2 > c_0^2/2$, giving

$$\xi_{1,2} = \frac{\tau}{2c_a} \pm \frac{\tau}{2}\sqrt{\frac{2}{c_0^2} - \frac{1}{c_a^2}}.$$

  If $c_0^2 \geq c_a^2$, the roots are positive and the obtained waves propagate along $(\mathbf{0}\boldsymbol{x})$ with positive wave number, that is to say toward negative x (from $+\infty$ to $-\infty$). If $c_0^2 < c_a^2$, two real waves exist whereas one propagates from $+\infty$ to $-\infty$ and the other from $-\infty$ to $+\infty$.
- *Case 2*. The determinant $\Delta = 0$ so $c_a^2 = c_0^2/2$, giving

$$\xi = \frac{\tau}{2c_a}.$$

  There are two waves with the same wave number, which propagate from $+\infty$ to $-\infty$ at the same speed.
- *Case 3*. The determinant $\Delta < 0$ so $c_a^2 < c_0^2/2$, giving

$$\xi_{1,2} = \frac{\tau}{2c_a} \pm \mathrm{j}\frac{\tau}{2}\sqrt{\frac{1}{c_a^2} - \frac{2}{c_0^2}}.$$

  There are two complex waves (one attenuated and one divergent), which propagate from $+\infty$ to $-\infty$. Along the $(\mathbf{0}\boldsymbol{x})$ we have

$$m(x,t) = m_0 e^{\pm\pi/2\sqrt{1/c_a^2 - (1/c_0^2)}x} e^{j(\omega t + (\pi/2c_a)x)}.$$

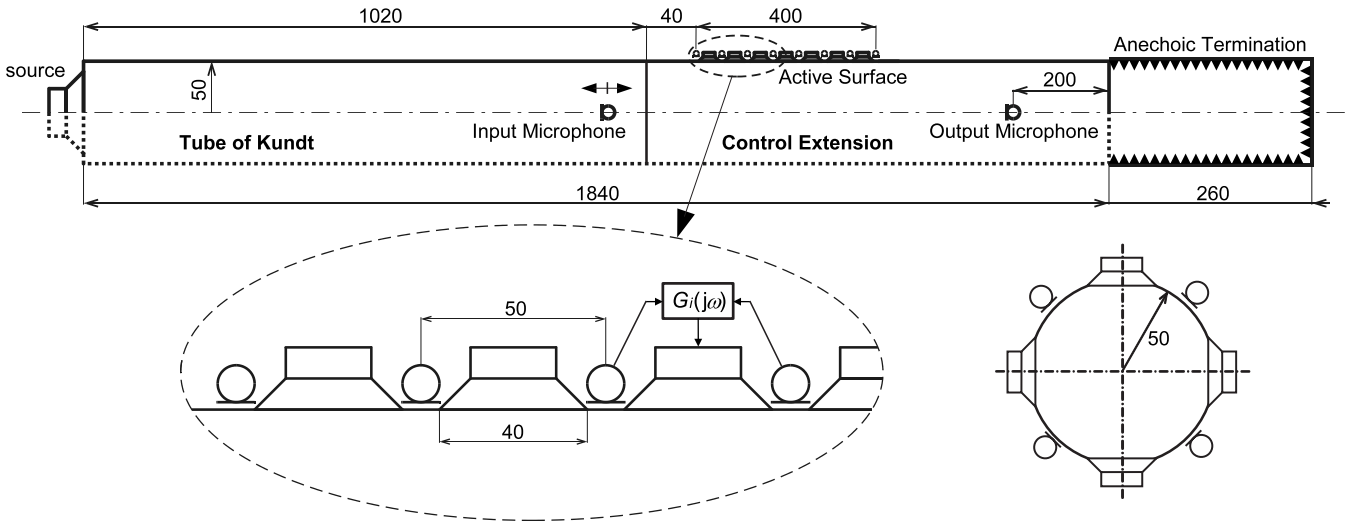


FIG. 9. Modeled vibroacoustic system.

One of these waves is unstable so we have to avoid using a control parameter c_a such that $c_a^2 < c_0^2/2$.

Using the Fourier transform of Eq. (8) and the expression of the control signal $u(x,t)$ given in Eq. (7), we obtain $\eta = \xi - \tau/c_a$ so one can find the values of η, the transversal component of wave numbers.

- *Case 1.* $c_a^2 > c_0^2/2$

$$\xi_{1,2} = \frac{\tau}{2c_a} \pm \frac{\tau}{2}\sqrt{\frac{2}{c_0^2} - \frac{1}{c_a^2}}$$

$$\Rightarrow \eta_{1,2} = -\frac{\tau}{2c_a} \pm \frac{\tau}{2}\sqrt{\frac{2}{c_0^2} - \frac{1}{c_a^2}}.$$

For $c_0^2 \geq c_a^2$ there are waves propagating from $-\infty$ to $+\infty$ on $(\mathbf{0}\boldsymbol{y})$ with total absorption on the surface. When $c_0^2 < c_a^2$, we get one reflected wave.

- *Case 2.* $c_a^2 = c_0^2/2$

$$\xi = \frac{\tau}{2c_a} \Rightarrow \eta = -\frac{\tau}{2c_a}.$$

- *Case 3.* $c_a^2 < c_0^2/2$

$$\xi_{1,2} = \frac{\tau}{2c_a} \pm \mathrm{j}\frac{\tau}{2}\sqrt{\frac{1}{c_a^2} - \frac{2}{c_0^2}}$$

$$\Rightarrow \eta_{1,2} = -\frac{\tau}{2c_a} \pm \mathrm{j}\frac{\tau}{2}\sqrt{\frac{1}{c_a^2} - \frac{2}{c_0^2}}.$$

Here the waves propagate from $-\infty$ to $+\infty$ on $(\mathbf{0}\boldsymbol{y})$. As one is situated on $y < 0$, the propagation along $(\mathbf{0}\boldsymbol{y})$ is amplitude decreasing when $y \to -\infty$.

One can conclude that, if we choose the control parameter c_a such that $c_a^2 \geq c_0^2/2$, we obtain waves with $\xi > 0$ whatever the frequency τ and the incidence angle. An acoustic field being propagated only in the decreasing x direction can thus be obtained, which means that there will be no waves transmitted to the right side of the domain. Our objective is therefore achieved and we underline here that the existence of the wave admitting a Fourier transform in all $\mathbb{R}^2 \times \mathbb{R}_t^{*+}$ is based on a formal demonstration. In fact, only the waves established on interaction with the active surface will have the stated properties. Schematically (Fig. 8) all waves intercepted by the control boundary will be reflected to the left of the normal axis at the incidence point.

Thus, by using the control strategy defined by Eqs. (8) and (7), we have designed a semi-infinite system where the acoustic energy is only transferred toward the negative x. The obtained active skin behavior also yields a very promising generalized impedance that could be used for acoustic power flow control. The design interface behavior is made by using standard first order differential operator in time and space that can be implemented by a distributed set of unit cell in configuration depicted by Fig. 5.

IV. NUMERICAL AND EXPERIMENTAL QUALIFICATIONS OF THE METHOD: APPLICATION TO THE TRANSMISSIBILITY CONTROL IN A TUBE

A. Description of the studied system

The acoustic system shown in Fig. 9 consists of a plane-wave tube of length $L = 1.84$ m and radius $R = 0.05$ m. A loudspeaker is placed on the left side of the tube to inject a sound disturbance into the system. At the distance $L_t = 1.06$ m begins an active zone of length $L_a = 0.4$ m composed alternately of seven circumferences of actuators and eight circumferences of sensors. These transducer channels are, in fact, implemented by a parallel distribution of four loudspeakers and four microphones. The measurements of pressure are then averaged by circumferential line and the control voltage applied to loudspeakers is injected in parallel on all four elements of each circumference. We thus preserve a quasiaxisymmetric architecture able to spatially filter waves with at least one nodal diameter. These waves will be quasi-insensitive to the control signals and quasinonmeasurable by the network of sensors. This principle of modal filtering is often used in structure dynamics to limit residual effects such as *spillover*. We can also adopt a completely

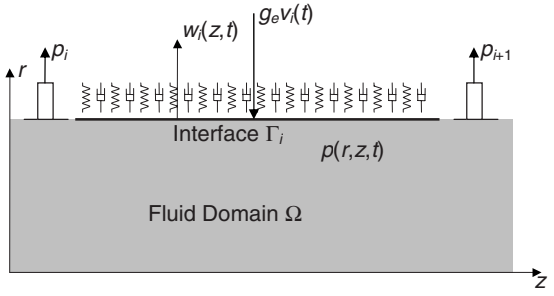


FIG. 10. Model of the control transducer.

axisymmetric model, composed of only one-half of the system, which allows a significant reduction in the calculation time.

The implemented control interconnection is presented in the introduction in Fig. 5. Each row of circumferential loudspeakers is controlled by a uniquely applied voltage computed by using two measurements picked up by the two adjacent rows of microphones. This distribution enables us to implement a discretized version of the continuous control law given by Eq. (8). For the finite element modeling we used COMSOL MULTIPHYSICS™ software and the postprocessing calculations related to the control were carried out in MATLAB®.

B. Modeling the control loudspeakers and fluid-structure interaction

The previously described system is composed of an acoustic medium coupled with a distributed active skin consisting of a set of transducers. As seen in Fig. 10, the measurement part, produced by microphones, does not induce a strong coupling effect and is simply introduced by a corresponding pressure state in the finite element modeling. On the contrary, actuators made of loudspeakers have to be carefully introduced in the model in order to take into account the whole set of strong electromechanical coupling effects. Indeed, such a system presents internal dynamics corresponding first to the pumping mode and second to all internal modes of the loudspeaker's membrane. Our model is based on a finite set of partial differential equations for each loudspeaker interface i such that

$$\bar{\rho}_{\mathrm{sp}}\ddot{d}_i + c_{\mathrm{sp}}\dot{d}_i + k_{\mathrm{sp}}d_i + T_{\mathrm{sp}}\Delta d_i = pl_{\mathrm{sp}} + g_e v_i. \tag{16}$$

This model corresponds to a cord of linear mass distribution $\bar{\rho}_{\mathrm{sp}}=900\times10^{-3}$ kg m^{-1} under a tension of $T_{\mathrm{sp}}=20.2\times10^{6}$ N m placed on a linear distribution of stiffness $k_{\mathrm{sp}}=20.2\times10^{6}$ N m^{-2} and damping $c_{\mathrm{sp}}=601$ N s m^{-2}. These mechanical characteristics were determined by updating the characterization curves of the microloudspeakers Monacor SP-11/2R, 8 Ω, 0.2 W. The external forces are induced by the pressure exerted over the length of the actuator pl_{sp} (l_{sp} is the width of the loudspeaker's membrane supposed here to be rectangular) and by the electrodynamic control efforts $g_e v_i$ (g_e represents the static gain of the corresponding electromechanical coupling relation including the RL circuit and the Laplace electromechanical coupling and v_i is the injected voltage).

We note that using a rectangular membrane does not modify the coupled system's behavior as long as we remain at the wavelengths longer than the membrane dimension. The frequency limit of our elementary actuator is given by the relations

$$l_{\mathrm{sp}} = \frac{\lambda}{2}, \quad f_c = \frac{c_0}{\lambda}, \tag{17}$$

so for $l_{\mathrm{sp}}=4$ cm we obtain $f_c \approx 4290$ Hz. This simple model of the control transducer provided the first few eigenfrequencies of the speaker's membrane. The first suspension mode (pumping mode) is situated at frequency 799 Hz and the first bending mode at 4033 Hz. The nominal frequency range of this acoustic generator is thus 800–4000 Hz. Our study concerns the frequencies between 100 and 3000 Hz in order to stay below the acoustic cutoff. The mechanical behavior of the system in this frequency band can be simplified by noting that the first suspension mode dominates in the expression of the membrane acceleration. As long as we work beyond this first internal modal frequency, we can write $d_i(r,t)\approx 1\times d_i^1(t)$ with

$$M_{\mathrm{sp}}\ddot{d}_i^1 = \int_{\Gamma_i} p(z,R,t)l_{\mathrm{sp}}\mathrm{d}z + g_e v_i l_{\mathrm{sp}}. \tag{18}$$

With no externally applied pressure, we thus obtain a system whose acceleration is directly proportional to the control voltage. It is then interesting to notice that we are working within the framework of the control expression given in Sec. III A. Nevertheless, to conserve a wide frequency band modeling of the loudspeaker, we use Eq. (16) in the global partial derivative system of equations driving the whole coupled axisymmetric problem (fluid, structure, and loudspeaker); hence,

$$\frac{r}{c_0^2}\frac{\partial^2 p}{\partial t^2} - \left[\frac{\partial}{\partial r}\left(r\frac{\partial p}{\partial r}\right) + \frac{\partial}{\partial z}\left(r\frac{\partial p}{\partial z}\right)\right] = 0 \quad \text{on } \Omega\times\mathbb{R}_t^{*+},$$

$$\frac{\partial p(0,r,t)}{\partial z} = F_{\mathrm{ext}} \quad \text{on } \Gamma_{\mathrm{in}}\times\mathbb{R}_t^{*+},$$

$$\frac{\partial p(z,R,t)}{\partial r} = -\rho_0\ddot{d}_i(z,t) \quad \text{on } \Gamma_i\times\mathbb{R}_t^{*+},$$

$$\frac{\partial p(z,R,t)}{\partial r} = 0 \quad \text{on } \Gamma_R\times\mathbb{R}_t^{*+},$$

$$\frac{\partial p(L,r,t)}{\partial z} = -\frac{1}{c_0}\frac{\partial p}{\partial t} \quad \text{on } \Gamma_{\mathrm{out}}\times\mathbb{R}_t^{*+},$$

$$\frac{\partial p(z,0,t)}{\partial r} = 0 \quad \text{on } \Gamma_{\mathrm{sym}}\times\mathbb{R}_t^{*+},$$

$$\bar{\rho}_{\mathrm{sp}}\ddot{d}_i + c_{\mathrm{sp}}d_i + k_{\mathrm{sp}}d_i + T_{\mathrm{sp}}\Delta d_i = pl_{\mathrm{sp}} + g_e v_i \quad \text{on } \Gamma_i\times\mathbb{R}_t^{*+}. \tag{19}$$

Γ_{in} is the source boundary on the left side of the tube, Γ_i is the ith boundary corresponding to the ith loudspeaker's

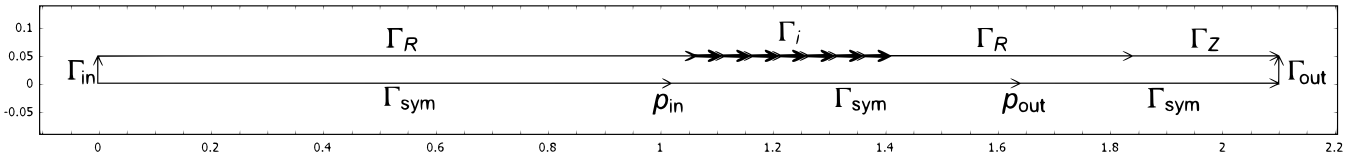


FIG. 11. Acoustic domain boundaries used in the model.

membrane, Γ_R is a rigid wall, Γ_{out} is the Sommerfeld radiation boundary on the right side of the tube, and Γ_{sym} is the symmetry boundary condition.

By assuming a monomodal response of each loudspeaker's membrane as in Eq. (18), with $M_{sp}=\int_{\Gamma_i}\bar{\rho}_{sp}dz$, the continuity equation of normal velocities on the fluid-membrane interface is

$$\frac{\partial p(z,R,t)}{\partial r} \approx -\frac{\rho_0}{M_{sp}}\left(\int_{\Gamma_i} p(z,R,t)l_{sp}dz + g_e v_i l_{sp}\right) \quad \text{on } \Gamma_i \times \mathbb{R}_t^{*+}. \tag{20}$$

There is also a static relation between the normal derivative of the pressure on the membrane interface and the applied control voltage on each of actuators. One can thus directly implement the theoretical control law defined in Sec. II B.

C. Implementation of a discrete form of the control equation

The continuous control equation (8) theoretically developed for a 2D semi-infinite domain can be directly applied to control the acoustical flow in the tube. Also, we need to implement the continuous control law in the discrete form onto the distributed set of acoustic transducers. To do so, we introduce a spatial discretization of the measured pressure in Eq. (8) so that $p(idz,t)=p_i$ and $\partial p(idz,t)/\partial t=\dot{p}_i$ as mentioned in Fig. 10. The spatial discretization depends on the distance between each row of microphones, in our case $dz=0.05$ m. The actuator length (diameter) is $l_{sp}=0.04$ m. We simply apply a first order Euler approximation scheme of the involved spatial derivative operator in Eq. (8) and we obtain the discrete control law

$$\frac{\partial p(z,R,t)}{\partial r} = -\left(\frac{1}{c_a}\frac{\dot{p}_{i+1}+\dot{p}_i}{2} - \frac{p_{i+1}-p_i}{dz}\right) \quad \text{for } z \in [idz,(i+1)dz]. \tag{21}$$

By assuming a monomodal response of loudspeakers as in Eq. (18), one can directly compute the imposed control voltage $v_i(t)$ so that relation (21) is satisfied; hence,

$$v_i(t) = -\frac{l_{sp}}{g_e}\mathrm{F1}\left(\frac{p_{i+1}+p_i}{2}\right) + \frac{M_{sp}}{\rho_0 g_e l_{sp}}\left[\frac{1}{c_a}\mathrm{D}\left(\frac{p_{i+1}+p_i}{2}\right) - \mathrm{F2}\left(\frac{p_{i+1}-p_i}{dz}\right)\right]. \tag{22}$$

F1 and F2 represent simple conditioning filters used for the practical implementation of the control, which is valid within the framework of frequency band assumptions explained previously. Transfer functions of these filters are given by Eqs. (23a) and (23b), where the optimal value of the variable ω_{opt} was identified during the experiments as $\omega_{opt}=2\pi \times 1450\ \mathrm{s}^{-1}$

$$\mathrm{F1}(s) = \frac{\omega_{opt}}{s+\omega_{opt}}, \tag{23a}$$

$$\mathrm{F2}(s) = \frac{\omega_{opt}^2}{(s+\omega_{opt})^2}. \tag{23b}$$

The time derivation of pressure is implemented by a filter with the transfer function

$$\mathrm{D}(s) = \frac{\omega_{der}^2 \omega_{opt} s}{(s+\omega_{der})^2(s+\omega_{opt})}, \tag{24}$$

where $\omega_{der}=2\pi\times 1500\ \mathrm{s}^{-1}$.

As the theoretical control law has been synthesized assuming an infinite distribution of active skin, it was necessary to limit the boundary effects of the finite set of active cells by weighting control signals v_i with a seven-coefficient Hanning window.

D. Numerical results

The system corresponding to the set of equations (19) is modeled with COMSOL MULTIPHYSICS™ software. The obtained LTI model is then exported into MATLAB® where control law (22) is implemented and the frequency response of control system is simulated. We underline here that the obtained system is numerically *stable*. The harmonic calculations are carried out in the frequency range 0–3000 Hz with a 20 Hz step. We present here only the pressure fields corresponding to the frequencies 500, 1000, and 2000 Hz. These selected frequencies correspond, respectively, to the harmonic responses computed (1) below the suspension resonance of the loudspeaker's membrane (800 Hz), (2) in the nominal frequency band of the plane-wave tube where assumption given in Eq. (18) is valid, and (3) in the higher frequency domain where the waves can no longer be considered as plane and where acoustic-mechanical coupling effects between loudspeaker's membranes and acoustic medium occur.

The acoustic domain, as well as partitions of the system boundaries, is presented in Fig. 11. To impose a pressure peak of 120 dB, which corresponds to the acoustic pressure $p_c=20$ Pa, the Neumann condition is applied on the source boundary Γ_{in} such as $F_{ext}=(\omega/c_0)p_c$.

Figures 12–14 show the pressure distributions over the domain with and without control. The control strategy does not induce any noticeable effect on the pressure amplitude at 500 Hz (Fig. 12) because this frequency is below the threshold of correct operation of the control loudspeakers. At the

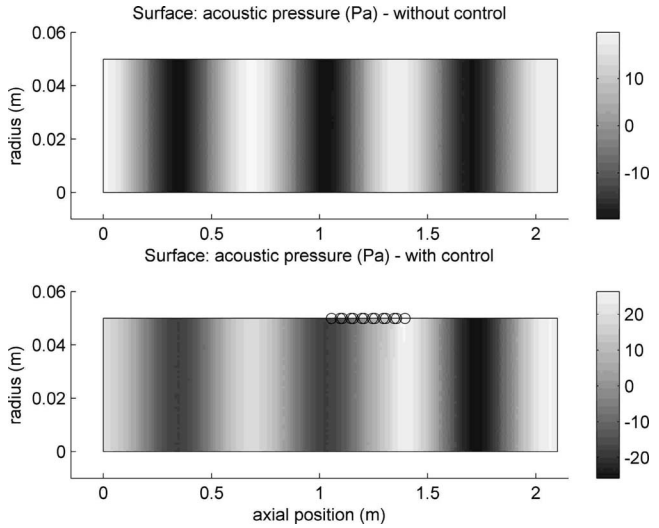


FIG. 12. Controlled and uncontrolled pressure distributions, frequency 500 Hz.

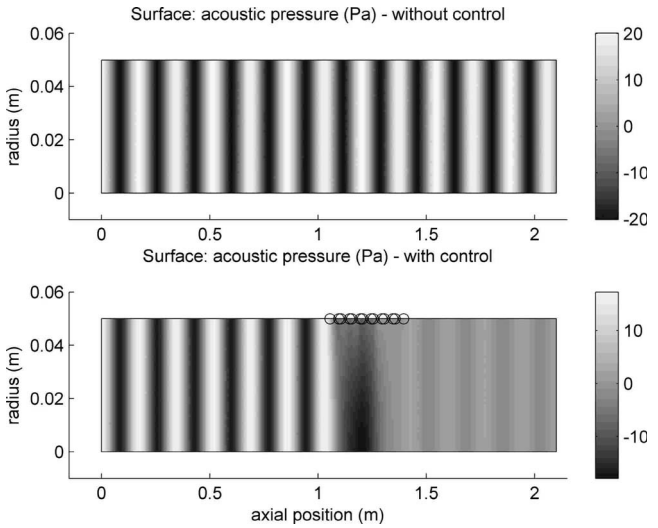


FIG. 14. Controlled and uncontrolled pressure distributions, frequency 2000 Hz.

midrange frequency (Fig. 13), a high decrease in pressure amplitude is clearly observed in front of the active zone, and Fig. 14 corresponding to the high frequency response at 2000 Hz shows a little less effect and a decrease in efficiency for higher frequencies.

In order to quantify the effectiveness of the control, we calculated in the postprocessing the radiated acoustic power through each domain boundary where the energy exchange occurs, i.e., Γ_{in}, Γ_i, and Γ_{out}. For our axisymmetric system, the harmonic expression of per cycle radiated power is given by

$$P_{\text{Ai}} = \frac{1}{2}\int_{\Gamma} \Re\left(r\frac{2\pi i}{\rho_0 \omega}\frac{\partial p}{\partial \vec{n}} p^* \right) \mathrm{d}\Gamma, \tag{25}$$

where $\Re$ indicates a real part of a complex number, $\vec{n}$ is normal vector of the corresponding boundary Γ, and * represents a complex conjugate. The results are presented in Fig. 15. Due to the physical limitations, our active distributed skin is not effective at 200 and 3000 Hz and the acoustic power is radiated through both input and output boundaries. At middle frequencies, where the system operates correctly, the input acoustic energy is attenuated by the control loudspeakers to almost zero at the output.

The Bode diagram in Fig. 16 shows the acoustic transfer function between F_{ext} and the pressure in the middle of the output section of the tube. We note that the system effectiveness is very good in the range 700–2000 Hz with significant loss of attenuation above the frequency limit $f_c \approx 4300$ Hz of the transducer's distribution. The considerable damping rate of transducers at the first suspension mode (800 Hz) allows the loudspeaker's membranes to passively absorb energy contained in the frequency band around that of resonance.

The membrane displacements of each control loudspeaker have also been carried out for different frequencies. The control displacements are more important at low frequencies and at higher frequencies the actuator membranes start to bend so they are no longer rigid pistons (see Fig. 17).

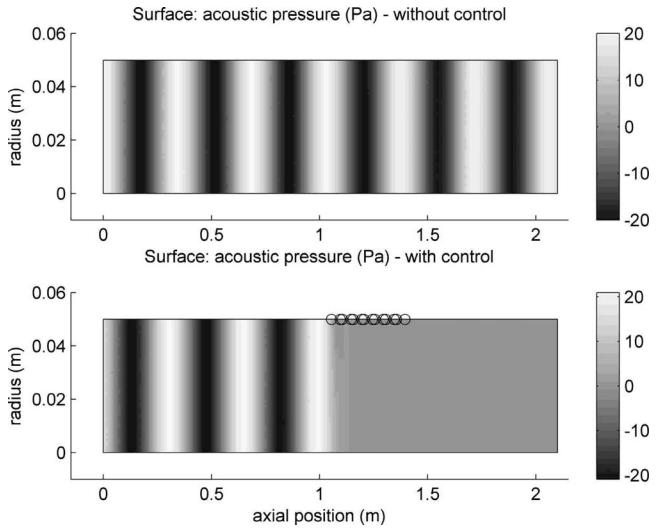


FIG. 13. Controlled and uncontrolled pressure distributions, frequency 1000 Hz.

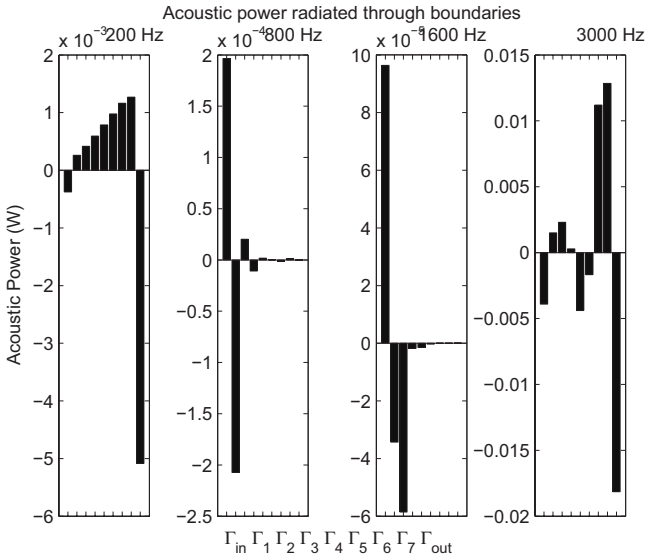


FIG. 15. Acoustic power radiated through the domain boundaries.

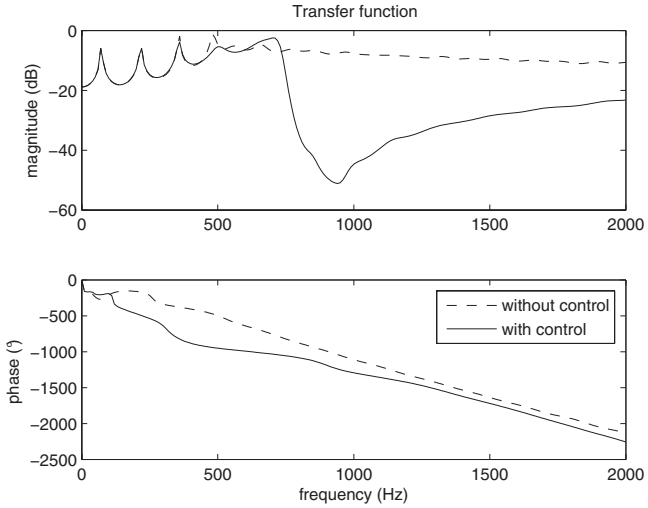


FIG. 16. Transfer function between the input and output acoustic pressures.

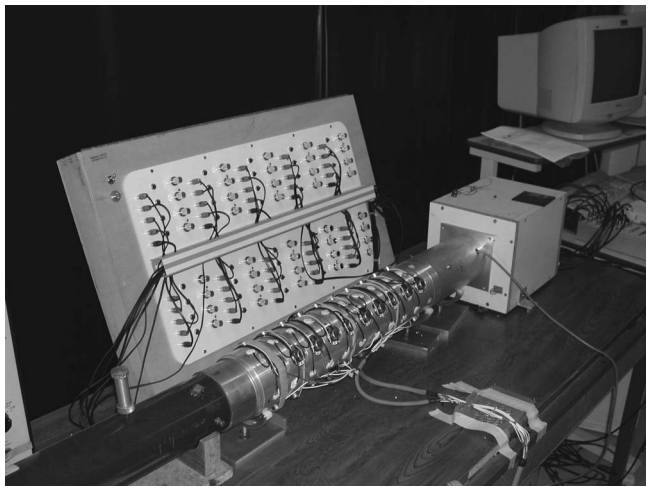


FIG. 18. Experimental setup.

This property is considerable especially without control and depends on the physical characteristics of the transducers, mainly on the internal stiffness introduced on these systems.

It is necessary to underline here that this internal stiffness is a very important dimensioning parameter that makes it possible to qualify the frequency band of the system. It affects the control effectiveness, in particular, compared to the *spillover* problems and must be as large as possible.

E. Experimental results

The breadboard construction carried out in our laboratory, shown in Fig. 18, implements exactly the control equations that we tested numerically. With the employed loudspeakers *Monacor* the frequency band with large effectiveness is reduced to between 700 and 1200 Hz. The limitations are due to the mechanical characteristics of the transducers, especially their first eigenmode at 800 Hz. Our hardware resources (dSPACE® DS1104 R&D Controller Board) provided seven channels of control loudspeakers and eight channels of microphones to implement the control strategy. Figure 19 shows the measured transfer function between the signal of disturbance applied to the input loudspeaker and the acoustic pressure measured at the tube's output.

One qualitatively observes the correct operation of the system compared to the numerical results presented previously. The reduction in the acoustic transfer is broadband and the form of the system response is comparable to that obtained theoretically (Fig. 16). We note that the magnitude peaks are caused by our anechoic termination, which is far from ideal, and their frequencies correspond to the stationary wave modes in the tube. The reasons for the *spillover* problems that appeared in practice at high frequencies can be the following.

- The realized experimental setup is obviously not perfectly axisymmetric. The induced coupling effects with these residual modes introduce high frequency uncontrolled perturbations being capable of producing instabilities. This behavior is totally comparable to the classical well known spillover effect.

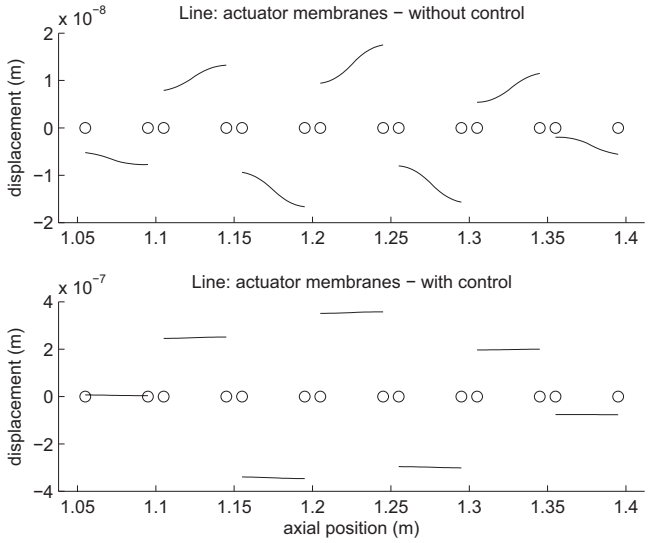


FIG. 17. Controlled and uncontrolled displacements of each speaker membrane, frequency 3000 Hz.

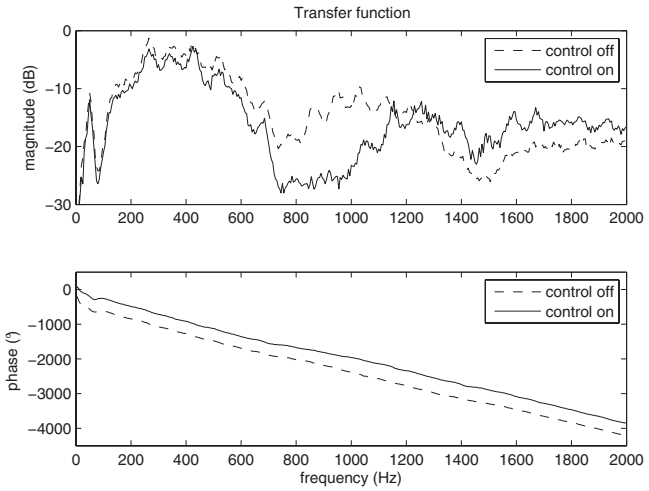


FIG. 19. Measured transfer function between the source loudspeaker voltage and the pressure at the output microphone.

- The time delay in the control loop is one of the major problems with such an application. As a completely centralized setup was used to implement the control algorithm, the time delay increases up to 0.2 ms. As a consequence, the control signal dephasing becomes large enough for frequency around 1250 Hz. This problem appears as the most constraining factor for experimental applications.
- The discretization used during the implementation of the partial derivative operator is also a limiting factor for our experimental tests.

There is not a perfect correlation between the experimental and numerical results but the main aim of this study was not to realize a precise model of the plane-wave tube with our "handmade" anechoic termination but to qualify the control strategy. It is possible to update a passive behavior of the numerical model by using experimental data but this work has not been carried out at this stage. Moreover, the simplified model we used allowed us to illustrate the main physical characteristic of the controlled system and to identify the most important design variables: suspension eigenfrequency, time delay of the control, transducer's discretization, internal dynamics of the control speakers, etc., which are the basic constraints limiting the efficient frequency band of the adaptive acoustical skin.

The applied experimental tests confirmed the expected functionality in the efficient frequency band without any instability problems. This underlines the high robustness of this distributed strategy, which appears as a fundamental property.

New calculations are currently being developed to introduce other aspects and find an optimum to handle all the difficulties of the control strategy. Another solution that will be investigated to deal with these induced constraints limiting the frequency band of interest is the hybrid control system using in parallel an optimal passive panel made with dedicated foam.

Moreover, it clearly appears that the band of effectiveness here from 700 to 1200 Hz depends mainly on the experimental space discretization of the transducers and on the time delay of the control loop. We are currently seeking to refine the test bench in order to show experimentally the influence of the space distribution on the effectiveness. In parallel, a totally decentralized implementation of the control will be carried out. It will be then possible to justify the development of a system with much more expensive microcomponent technologies. In fact, for the control of audible frequencies, there is no necessity of very high discretization of the network of acoustic transducers involving, microelectromechanical system (MEMS) technology, but we need the capability of designing and making highly integrated electroacoustical devices comprising not only transducers (microphone and loudspeaker) but also a distributed numerical control, enabling an improvement in the architecture and properties, compared to the centralized system. The microelectronic or MEMS technology in our study is not important for developing the transducers but for integrating the control electronics.

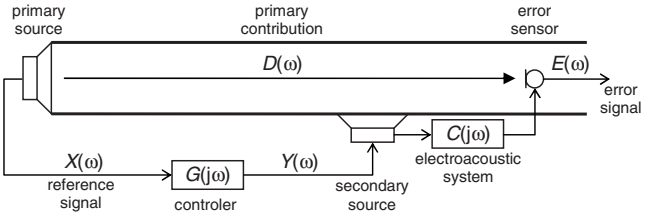


FIG. 20. Feedforward active control system.

V. COMPARISON WITH AN OPEN-LOOP CONTROL ALGORITHM

The same 2D finite element model designed previously using COMSOL MULTIPHYSICS™ software was used as a source for the feedforward algorithm developed in MATLAB® allowing the theoretical comparison with our distributed control. The feedforward method is widely used when the reference signal from the source is available. A single channel active control system is schematically illustrated in Fig. 20. It is important to note here that all of the system elements are considered to be linear. With this assumption we can express the frequency response of the error path using Fourier transforms

$$C(\mathrm{j}\omega) = \frac{E(\omega)}{Y(\omega)}, \tag{26}$$

where the contribution of the primary source is assumed to be $D(\omega)=0$. We can write for the total response at the error sensor

$$E(\omega) = D(\omega) + G(\mathrm{j}\omega)C(\mathrm{j}\omega)X(\omega). \tag{27}$$

Equation (27) was implemented on each of the seven channels of control transducers and the transfer function from the source to output was determined to compare the results with our distributed control. It can be seen in Fig. 21 that the efficiency of the feedforward algorithm remains similar in the entire frequency band but above $f=1$ kHz it is lower compared to that obtained with our distributed strategy.

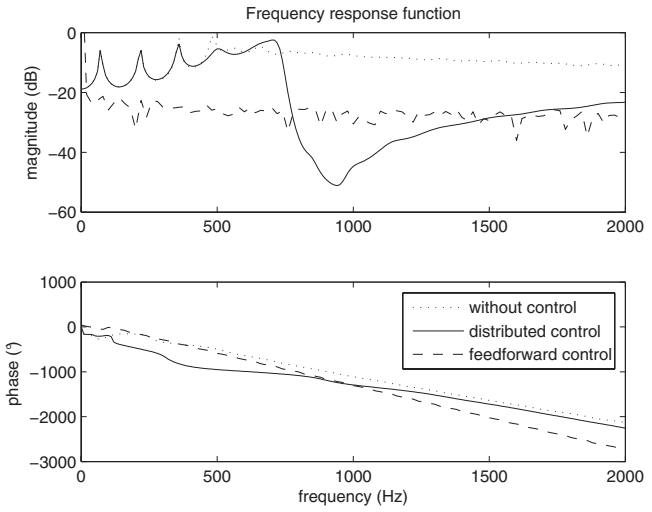


FIG. 21. Comparison of the transfer function for the distributed and feedforward algorithm.

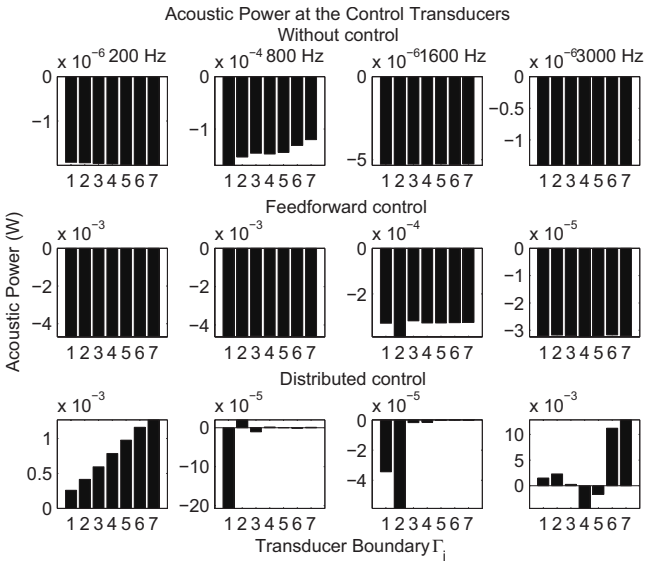


FIG. 22. Acoustic intensity for the noncontrolled, distributed, and feedforward system.

To compare both methods we calculated the acoustic power transmitted by each control loudspeaker using Eq. (25). As shown in Fig. 22, the energy necessary for our distributed control is less than for the feedforward algorithm especially at middle frequencies, where our distributed method absorbs the propagated acoustic energy with better efficiency. Furthermore, it can be seen in this figure also the effect of passive attenuation even without control at the resonant frequency of the control loudspeaker (800 Hz).

VI. CONCLUSION

A new methodology for designing control strategies based on distributed transducer network is presented in this paper. With the objective of controlling an acoustic energy transfer through a tube, we have demonstrated on a simple example the type of results that can be obtained. Although these results show clearly the potential benefits, they also present the difficulties of setting up the experimental realization, in particular, those introduced in high frequencies by the internal stiffness of the actuators and the coupling low frequencies on the suspension mode. The obtained experimental results are very encouraging and point out the potential of such a method.

The presented procedure has great merit of being very effective and allows an important attenuation of the acoustic transfer without the necessity of very strong displacements of the membrane interfaces. It is thus possible to use this technique with transducers produced on a silicon base with thin deposits of lead zirconate titanate layers as actuator elements.

Compared to traditional feedforward algorithm, the developed distributed control system is more efficient in the frequency band around 1 kHz and the power needed to minimize the acoustic transfer is less, thanks to the natural passive attenuation property of our system.

The proposed method appears to be suitable for use with distributed MEMS transducers. The effectiveness was shown on a simple example and the numerical modeling was validated by experimental results.

This paper shows certain results, aimed at the distributed control system development, completely different from other methods used up to now. We consider the fabrication of a network of miniactuators and sensors, which can be used to cover a surface and act on the noise transmission across this surface. These technologies open totally new perspectives in terms of noise suppression but to be fully exploited, it is necessary to completely reconsider the associated control strategies to quantify the effectiveness that can be expected from these new technologies for large frequency band noise reduction.

The next step in this work, which is running at the moment, is the development of individual cells with an integrated microcontroller, sensor and actuator, power supply, signal amplification and conditioning, and especially the possibility of interconnection of several or even many such cells in the area network, a so-called active antinoise skin. The distributed control algorithm of every cell will use as input signal its own sensor and several sensors from neighbor cells, which is necessary for the implementation of 2D spatial operators. The global behavior of the entire network will then be supervised and modified, if necessary, by one *master* microcontroller.

[1]N. Atalla, R. Panneton, F. C. Sgard, and X. Olny, "Acoustic absorption of macro-perforated porous materials," J. Sound Vib. **243**, 659–678 (2001).
[2]N. Sellen, M. Cuesta, and M. A. Galland, "Passive layer optimization for active absorbers in flow duct applications," Ninth AIAA/CEAS Aeroacoustics Conference, 2003, AIAA Paper No. 2003-3186.
[3]R. Ramakrishnan and W. R. Watson, "Design curves for rectangular splitter silencers," Appl. Acoust. **35**, 1–24 (1992).
[4]C. Yilmaz and N. Kikuchi, "Analysis and design of passive low-pass filter-type vibration isolators considering stiffness and mass limitations," J. Sound Vib. **293**, 171–195 (2006).
[5]H. Zheng, C. Cai, G. S. H. Pau, and G. R. Liu, "Minimizing vibration response of cylindrical shells through layout optimization of passive constrained layer damping treatments," J. Sound Vib. **279**, 739–756 (2005).
[6]M. L. Munjal, "Analysis and design of mufflers—An overview of research at the Indian Institute of Science," J. Sound Vib. **211**, 425–433 (1998).
[7]J. S. Vipperman, R. A. Burdisso, and C. R. Fuller, "Active control of broadband structural vibration using the LMS adaptive algorithm," J. Sound Vib. **166**, 283–299 (1993).
[8]P. Gardonio, E. Bianchi, and S. J. Elliott, "Smart panel with multiple decentralized units for the control of sound transmission. Part I: Theoretical predictions," J. Sound Vib. **274**, 163–192 (2004).
[9]P. Gardonio, E. Bianchi, and S. J. Elliott, "Smart panel with multiple decentralized units for the control of sound transmission. Part II: Design of the decentralized control units," J. Sound Vib. **274**, 193–213 (2004).
[10]P. Gardonio, E. Bianchi, and S. J. Elliott, "Smart panel with multiple decentralized units for the control of sound transmission. Part III: Control system implementation," J. Sound Vib. **274**, 215–232 (2004).
[11]P. A. Nelson and S. J. Elliott, *Active Control of Sound* (Academic, London, 1992).
[12]T. M. Kostek and M. A. Franchek, "Hybrid noise control in ducts," J. Sound Vib. **237**, 81–100 (2000).
[13]A. Benjeddou, "Advances in hybrid active-passive vibration and noise control via piezo-electric and viscoelastic constrained layer treatments," J. Vib. Control **7**, 565–602 (2001).
[14]R. L. Clark and C. R. Fuller, "Experiments on active control of structurally radiated sound using multiple piezoceramic actuators," J. Acoust. Soc. Am. **91**, 3313–3320 (1992).
[15]A. Preumont, A. Francois, F. Bossens, and A. Abu-Hanieh, "Force feedback versus acceleration feedback in active vibration isolation," J. Sound

Vib. **257**, 605–613 (2002).
[16]M. Furstoss, D. Thenail, and M. A. Galland, "Surface impedance control for sound absorption: Direct and hybrid passive/active strategies," J. Sound Vib. **203**, 219–236 (1997).
[17]B. Mazeaud, "Developing of an intelligent sound coating for a duct in the presence of flow," Ph.D. thesis, Laboratory of Fluid mechanics and Acoustics, Centrale Lyon, 2005.
[18]D. Guicking and K. Karcher, "Active impedance control for one-dimensional sound," ASME J. Vib., Acoust., Stress, Reliab. Des. **106**, 393–396 (1984).
[19]D. Guicking, K. Karcher, and M. Rollwage, "Coherent active methods for applications in rooms acoustics," J. Acoust. Soc. Am. **78**, 1426–1434 (1985).
[20]F. O. Bustamante and P. A. Nelson, "An adaptive controller for the active absorption of sound," J. Acoust. Soc. Am. **91**, 2740–2747 (1992).
[21]O. Lacóur, M. A. Galland, and D. Thenail, "Preliminary experiments on noise reduction in cavities using active impedance changes," J. Sound Vib. **230**, 69–99 (2000).
[22]H. F. Olson and E. G. May, "Electronic sound absorber," J. Acoust. Soc. Am. **25**, 1130–1136 (1953).
[23]D. Guicking and E. Lorentz, "An active sound absorber with porous plate," ASME J. Vib., Acoust., Stress, Reliab. Des. **106**, 389–392 (1984).
[24]M. A. Galland, B. Mazeaud, and N. Sellen, "Hybrid passive/active absorbers for flow ducts," Appl. Acoust. **66**, 691–708 (2005).
[25]G. Montseny, "Diffusive representation of pseudo-differential time-operator," ESAIM: Proceedings, fractional differential systems: Models, methods and applications, **5**, 159–175 (1998).
[26]D. Matignon, J. Audounet, and G. Montseny, "Fractional integrodifferential boundary control of the Euler-Bernoulli beam," Conference on Decision and Control, IEEE-CSS, 1998, pp. 4973–4978.
[27]D. Matignon, J. Audounet, and G. Montseny, "Smart energy decay for wave equations with damping of fractional order," Fourth International Conference on Mathematical and Numerical Aspects of Wave Propagation Phenomena, 1998, pp. 638–640.
[28]I. Lasiecka and R. Triggiani, "Exact controllability of the wave equation with Neumann boundary control," Appl. Math. Optim. **19**, 243–290 (1989).
[29]N. Tanaka and Y. Kikushima, "Optimal vibration feedback control of an Euler-Bernoulli beam: Toward realization of the active skin method," J. Vibr. Acoust. **121**, 174–182 (1999).
[30]N. Tanaka and H. Sakano, "Cluster power flow control of a distributed-parameter planar structure for generating a vibration-free zone," Smart Mater. Struct. **16**, 47–56 (2007).
[31]A. C. Galucio, J. F. Deü, and R. Ohayon, "A fractional derivative viscoelastic model for hybrid active-passive damping treatments in time domain—Application to sandwich beams," J. Intell. Mater. Syst. Struct. **16**, 33–45 (2005).
[32]M. Á. Fernández and P. Le Tallec, "Linear stability analysis in fluidstructure interaction with transpiration. Part I: Formulation and mathematical analysis," Comput. Methods Appl. Mech. Eng. **192**, 4805–4835 (2003).

Children's annoyance reactions to aircraft and road traffic noise

Elise E. M. M. van Kempen,[a] Irene van Kamp,[b] and Rebecca K. Stellato[c]
National Institute for Public Health and the Environment, Centre for Environmental Health Research, P.O. Box 1, 3720 BA Bilthoven, The Netherlands

Isabel Lopez-Barrio[d]
Institution de Acustico, Consejo Superior De Investigaciones Cientificas (CSIC), C/Serrano 144, E-28006 Madrid, Spain

Mary M. Haines[e]
The Sax Institute, 235 Jones Street, Ultimo, Sydney, New South Wales 2000, Australia

Mats E. Nilsson[f]
Department of Psychology, Institute of Environmental Medicine, and Karolinska Institute, Stockholm University, SE-106 91 Stockholm, Sweden

Charlotte Clark[g]
Barts and the London, Queen Mary's School of Medicine and Dentistry, University of London, London EC1M 6BQ, United Kingdom

Danny Houthuijs[h]
National Institute for Public Health and the Environment, Centre for Environmental Health Research, P.O. Box 1, 3720 BA Bilthoven, The Netherlands

Bert Brunekreef[i]
Institute for Risk Assessment Sciences (IRAS), and Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, University of Utrecht, P.O. Box 80178, 3508TD Utrecht, The Netherlands

Birgitta Berglund[j]
Department of Psychology, Institute of Environmental Medicine, and Karolinska Institute, Stockholm University, SE-106 91 Stockholm, Sweden

Stephen A. Stansfeld[k]
Barts and the London, Queen Mary's School of Medicine and Dentistry, University of London, London EC1M 6BQ, United Kingdom

(Received 10 June 2008; revised 27 October 2008; accepted 6 December 2008)

Since annoyance reactions of children to environmental noise have rarely been investigated, no source specific exposure-response relations are available. The aim of this paper is to investigate children's reactions to aircraft and road traffic noise and to derive exposure-response relations. To this end, children's annoyance reactions to aircraft and road traffic noise in both the home and the school setting were investigated using the data gathered in a cross-sectional multicenter study, carried out among 2844 children (age 9–11 years) attending 89 primary schools around three European airports. An exposure-response relation was demonstrated between exposure to aircraft noise at school ($L_{Aeq,7–23\ h}$) and severe annoyance in children: after adjustment for confounders, the percentage severely annoyed children was predicted to increase from about 5.1% at 50 dB to about 12.1% at 60 dB. The findings were consistent across the three samples. Aircraft noise at home ($L_{Aeq,7–23\ h}$) demonstrated a similar relation with severe annoyance. Children attending schools with higher road traffic noise ($L_{Aeq,7–23\ h}$) were more annoyed. Although children were less annoyed at levels above 55 dB, the shapes of the exposure-response relations found among children were comparable to those found in their parents.
 [DOI: 10.1121/1.3058635]

PACS number(s): 43.50.Qp, 43.50.Lj, 43.66.Lj [BSF] Pages: 895–904

I. INTRODUCTION

Annoyance is one of the most widespread and well-documented responses to noise. It is a collective term for several negative reactions such as irritation, dissatisfaction, or anger, which appear when noise disturbs someone's daily activities.[1] While adult reactions to noise have been well described,[2–4] this is not so for noise annoyance in children. In comparison with adults, children may be particularly vulnerable to the effects of noise because they have less capacity to anticipate, understand, and cope with stressors.[5]

Exposure-response relationships for noise annoyance among adults have been widely studied, and large datasets have allowed the construction of generalized curves.[6–11] For children, generalized exposure-response relationships are lacking. According to Lercher,[12] this omission is due to a lack of a standard methodology for measuring annoyance in children and insufficient representative data on which to base a generalized exposure-response relationship. Four previous studies have assessed residential noise annoyance in children in a quantitative and systematic manner: the Munich airport study,[13,14] the Heathrow studies,[15–17] and the Tyrol studies.[18,19] In these studies, children living in noisier areas in their community were significantly more annoyed by noise than children living in quieter areas.

Most studies have only focused on exposure at school when investigating the effects of noise exposure in children. This is a gap in research since the impact of noise on children's health can occur in different environments over a 24 h period: at home and at school, indoors and outdoors and over different times of the day.[20]

Among adults, annoyance is usually measured by means of one or more questions as part of a questionnaire or interview.[21] In the past, a wide variety of questions and scaling methods has been employed to measure annoyance.[11] As with adult studies, different methods have been used to measure children's annoyance reactions. Although each of the studies purports to measure annoyance, it is not fully clear what is being measured. Some studies[13,14] define annoyance as an affective response that indicates a chronic decline in well-being; others[22] conclude that noise annoyance in children pertains to the same construct as in adults, since the emotional response to aircraft noise was consistent with adult reactions. In previous studies among adults,[23,24] interference and annoyance were highly related while well-being formed a separate dimension. It is uncertain whether children are also able to make such distinctions and thus show a comparable pattern to adults.

The primary goal of this paper is to investigate children's annoyance reactions and the existence of exposure-response associations to aircraft and road traffic noise in both the home and the school setting, using data collected from children living around three European airports, gathered in the framework of the European Fifth Framework Project Road Traffic and Aircraft Noise Exposure and Children's Cognition and Health (RANCH). A secondary goal was to compare children's annoyance reactions with those of their parents. Some results of RANCH have already been reported elsewhere,[25] focusing on the effects of noise exposure at school on cognition and health.

II. METHODS

A. Selection and recruitment

Children aged 9–11 years were recruited from primary schools in areas around Heathrow airport (London, UK), Schiphol airport (Amsterdam, The Netherlands), and Madrid-Barajas airport (Spain). Schools were selected according to the modeled noise exposure of the school area (expressed as $L_{Aeq,7–23\ h}$) and were matched on indicators of socioeconomic status (SES) and ethnicity. Out of 767 primary schools available, 134 were invited to participate and 89 agreed. The parents or caregivers of 3207 children were approached through the schools by letter to give consent for their children to participate. Written consent was additionally obtained from the children. The final sample contained 2844 children. For full details of selection and recruitment, see Ref. 25.

B. Procedure

The children completed a self-administered questionnaire on annoyance as part of a 2 h group testing session which also included various paper-and-pencil tests measuring cognitive abilities.[25] The children were also given a questionnaire to take home for one of their parents or caregivers to complete, and requested information on the health and behavior of the child, on hearing transportation sounds and their annoyance, and potential confounding factors such as glazing of the child's home, length of residency, indicators for SES (employment status, crowding, maternal education, and parental home ownership), ethnic origin, and main language spoken at home. These variables were only available for those children whose parents also completed the questionnaire, so parents' participation served as a criterion for inclusion in the statistical analysis. To ensure accurate conceptual translation, all questionnaires were translated from English into Dutch and Spanish and subsequently back-translated. Before data collection, all procedures and materials were tested in a pilot study in October 2001. In all three participating countries, ethical approval of the study was obtained.

[a]Author to whom correspondence should be addressed. Electronic mail: elise.van.kempen@rivm.nl
[b]Electronic mail: irene.van.kamp@rivm.nl
[c]Electronic mail: r.k.stellato@bio.uu.nl
[d]Electronic mail: iaclb41@ia.cetef.csic.es
[e]Electronic mail: mary.haines@saxinstitute.org.au
[f]Electronic mail: mnn@psychology.su.se
[g]Electronic mail: c.clark@qmul.ac.uk
[h]Electronic mail: danny.houthuijs@rivm.nl
[i]Electronic mail: b.brunekreef@iras.uu.nl
[j]Electronic mail: birber@mbox.ki.se
[k]Electronic mail: s.a.stansfeld@qmul.ac.uk

C. Noise exposure assessment

In each country, predicted levels of aircraft noise exposure for both the *school* and children's *home* address were obtained from nationally available noise contours or grids. The United Kingdom and Spain both used noise contours predicting $L_{Aeq,16\ h}$, which predicts average noise exposure from 0700 to 2300 h for a three month period for the year 2000. In The Netherlands, modeled aircraft noise levels ($L_{Aeq,7-23\ h}$) for the period of a year with a resolution of 250×250 m^2 grids were obtained. The contours in the UK and Spain were provided by the British Civil Aviation Authority and the Spanish Airports and Air Navigation, respectively. The Netherlands used noise data from the year 2001 provided by the Dutch National Aerospace Laboratory (NLR).

Road traffic noise levels were only available for the school situation since it was not possible to estimate road traffic noise exposure at home in a reliable way for this study. Predictions of road traffic noise at school (expressed in $L_{Aeq,7-23\ h}$) were made using different methods in each country, as the data detailing road traffic noise in each country varied enormously. In the UK, road traffic noise exposure was predicted using calculation of road traffic noise (CRTN).[26] This method involves obtaining the traffic flow data for the road section nearest to the school and calculating noise exposure. The traffic flow data covered the period 0700 to 2300 h. In The Netherlands, modeled composite data from 2000 and 2001, with a resolution of 25×25 m^2 grids, were linked to school addresses using a geographic information system (GIS).[27] In Spain, direct external measurements were taken of road noise during school visits. Taking into account factors such as traffic flow, speed limits, and distance to the street, these were transformed into 7–23 h L_{Aeq}-values.

D. Child and parent noise annoyance

For both children and parents, annoyance was measured as part of a self-administered questionnaire by means of standard questions. For children the following wording was used: "Thinking about the last year, when you are at [school] [home], how much does the noise from [aircraft] [road traffic] bother, disturb, or annoy you?" Answers were indicated on a five-point category scale ("not at all, a little, quite a bit, very much, extremely"). For parents the following wording was used: "Thinking about the last 12 months, when you are at home, how much does noise from [aircraft] [road traffic] noise bother, disturb, or annoy you?" Answers were indicated on a five-point category scale ("not at all, slightly, moderately, very, and extremely").[21]

Children and parents were also asked how frequently they heard the noise from road traffic or aircraft when they were at school or home: "Do you hear noise from [aircraft] [road traffic] when at [school] [home]?" Answers were indicated on a four-point category scale ("never, sometimes, often, and always"). If parents indicated never hearing noise while indicating "slightly," "moderately," "very," or "extremely" annoyed, their answer on the question measuring annoyance was recoded to "not at all annoyed." Since we could not necessarily expect children to answer these questions in such a consistent way, this transformation was not used for the children. For both children and parents, the answers of the annoyance questions were subsequently dichotomized, with the two highest categories ("very much" and extremely annoyed) defining "severely annoyed." This cutoff corresponds to a lower cutoff than used for defining "highly annoyed" in the pooled analyses of Miedema and co-workers.[9–11] In their analyses, the annoyance scale is transformed to a 0–100 scale and the cutoff for highly annoyed corresponds to a value of 72.[9–11] Our cutoff point would correspond to a cutoff at 60 on the same scale.

E. Interference with activities

Interference with activities at school and at home was measured by asking the children whether noise from road or aircraft noise interfered with (i) playing outdoors, (ii) working in a group, (iii) working individually, (iv) listening to the teacher, (v) listening to TV, radio, or music, (vi) talking, or (vii) reading or doing homework. Answers were indicated on a four-point category scale ("never, sometimes, often, and always").

F. Perceived health

In order to measure perceived health, the children were asked how often they had the following symptoms during the past month: headache, vomiting, stomachache, difficulty falling asleep, and the number of times woken at night or felt sleepy during the day. Answers were indicated on a five-point category scale ("never, a few times, once a week, a few times a week, and every day or night").

G. Data analysis

In order to test the convergent and divergent validity of the annoyance scale, a principal component analysis (PCA) was carried out using SPSS for Windows (version 12.0.1) on the annoyance and interference questions for both the school and home situation and perceived health. Home and school annoyance and interference questions were combined in the PCA, and subjective health symptoms were included, in order to determine whether children could distinguish between the home and school situation, and between annoyance interference and subjective health. We expected high correlations between annoyance and/or interference at school and at home for aircraft noise, respectively, but not necessarily for road traffic noise. Only components that accounted for variances with eigenvalues greater than 1 were included in the following presentation. To make the components more interpretable a rotation with the Varimax method was performed. However, on the basis of age, gender, etc., one would expect a certain correlation between the components. As a kind of sensitivity analysis an oblique rotation (with delta=0) was performed in addition to Varimax rotation, assuming that the resulting components may be correlated. Cronbach's alphas were calculated to test the internal consistency of the obtained components.

To assess the association between aircraft and road traffic noise exposure and severe annoyance, multilevel logistic regression analyses by means of generalized linear mixed models were carried out using the GLIMMIX procedure in SAS version 9.1. The advantage of multilevel modeling compared to a simple logistic regression approach is its ability to take into account effects at the level of center, school, and pupil simultaneously. Two-level (pupil and school) random intercept models were used, and country was included as a fixed effect. Coefficients (B) and standard errors were estimated under residual pseudolikelihood estimation. In all models, aircraft or road traffic noise exposure (either at school or at home) was the main independent variable and was included as a continuous variable. For the association with aircraft noise, a quadratic term for aircraft noise was also included because this increased the model fit (see also Ref. 25). The logistic regression models included age (years), sex, ethnicity (white/nonwhite), school glazing (single, mixed, double, and triple) or double glazing at home (yes/no), length of school enrolment (<1, 1–2, 3–6, and >6 years) or residency (<1, 1–5, 6–10, and >10 years), and indicators of SES (crowding, home ownership, parental employment, and mother's education) as potential confounders. Models were estimated for the pooled data. Heterogeneity in the exposure-response relationships among countries was tested in the models on the pooled data by examining the interaction between country and noise exposure. Statistical significance of a coefficient was tested under maximum pseudolikelihood estimation, using a Wald chi-square test.

III. RESULTS

A. Sample information

The British sample contains fewer employed parents, fewer home owners, and more nonwhite children than the Dutch and Spanish samples. The prevalence of severe annoyance due to aircraft and road traffic noise in the Dutch and Spanish samples was somewhat lower than in the British sample. There were also differences between the samples in terms of length of time at school and glazing (Table I).

B. Construct validity

The PCA on interference, annoyance, and perceived health yielded five components with eigenvalues greater than 1 (Table II). The total percentage of variance explained by these five components was 56%. The values of the Cronbach's alphas indicate that the components have a high reliability. Items referring to annoyance and interference from aircraft noise annoyance (without distinction between home and school situations) loaded highly on the first component, whereas items regarding annoyance and interference from road traffic noise at school loaded highly on the second component. Items referring to interference at home from road traffic noise loaded highly on the third component, and items on self-reported health symptoms loaded highly on the fourth component. Items loading highly on the fifth component referred to interference when playing outdoors at home, and school due to aircraft and road traffic noise, and annoyance from road traffic noise at home. The oblique rotation resulted in the similar grouping of variables as the Varimax rotation, and the interpretation of the components did not change.

C. Aircraft noise exposure *at school*

Aircraft noise exposure at school was significantly related to severe annoyance ($\chi^2=52.7$, $df=2$, $p<0.0001$): in schools in areas with higher aircraft noise exposure the proportion severely annoyed children was significantly higher. The percentage severely annoyed children was predicted to increase from about 5% at 50 dB to about 12% at 60 dB (Fig. 1). The only potential confounder that had a significant effect on annoyance was mother's education ($\chi^2=6.8$, $df=1$, $p=0.009$); children of mothers with a higher level of education were more annoyed by aircraft noise at school; an odd ratio (OR) of 2.24 (95%CI: 1.22–4.12) was estimated. Country did not have a significant effect on annoyance ($\chi^2=1.6$, $df=2$, $p=0.457$). Although the proportion of severely annoyed children in the Dutch sample was higher compared to the British and Spanish samples at aircraft noise levels ($L_{Aeq,7-23\ h}$) of 63 dB and higher, the change in the percentage severely annoyed per 1 dB increase in the noise did not differ significantly between the three countries (test of heterogeneity: $\chi^2=8.9$, $df=4$, $p=0.064$).

D. Aircraft noise exposure *at home*

Aircraft noise exposure at home was significantly related to severe annoyance ($\chi^2=50.5$, $df=2$, $p<0.0001$): the proportion of severely annoyed children was higher in areas with higher aircraft noise levels. The percentage severely annoyed children was predicted to increase from about 7% at 50 dB to about 15% at 60 dB (Fig. 2). Country did not have a significant effect on annoyance. The only potential confounder that had a significant effect on annoyance was gender: girls were less annoyed due to aircraft noise at home than boys ($\chi^2=8.3$, $df=1$, $p=0.004$) [OR=0.62 (95%CI: 0.45–0.86)]. The difference in the effect size at different noise levels for each country was statistically not significant (test of heterogeneity: $\chi^2=5.9$, $df=4$, $p=0.209$). Comparison between Figs. 1 and 2 indicates that the exposure-response relationships for the home and school situations are similar.

E. Road traffic noise exposure *at school*

Road traffic noise exposure at school was significantly related to severe annoyance from road traffic noise at school: children attending schools with higher road traffic noise were more annoyed ($\chi^2=7.4$, $df=1$, $p=0.007$). The percentage severely annoyed children was predicted to increase from about 4% at 50 dB to about 6% at 60 dB (see also Fig. 3). Potential confounders that had a significant effect on annoyance were mother's educational attainment ($\chi^2=16.6$, $df=1$, $p<0.0001$) [OR=5.04 (95%CI: 2.28–11.8)], school enrolment ($\chi^2=8.4$, $df=3$, $p=0.040$), and school glazing ($\chi^2=7.2$, $df=2$, $p=0.028$). There was no significant difference in the change in the percentage severely annoyed per 1 dB increase in the noise between the three countries (test of heterogeneity: $\chi^2=0.70$, $df=2$, $p=0.704$). Comparison between Figs. 1 and 3 indicates that the exposure-response re-

TABLE I. General characteristics of the children and their parents included in the analysis. (Abbreviations: *N*, sample size; SD, standard deviation;%, percentage; L_{Aeq},7–23 h, equivalent noise level from 7 to 23 h).

Characteristic	UK (*N*=863)	The Netherlands (*N*=612)	Spain (*N*=553)
No. of participating schools	29	33	27
Girls (%)	54.5	50.1	53.0
Mothers (%)	93.3	90.8	92.0
Mean age (SD)			
Children	10.3 (0.3)	10.5 (0.6)	10.9 (0.4)
Parents	37.7 (5.5)	40.9 (4.1)	39.6 (5.0)
Socioeconomic status			
Crowding in the home (%)	22.1	31.4	9.6
Parental home ownership (%)	59.4	81.7	85.5
Employed parents (%)	78.5	93.0	89.3
Mean mother's education (SD)[a]	0.5 (0.3)	0.5 (0.3)	0.5 (0.3)
White British/Dutch/Spanish (%)	66.2	89.4	91.5
Length of time at school (%)			
Less than 1 year	3.5	0.2	0.2
1–2 years	10.4	6.6	3.4
3–6 years	49.7	21.3	9.0
More than 6 years	36.5	72.0	87.5
Length of residence (%)			
Less than 1 year	7.0	4.1	7.1
1–5 years	33.8	19.8	21.3
6–10 years	26.7	19.3	17.5
More than 10 years	32.5	56.9	54.1
Severe annoyance (%)			
Aircraft noise at school, children	10.9	9.4	7.4
Aircraft noise at home, children	12.2	7.1	7.8
Road traffic noise at school, children	6.5	4.0	4.9
Aircraft noise at home, parents	11.6	8.7	6.2
Mean modeled noise exposure ($L_{Aeq,7–23\ h}$) levels [dB(A)] (range)[b]			
Aircraft noise at school	53.0 (34.0–68.0)	54.2 (41.0–68.0)	46.1 (30.0–77.0)
Aircraft noise at home	53.1 (33.9–72.8)	49.1 (34.5–64.5)	46.4 (31.9–72.8)
Road traffic noise at school	50.6 (37.0–67.0)	49.3 (34.0–62.0)	54.1 (43.0–71.0)
Insulation			
School glazing (%)			
Single	51.9	44.3	71.3
Mixed	9.0	–	–
Double	39.1	46.6	28.8
Triple	–	9.2	–
Double glazing at home (%)	82.2	58.1	57.1

[a]Ranked index of standard qualification in every country.
[b]The range runs from the minimum value to the maximum value.

lationships for road traffic noise and aircraft noise differ from each other: the relation with annoyance is much stronger for aircraft noise.

F. Comparison of childrens' and parents' annoyance reaction to aircraft noise

In Fig. 4 the exposure-response relationships for both children and parents are presented for the exposure to aircraft noise at home. The percentage severely annoyed children was predicted to increase from about 7% at 50 dB to about 15% at 60 dB; the percentage severely annoyed parents was predicted to increase from about 5% at 50 dB to about 17% at 60 dB. There was a significant difference in exposure-response gradient between the children and their parents ($\chi^2=18.7$, $df=2$, $p<0.0001$). At levels above 55 dB the percentage of severely annoyed children is lower than the percentage of severely annoyed parents, but below 45 dB the percentage of severely annoyed children is slightly higher than the percentage of severely annoyed parents.

TABLE II. Component loading matrix (N=2185) using Varimax rotation. (Only loadings >0.400 are shown. Loadings within "⟨ ⟩" were not used when interpreting the data. N is the sample size.)

	Component				
Item	I	II	III	IV	V
Annoyed road traffic at school		0.617			
Annoyed aircraft at school	0.740				
Annoyed road traffic at home					0.410
Annoyed aircraft at home	0.655				
Road traffic interferes play outdoors at school					0.665
Road traffic interferes group work at school		0.692			
Road traffic interferes own work at school		0.736			
Traffic interferes listening to teacher at school		0.699			
Aircraft interferes play outdoors at school	0.595				
Aircraft interferes group work at school	0.702				
Aircraft interferes own work at school	0.700				
Aircraft interferes listening to teach at school	0.691				
Road traffic interferes play outdoors at home					0.740
Road traffic interferes TV at home			0.701		
Road traffic interferes talking at home			0.635		
Road traffic interferes reading at home			0.685		
Aircraft interferes play outdoors at home					0.564
Aircraft interferes TV at home	⟨*0.539*⟩		0.575		
Aircraft interferes talking at home	0.571		⟨*0.504*⟩		
Aircraft interferes reading at home	0.565		⟨*0.558*⟩		
Headaches				0.610	
Vomiting				0.620	
Stomachache				0.579	
Difficult to sleep				0.590	
Times awake				0.641	
Sleepy in day				0.573	

Component	Interpretation	Variance explained	Alpha[a]
I	Disturbance and annoyance due to aircraft noise	32.471	0.89
II	Disturbance and annoyance due to road traffic noise at school	7.512	0.79
III	Interference from road traffic noise at home	6.220	0.77
IV	Health	5.225	0.67
V	Interference playing outdoors	4.843	0.69
Total		56.272	

[a]Cronbach's alpha (standardized): function of the item intercorrelation and the number of items included in the scale based on the items given in the component loading matrix.

IV. DISCUSSION

We found significant associations between aircraft and road traffic noise exposure and annoyance among school children living near three major European airports. This is consistent with results of previous studies investigating children's reactions to aircraft and road traffic noise,[13–18] which demonstrated that annoyance was significantly higher among children in high noise schools and areas compared with low noise schools and areas.

A. Measurement of children's annoyance

The results of the PCA show that children can make a clear distinction between annoyance and perceived health as measured by means of self-reported symptoms. As in adults,[23,24] the correlation found between annoyance and interference or disturbance of activities was high. This is consistent with the findings of a survey among 207 children (aged 13–14 years) investigating the effects of road traffic noise.[28] Our results are consistent with Haines and Stansfeld[29] investigating the effects of aircraft noise; they found that severely annoyed children agreed more often that "noise makes it hard to work" than children who were less annoyed. However, aircraft noise annoyance at school was not found to be associated with other aspects of classroom interference.

Our data suggest that the children were able to distinguish between indoors and outdoors rather than between school and home. This is clear from the grouping of items in Table II, where the items "road traffic interferes play out-

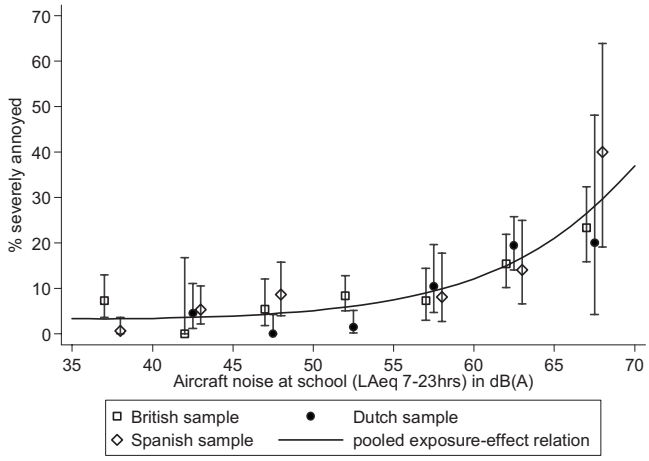


FIG. 1. The country-specific percentage severely annoyed children by 5 dB bands of aircraft noise ($L_{Aeq,7-23\ h}$) at school and the relationship between aircraft noise at school and the percentage of children severely annoyed derived after pooling the data and adjustment for confounders. The vertical lines correspond to the 95% confidence interval.

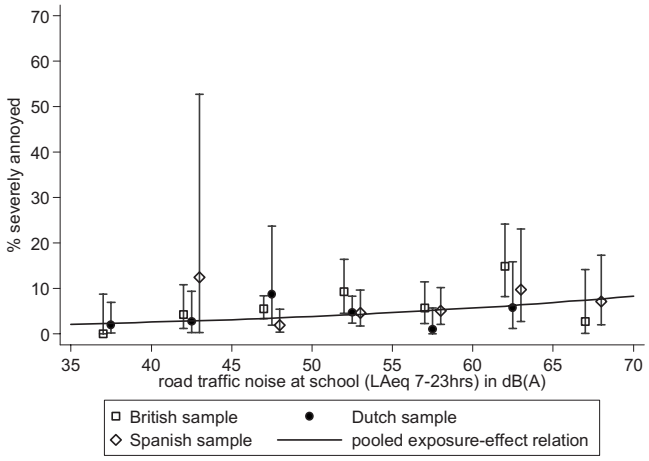


FIG. 3. The country-specific percentage severely annoyed children by 5 dB bands of road traffic noise ($L_{Aeq,7-23\ h}$) at school and the relationship between road traffic noise at school and the percentage of children severely annoyed derived after pooling the data and adjustment for confounders. The vertical lines correspond to the 95% confidence interval.

doors at school," road traffic interferes play outdoors at home," and "aircraft interferes play outdoors at home" loaded high in the same component. Also, there was a clear distinction between annoyance from aircraft and road traffic noise, but not between annoyance from aircraft noise at school and at home. The latter result was not the case for road traffic noise. The observed high correlation between annoyance from aircraft noise at school and at home is consistent with the distribution of aircraft noise exposure levels: aircraft noise levels at school and at home were also highly correlated in each of the countries ($r \sim 0.85-0.93$).[30] Since primary schools are usually located in the residential area of the child, we expect a great similarity in exposure levels between the school and home situations. Children aged 9–12 years appear to be able to discriminate their annoyance responses to road and aircraft noise sources and are consistent in their annoyance responses to aircraft noise across contexts such as school and home. Our results also indicate that children clearly distinguish between sources of noise as well as between annoyance and other indicators of well-being.

B. Annoyance reactions to aircraft and road traffic noise

After pooling the data, noise exposure levels of both aircraft and road traffic were significantly related to the percentage of severely annoyed children. No significant differences were found in the fraction severely annoyed at different exposure levels between countries. This is different from the variability earlier observed in a review evaluating studies that investigated adults' noise annoyance reactions.[31]

Another finding of our study was that the association with annoyance in children is stronger for aircraft than for road traffic noise, similar to adults. First, it is likely that aircraft noise has a greater effect on children's annoyance reactions than road traffic noise amongst others because of

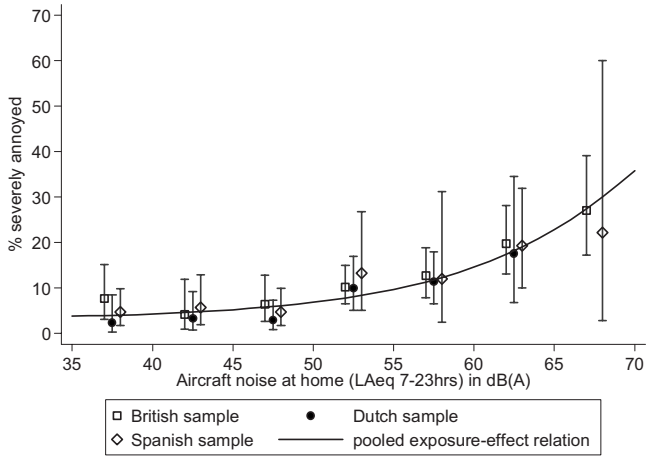


FIG. 2. The country-specific percentage severely annoyed children by 5 dB bands of aircraft noise ($L_{Aeq,7-23\ h}$) at home and the relationship between aircraft noise at home and the percentage of children severely annoyed derived after pooling the data and adjustment for confounders. The vertical lines correspond to the 95% confidence interval.

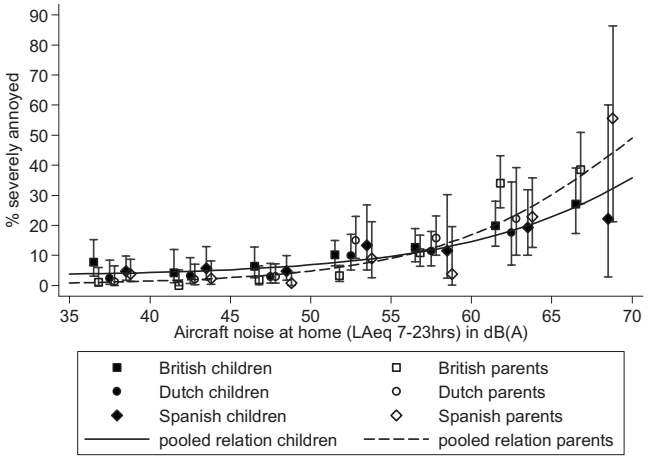


FIG. 4. Comparison between children and their parents: the country-specific percentage severely annoyed children and parents by 5 dB bands of aircraft noise ($L_{Aeq,7-23\ h}$) at home and the relationship between aircraft noise at home and the percentage of children and parents severely annoyed derived after pooling the data and adjustment for confounders. The vertical lines correspond to the 95% confidence interval.

its intensity, its variability, and unpredictability in comparison with road traffic noise.[1,32] Second, exposure misclassification may also have occurred because classrooms were at varying distance from the façade of the school building.[30] A third possible explanation is that the combined exposure to aircraft noise and road traffic noise might have affected children's annoyance response: children in high aircraft noise areas might report more annoyance from aircraft noise in high road traffic noise areas than children in low road traffic noise areas and vice versa. Fourth, differences in school systems and teachers' attitudes and/or responses toward noise might have differential effects on the children's reactions to noise sources at school. There might be differences in frequency and type of insulation of both schools and homes, which could result in different annoyance reactions, even though both design and analysis accounted for the influence of insulation. Finally, with the current available methods and data it is more dubious to predict road traffic noise exposure accurately. Different countries use different methods for the CRTN exposure. Previous comparisons of different national calculation methods for certain road traffic situations revealed that differences up to 15 dB may exist.[33] However, since the exposure-response functions for annoyance did not differ much between countries, this seems to indicate that the different methods for assessing the exposure were robust enough in the RANCH study. Unfortunately, most of these possible explanations cannot be further investigated with the RANCH data.

C. Annoyance reactions of children and parents

In general, the exposure-response relationships of children and their parents display a comparable trend in spite of some significant differences: children have lower response frequencies of severely annoyed than their parents at higher noise levels. This is consistent with earlier findings from the Tyrol Mountains study[18] which investigated the relationship between road and rail traffic noise and annoyance in children and their parents.

Possible explanations for the difference between children and adults' annoyance response to noise could be sought in nonacoustical factors[34] such as noise sensitivity, attitudes toward the noise source, perceived control, expectations, and coping behavior. Boman and Enmarker[35] observed that teachers were more annoyed due to road traffic noise than children and perceived the noise of road traffic noise to be more unpredictable than their pupils. In addition, the teachers described themselves as more sensitive to noise than the children. Conversely, teachers perceived more personal control over the noise than did the children. Unfortunately, the RANCH data do not enable us to analyze the influence of such nonacoustical factors.

The observation that children have significantly lower responses than their parents at higher noise levels could also mean that children are more sensitive at lower noise levels and that children's annoyance at higher noise levels is less influenced by nonacoustical factors than the annoyance of adults. To what extent this is the case for children cannot be determined based on the RANCH data.

D. Strengths and limitations

This study represents an improvement on previous studies[12–19] due to its large sample size both in the number of participants and number of schools. Despite the heterogeneity of the countries, results for noise and annoyance were rather similar across the three countries; i.e., we noted cross-cultural replication of the findings. A further strength of this study is the comprehensive inclusion of potential confounders and determinants. The hierarchical structure of the data (children within schools) has been taken into account, which was not the case in analyses of previous studies. The participants were distributed over a broad exposure range, and a continuous noise exposure measure was used, adding to the statistical power of the study. Most studies investigating the impact of noise exposure have involved between-group comparisons (high versus low): results of these studies may be sensitive to decisions about cutoff points used to categorize continuous exposure variables and the method used to assign scores to exposure categories.[36]

As already indicated the estimation of exposure to road traffic noise remains problematic: during their time at school, road traffic noise exposure changes as children move to a different classroom each year. Thus, the road traffic noise levels at the façade of their current classroom might not reflect the average level of exposure during their time at school.

E. Implications

The WHO guidelines for noise suggest that children are more sensitive to noise than adults because they are exposed to noise during critical developmental periods.[1] Children may also have fewer possibilities for controlling noise or have a less developed coping repertoire than adults.[5] However, we found that the exposure-response relationships for children were broadly comparable to those for their parents; if anything, the frequencies of severe annoyance at high exposures were lower among the children. Furthermore, annoyance is not the only indicator of the impact on children's health and well-being due to community noise. As demonstrated in the different publications of the RANCH study, cognitive,[25,30] behavioral,[37] and physiological measures[38–40] are necessary to fully describe the impact of environmental noise on children. For annoyance, the WHO guidelines recommend a L_{Aeq} of 55 dB for noise from external sources outdoors at school during play and for noise outdoors in the living area.[1] Our results (Figs. 1 and 2) indicate that some children were already severely annoyed due to aircraft noise at home and at school at lower levels ($L_{Aeq,7–23.outdoors}$ 45 dB), which suggests that the WHO community guideline values should be lowered to protect these children.

V. CONCLUSIONS

Children's annoyance can be reliably measured within a questionnaire. Exposure-response relationships were demonstrated between aircraft and road traffic noise exposure and severe annoyance among primary school children. Although

children were less annoyed at levels above 55 dB, these relationships were broadly comparable to those among their parents.

ACKNOWLEDGMENTS

Funding was provided by the European Community (QLRT-2000-00197), DEFRA in the UK, the Dutch Ministries of VROM, VWS, and VW. We thank the children, their parents, and the schools, our colleagues at RIVM, Queen Mary, and CSIC for helping with collecting the data, and the other members of the RANCH team for their consultation about the study. None of the authors have any competing financial interests.

[1]World Health Organization, *Guidelines for Community Noise*, edited by B. Berglund, T. Lindvall, D. H. Schwela, and K. T. Goh, World Health Organization, Geneva, 1999.
[2]R. F. S. Job, "Noise sensitivity as a factor influencing human reactions to noise," Noise Health **3**, 57–68 (1999).
[3]R. Guski, U. Felscher-Suhr, and R. Schuemer, "The concept of noise annoyance: How international experts see it," J. Sound Vib. **223**, 513–527 (1999).
[4]J. M. Fields, R. G. deJong, T. Gjestland, I. H. Flindell, R. F. S. Job, S. Kurra, P. Lercher, M. Vallet, T. Yano, R. Guski, U. Felscher-Suhr, and R. Schumer, "Standardized general purpose noise reaction questions for community noise surveys: Research and recommendation," J. Sound Vib. **242**, 641–679 (2001).
[5]M. L. Bistrup, "Prevention of adverse effects of noise on children," Noise Health **5**, 59–65 (2003).
[6]T. J. Schultz, "Synthesis of social surveys on noise annoyance," J. Acoust. Soc. Am. **64**, 377–405 (1978).
[7]S. Fidell, D. S. Barber, and Th. J. Schultz, "Updating a dosage-effect relationship for the prevalence of annoyance due to general transportation noise," J. Acoust. Soc. Am. **89**, 221–233 (1991).
[8]L. S. Finegold, C. S. Harris, and H. E. von Gierke, "Community annoyance and sleep disturbance: Updated criteria for assessing the impacts of general transportation noise on people," Noise Control Eng. J. **42**, 25–30 (1994).
[9]H. M. Miedema and H. Vos, "Exposure-response relationships for transportation noise," J. Acoust. Soc. Am. **104**, 3432–3445 (1998).
[10]H. M. E. Miedema and H. Vos, "Noise annoyance from stationary sources: Relationships with exposure metric day-evening-night level (DENL) and their confidence intervals," J. Acoust. Soc. Am. **116**, 334–343 (2004).
[11]H. M. E. Miedema and C. G. M. Oudshoorn, "Annoyance from transportation noise: Relationships with exposure metrics, DNL and DENL and their confidence intervals," Environ. Health Perspect. **109**, 409–416 (2001).
[12]P. Lercher, "Annoyance, disturbance and severances in children exposed to transportation noise," in *Proceedings of the Eighth International Congress on Noise as a Public Health Problem*, 29 June–3 July 2003, edited by R. G. de Jong, T. Houtgast, E. A. M. Franssen, and W. Hofman, (International Commission on Biological Effects of Noise, Rotterdam, The Netherlands, 2003), pp. 241–248.
[13]G. W. Evans, S. Hygge, and M. Bullinger, "Chronic noise and psychological stress," Psychol. Sci. **6**, 333–338 (1995).
[14]G. W. Evans, M. Bullinger, and S. Hygge, "Chronic noise exposure and physiological response: A prospective study of children living under environmental stress," Psychol. Sci. **9**, 75–77 (1998).
[15]M. M. Haines, S. A. Stansfeld, R. F. S. Job, B. Berglund, and J. Head, "Chronic aircraft noise exposure, stress responses, mental health and cognitive performance in school children," Psychol. Med. **31**, 265–277 (2001).
[16]M. M. Haines, S. A. Stansfeld, S. Brentnall, J. Head, B. Berry, M. Jiggins, and S. Hygge, "The West-London School Study: The effects of chronic aircraft noise exposure on child health," Psychol. Med. **31**, 1385–1396 (2001).
[17]M. M. Haines, S. A. Stansfeld, R. F. S. Job, B. Berglund, and J. Head, "follow-up of chronic aircraft noise exposure on child stress responses and cognition," Int. J. Epidemiol. **30**, 839–845 (2001).
[18]P. Lercher, G. Brauchle, W. Kofler, U. Widmann, and M. Meis, "The assessment of noise annoyance in schoolchildren and their mothers," in *Proceedings of the 29th International Congress and Exhibition on Noise Control Engineering*, 27–30 August 2000, edited by D. Casserau (Société Française d'Acoustique, Nice, France, 2000), Vol. **4**, pp. 2318–2322.
[19]H. Sukowski, M. Meis, P. Lercher, D. Heydinger, and A. Schick, "Noise annoyance of children exposed to chronic traffic noise: Results from the Tyrol School Study II," in *Contribution to Psychological Acoustics: Results of the eighth Oldenburg Symposium on Psychological Acoustics*, edited by A. Schick, J. Hellbrück, H. Höge, M. Klatte, G. Lazarus-Mainka, M. Meis, C. Reckhardt, K. P. Walcher, and R. Weber Bibliotheks—und Informationssystem der Universiät Oldenburg, 2000, pp. 571–580.
[20]B. Berglund, T. Lindvall, and M. E. Nilsson, "The children's community response to environmental noise," Revista de Acústica **38**, Paper ENV 02-003(2007).
[21]ISO/TS 15666, *Acoustics—Assessment of noise annoyance by means of social and socio-acoustic surveys*, International Organization for Standardization, Geneva, Reference No. ISO/TC 43/SC 1 N 1313, 2003.
[22]M. M. Haines, S. L. Brentnall, S. A. Stansfeld, and E. Klineberg, "Qualitative responses of children to environmental noise," Noise Health **5**, 19–30 (2003).
[23]I. van Kamp, A. E. M. Franssen, and R. G. de Jong, "Indicators of annoyance: A psychometric approach; the measurement of annoyance and interrelations between different measures," in *The 2001 International Congress and Exhibition on Noise Control Engineering* (Acoustical Society of The Netherlands, The Hague, The Netherlands, 2001), pp. 27–30.08.
[24]I. van Kamp, *Coping With Noise and Its Health Consequences* (Styx & pp, Groningen, 1990).
[25]S. A. Stansfeld, B. Berglund, C. Clark, I. Lopez-Barrio, P. Fischer, Ohrström, E., M. M. Haines, J. Head, S. Hygge, I. van Kamp, and B. F. Berry, "Aircraft and road traffic noise and children's cognition and health: A cross-national study," Lancet **365**, 1942–1949 (2005).
[26]*Calculation of Road Traffic Noise (CRTN)*. (HSMO, London, 1998).
[27]A. G. Dassen, J. Jabben, and J. H. J. Dolmans, "Development and use of EMPARA: A model for analysing the extent and effects of local environmental problems in the Netherlands," in *The 2001 International Congress and Exhibition on Noise Control Engineering*, (Acoustical Society of The Netherlands, The Hague, The Netherlands, 2001), pp. 27–30.08.
[28]I. Enmarker, and E. Boman, "Noise annoyance responses of middle school pupils and teachers," J. Environ. Psychol. **24**, 527–536 (2004).
[29]M. Haines, and S. Stansfeld, "Measuring annoyance and health in child social surveys," in *The 2000 International Congress and Exhibition on Noise Control Engineering*, (Société Française d'Acoustique, Nice/France, 2000).
[30]C. Clark, R. Martin, E. van Kempen, T. Alfred, J. Head, H. W. Davies, M. M. Haines, I. Lopez-Barrio, M. Matheson, and S. A. Stansfeld, "Exposure-effect relations between aircraft and road traffic noise exposure at school and reading comprehension. The RANCH project," Am. J. Epidemiol. **163**, 27–37 (2006).
[31]R. Guski, "How to forecast community annoyance in planning noisy facilities," Noise Health **6**, 59–64 (2004).
[32]European Commission Working Group Assessment of Exposure to Noise (WG-AEN), "Good practice guide for strategic noise mapping and the production of associated data on noise exposure," Position paper, Version **2**, 13 August 2007, Report No. WG-AEN 004.2007. doc, available at http://ec.europa.eu/environment/noise/pdf/gpg2.pdf. Last viewed 1/9/2009.
[33]H. A. Nijland and G. P. van Wee, "Traffic noise in Europe: A comparison of calculation methods, noise indices and noise standards for road and railroad traffic in Europe," Transport Rev. **25**, 591–612 (2005).
[34]P. J. M. Stallen, "A theoretical framework for environmental noise annoyance," Noise Health **3**, 69–79 (1999).
[35]E. Boman, and I. Enmarker, "Factors affecting pupils' noise annoyance in schools: The building and testing of models," Environ. Behav. **36**, 207–228 (2004).
[36]D. B. Richardson and D. Loomis, "The impact of exposure categorisation for grouped analyses of cohort data," J. Occup. Environ. Med. **61**, 930–935 (2004).
[37]S. A. Stansfeld, C. Clark, R. M. Cameron, M. M. Haines, I. van Kamp, E. van Kempen, and I. Lopez-Barrio, "Aircraft and road traffic noise exposure and children's mental health in the RANCH study," in *Proceedings of the 33rd International Congress and Exposition on Noise Control Engineering*, 22–25 Aug. (Czech Acoustical Society, Prague, 2004).
[38]E. van Kempen, I. van Kamp, P. Fischer, H. Davies, D. Houthuijs, R. Stellato, C. Clark, and S. Stansfeld, "Noise exposure and children's blood

pressure and heart rate: The RANCH-project," J. Occup. Environ. Med. **63**, 632–639 (2006).
[39]A. G. Gunnarsson, B. Berglund, M. Haines, M. E. Nilsson, and S. A. Stansfeld, "Psychological restoration in noise-exposed children," *Proceedings of the eighth International Congress on Noise as a Public Health Problem*, 29 June–3 July 2003. (International Commission on Biological Effects of Noise, Rotterdam, 2003), pp. 159–160.
[40]E. Öhrström, E. Hadzibajramovic, M. Holmes, and H. Svensson, "Effects of road traffic noise on sleep: Studies on children and adults," J. Environ. Psychol. **26**, 116–126 (2006).

Response to a change in transport noise exposure: Competing explanations of change effects

A. L. Brown[a)]
Urban Research Program, Griffith School of Environment, Griffith University, Nathan, Brisbane, Queensland 4111, Australia

Irene van Kamp[b)]
Centre of Environmental Health Research, National Institute for Public Health and the Environment, P.O. Box 1, 32700 BA Bilthoven, The Netherlands

(Received 21 April 2008; revised 24 September 2008; accepted 6 December 2008)

Annoyance response to a change in noise exposure appears to demonstrate an excess response relative to those predicted from exposure-response curves obtained under steady-state conditions. This change effect also appears to persist well after the change. Numerous explanations have been postulated for this phenomenon. This paper catalogs the different explanations and reviews the evidence for each. The evidence is of limited and variable quality but, while inadequate to endorse any one explanation, is sufficient to reject some notions and to identify a residual set of plausible explanations. These include two explanations based on modifiers of exposure-response relationships that potentially change between before and after conditions, an explanation based on differential response criteria of respondents chronically exposed to different steady-state levels of noise, and an explanation based on retention of coping strategies. All have ramifications for the assessment of human response (annoyance) where noise exposure changes, and some have wider implications for the interpretation of generalized exposure-response curves obtained in the steady state.
 [DOI: 10.1121/1.3058636]

PACS number(s): 43.50.Qp, 43.66.Lj [BSF] Pages: 905–914

I. INTRODUCTION

The literature of human response to a change in transport noise exposure suggests that human response to changed exposure includes both an *exposure effect* and a *change effect*. The change effect is manifest as an excess response to the new noise exposure over that predicted from steady-state exposure-response curves. Excess response was found unambiguously for changes in road traffic noise where the change in exposure resulted from an increment or decrement in source levels (termed type 1 changes) rather than from the insertion of barriers or other path mitigation interventions (type 2 changes) (Brown, 2009).

This paper catalogs the surprising number of putative explanations advanced for the phenomenon of change effect and reviews evidence for each of the explanations. For some of the explanations, this often required identification and, where possible, clarification of confusion and ambiguity in various concepts and terms found in the literature on response to change.

Why excess response has spawned so many alternative explanations is unclear, but given that situations where noise exposure changes as a result of infrastructure changes are contentious, explication of the change effect and its implications for assessing change is of practical importance to communities and authorities. Further, some have implications for interpreting exposure-response relationships obtained under steady-state noise conditions. These need to be considered alongside other current debates concerning the *durability* of generalized exposure-response curves and assessing (trends in) annoyance (for example, Guski, 2004; Fidell and Silvati, 2004; van Kempen and van Kamp, 2005; Brooker, 2008).

A. Explanations of the change effect

The explanations of change effects are diverse. While we use the term "explanations," some are more appropriately seen as opinions. Others call on theory. Suggested explanations are discussed in Secs. II A–II K:

A. Any change effect is only *transient* as people *adapt* to the change.
B. Respondents' *anticipation* or *expectation* of change.
C. Respondents' *attitudes toward the source/authorities* change.
D. The combined effects of changes in other environmental attributes—an *area effect* or a *halo effect*, though we prefer the term *surrogate effect*.
E. *Demand-response bias* generated by repeated questioning of respondents.
F. Found in *adaptation-level theory*.
G. *Partial retention of behavioral coping strategies*.
H. *Differential response criteria* (*response bias*) in responding to annoyance scales.
I. *Memory distortion*.
J. *Self-selection*.
K. While not explaining the change effect itself, *perceptual*

[a)]Electronic mail: lex.brown@griffith.edu.au
[b)]Electronic mail: irene.van.kamp@rivm.nl

constancy or loudness constancy may explain differences between type 1 and type 2 changes.

Below we examine the diverse claims and counterclaims for the different explanations and evidence from empirical studies and seek to clarify their underlying mechanisms.

II. THE EXPLANATIONS

A. Any change effect is only transient as people adapt to the change

There is a prevailing notion (part of the folklore of noise) that people adapt, *habituate*, or *get used to* new conditions after a change in exposure. There is confusion in the use of these terms. *Habituation* is a decrease in response to a stimulus after repeated presentations; *adaptation* an adjustment of response to a stimulus to reach a new equilibrium. In the literature of environmental noise, these terms have generally been applied without discrimination and without necessarily referring to component emotional, physiological, behavioral, and cognitive dimensions. Responses in change situations have been described as transient, and respondents assumed, in lay terms, to get used to the new conditions, but mechanisms for these processes/outcomes are not invoked. *Desensitization* of response to a noise stimulus (Raw and Griffiths, 1985) is another "adaptation-related" term and is included in several of the explanations described below. Weinstein (1982) used the antonym *sensitization* for an increase in response over time.[1] Sensitization is an example of nonassociative learning in which the progressive amplification of a response follows repeated administrations of a stimulus (Bell *et al.*, 1995). As an example of claims for adaptation to noise, Nimura *et al.* (1973) implied that it was adaptation nearer the longer established of two new Shinkansen rail lines that was responsible for a much lesser response beside an older line as against one opened more recently.

There is ambiguity in the way different authors define adaptation. Particular interest here is in whether there is adaptation of any observed change effect, and we use the definition in Horonjeff and Robert (1997). Figure 1 illustrates two hypothetical chronologies of response to a step increase in noise exposure. The change effect is the difference between the attitudinal response after the change and the long-term baseline response to the new after-change exposure conditions (expected from steady-state exposure-response curves). The solid-line trajectory illustrates adaptation of the change effect, with regression of response to the *new* baseline level for the postchange exposure. The broken line trajectory suggests a persistent change effect, with no long-term adaptation.

Weinstein (1982) found that after-change responses remained constant between 4 and 16 months after the change following a large increment in noise exposure (no examination for change effect). By contrast, Lambert *et al.* (1996) reported adaptation near a new TGV line, but their approach was to ask respondents if they had become used to the noise 3–4 years after the line opened (85% indicated that they had, 75% within 1 year). This implicitly defines adaptation as responses returning to the *prechange* conditions, not the expected *postchange* steady-state response as in Fig. 1. There is a related ambiguity in studies that use self-reported adaptation as, for example, in the study of Hatfield *et al.* (2001) on response to aircraft noise in Sydney where respondents were asked if they were *more used to* (24% indicated that they were) or *more sensitive to* the noise. These very different definitions of adaptation are what Fields and Hall (1987) described as using the same label for measuring quite different concepts. The conclusion is that caution is required in synthesizing across different studies that reported that adaptation was/was not present.

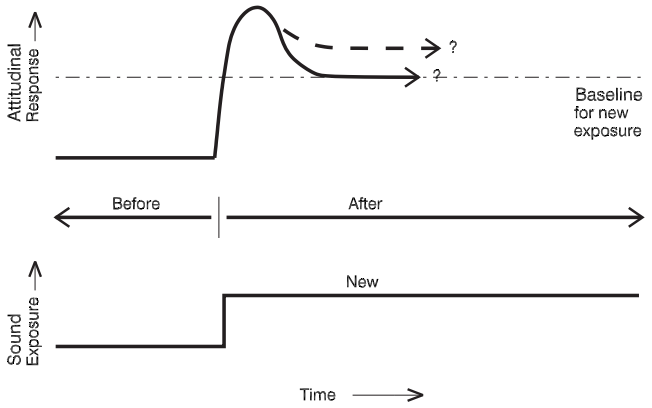


FIG. 1. Hypothetical chronologies of attitudinal response to a step-change in exposure. The light horizontal broken line is the expected response to the after-change conditions (from steady-state exposure-response curves). The two attitudinal response trajectories shown in bold are different possible responses to an increase in noise exposure (after Horonjeff and Robert, 1997).

The weight of evidence in empirical studies was that there is no adaptation of the change effect months, even years, after the change. Griffiths and Raw (1989) did find partial attenuation 7–9 years out, and Breugelmans *et al.* (2007) found that the change effect temporarily disappeared at the fifth of sixth resurveys of a panel around Schiphol but returned at the sixth resurvey. Most of the available results confirm the broken trajectory of Fig. 1 in the longer term. Could there be adaptation of at least part of the change effect immediately after the change, as suggested in Fig. 1? This has been postulated by various authors (Hatfield *et al.*, 2001; Jonsson and Sörensen, 1973; Langdon and Griffiths, 1982; Lawson and Walters, 1973), but there is no empirical evidence of adaptation to the change effect in days or weeks following a change. Evidence of very rapid adaptation to noise, within hours and minutes, is primarily based on laboratory studies of perceptual change (e.g., Quehl and Basner, 2006), not field studies, but has no clear relevance in the current context. Overall, there is no evidence that the change effect is a transient phenomenon.

B. Anticipation or expectation of change

It is suggested that respondents' expectation (or anticipation) of change could explain the change effect. We note, at the outset, Guski's (1999) statement that, "There are no meaningful data on this topic." At issue is that the purported

anticipation was sometimes merely assumed and that there is no standardization of what it is that respondents anticipate/expect. Different concepts have been used: assumed *expectation of increased annoyance* (Öhrström, 1997; van Dongen and van den Berg, 1983), measurements variously operationalized as *expectation of an improvement, deterioration, or remaining the same* (Job *et al.*, 1996), *expected effect of a change* (Mackie and Davies, 1981), and expectation of increased annoyance (Schreckenberg *et al.*, 2001); and measurements, longitudinally, of *expectation about future noise levels* (Breugelmans *et al.*, 2007). The assumption that all residents, living in an area that is known (by authorities, through the media, etc.) to be one in which noise will change, will share these expectations is open to question. Flindell and Witter (1999) provided evidence from Heathrow that a minority of respondents were actually aware of proposed changes despite prospective changes to aircraft flight patterns being extensively publicized.

Observations on the role of expectation in situations of change include the following: Öhrström (1997) speculated that expectation of an increase in annoyance could explain more annoyance before extension of a railway line than 3.5 years after; van Dongen and van den Berg (1983) suggested that expectations regarding a new railway line were the cause of before-change annoyance being higher than the actual annoyance levels afterward; Mackie and Davies (1981) provided empirical evidence that people expect changes to have more effect than they actually do; and Schreckenberg *et al.* (2001) reported that before a change, respondents' expectation of what annoyance would be after the change was higher than both before-change annoyance and after-change annoyance (actual change in noise levels proved to be zero). None of these observations shed any light on the role of expectation in explaining a change effect.

Recent evidence from Schiphol does not support this explanation; observed excess response in a subgroup experiencing an increase became apparent only after the new runway became fully operational despite prior extensive media coverage of this upcoming change (Breugelmans *et al.*, 2007; Ministry of Transport, Public Works and Water Management, 2005). Repeat measurements of expectations regarding future noise levels (improvement versus deterioration) were also available in this study. Expectation of a deterioration was shown to be a strong predictor of annoyance (Houthuijs *et al.*, 2007), but more detailed analyses would be needed to evaluate the precise influence of expectations (independently or via changed attitudes) on the measured change effect.

Based on the lack of other evidence and on the Schiphol study results, expectation *per se* has little support as direct explanation of the change effect. However, data from the Sydney Airport study (Job *et al.*, 1996) do provide a potential link between expectation and another explanation—change in attitudes toward the authorities/source (see Sec. II C). If certain attitudes do change in situations where noise exposure changes, the role of expectation may partly be to shift the occurrence of attitudinal change, temporally, to before the change in exposure.

C. Attitudes toward the source/authorities change

A range of attitudes correlate with annoyance responses (Fields, 1993). This explanation postulates that certain attitudes (to authorities and to noise preventability) may change, becoming more negative where noise increases and more positive where noise levels decrease, thus modifying the exposure-effect relationship and resulting in the observed excess response in change studies (Job, 1988b). Langdon and Griffiths (1982) noted such a possibility for residents' perception of the public policy of authorities and that this possibility was first suggested by Scholes (1977). Fidell *et al.* (1985) suggested that heightened community awareness might plausibly be sufficient to account for a greater prevalence of self-reported annoyance.

Kastka (1981) could not attribute large excess response solely to the change in noise level, and based on no overall correlation between the magnitude of the reported effect at his 50 change sites and the change in levels at those sites, he suggests, though without evidence, that this may be due to goodwill to the authorities resulting from advertising campaigns for traffic calming schemes (also see Schreckenberg *et al.*, 2001). Fields and Hall (1987) noted that there is some limited empirical evidence that attitudes of a population can be manipulated so as to alter annoyance responses. A more recent experiment (Djokvucic *et al.*, 2004) indicates that (manipulated) attitude toward the source does indeed directly influence responses to noise independent of the noise effect. Performance impairment was shown to be indirectly influenced by (manipulated) attitude via reaction. Likewise, Maris *et al.* (2007) demonstrated that fairness of the exposure procedure (sound management) can be used as an instrument to reduce noise annoyance. Raw and Griffiths (1990) expressed a contrary view regarding the role of attitudes in the change effect, based on no observed differences in the exposure-response relationship for before conditions between their increase sites and decrease sites. They noted that it would also be necessary to assume that similar changes in attitude occurred in all of their field sites (with very different characteristics and histories), and these would have to be maintained for years after the change.

Is there evidence of attitudes toward the source/authorities changing with a change in exposure and, if so, any evidence that these are linked to measured change effects? In attempting to address the first of these questions, Job *et al.* (1996) and, more recently, Schreckenberg and Meis (2007) suggested that change in attitude can be observed in locations where changes to future noise exposures have been forecast. Job *et al.* (1996) found that negative attitudes were higher in areas near the Sydney Airport in which an increase in noise was *anticipated* (by the authorities) and lower in areas where a decrease was anticipated. Schreckenberg and Meis (2007) reported that for the Frankfurt Airport, where a fourth runway was being planned, (mis)trust of the aviation industry was positively related to annoyance. However, both of these were cross-sectional studies *before* change occurred, not longitudinal evidence of attitudes changing systematically with changes in exposure. In fact, they provided no evidence that attitudes to the source/authorities actually

changed, only that there was a difference in these attitudes in different areas (and for the Sydney data, a lack of clarity regarding their actual noise exposures) or at different levels of annoyance.

Houthuijs *et al.* (2007) provided data in which the link of these attitudes to change effects can be tested. A longitudinal study around the Schiphol Airport commenced a year before the planned opening of a new runway through three years afterward, with yearly repeated measures in a panel ($N=650$) that was part of a large before-after survey of some six thousand respondents. They reported a systematic relationship between noise exposure and a composite measure of attitudes toward the airport and its expansion. Breugelmans *et al.* (2007) used the panel data from this study to analyze the role of changes in nonacoustical factors—including repeated measurements of attitudes to the authorities, the airport, and its expansion—reporting that the latter did not explain the excess response in the noise-increase subgroup of the panel. They concluded that the excess response can be considered to be a (change in) noise effect. Their model for repeated measurements included the noise level exposure (L_{den}) as well as the change in noise level (Delta L_{den}), as predictors of response.

Despite strong evidence from Breugelmans *et al.* (2007) that changes in negative attitudes cannot explain observed excess response, the emphasis that has been given to this explanation in the past suggests that it should perhaps not be rejected at this stage without further confirmation

D. Surrogate effect

Reductions or increases in road traffic noise exposure may be associated with parallel changes in air pollution, congestion, accidents, housing value, presence of trucks, and general appearance of the area. Similarly, changes in aircraft noise exposure may be associated with changes in fear of crashes. The noise stressor will therefore change in association with change in many other stressors, and the combination may contribute to the observed change effects in noise response.

A credible mechanism for this explanation would be through these changes altering respondents' overall opinion of neighborhood quality—a known modifier in the exposure-response relationship (Langdon, 1976). Vincent and Champelovier (1993) provided evidence of a quantitative increase in perception of neighborhood quality with the installation of a noise barrier, as did Öhrström (2004) following a major reduction in traffic flows. However, Griffiths and Raw (1989) did not find a difference in opinion of the area between newcomers (who had relocated to the area in the previous six years) and long-term residents who had experienced the change.

We suggest that this explanation be termed a surrogate effect, though a variety of terms has been used. Klæboe *et al.* (1998) explained excess response to traffic improvements in Oslo in terms of an *areawide* effect—simultaneous improvements in air pollution and volumes of road traffic amongst other factors (see also Klæboe *et al.*, 2000). The Brown (2009) finding of a large excess response to noise in the major traffic improvement study by Öhrstöm (2004) could also be explained by the simultaneous improvements in other environmental factors as well as a change in noise exposure. A surrogate effect explanation would also fit the excess-response change effect reported by Kastka (1981). Horonjeff and Robert (1997) referred to Kastka's finding as a halo effect. The surrogate effect explanation remains a plausible mechanism to explain excess response.

E. Demand-response bias generated by repeated questioning

This explanation suggests that where panel designs involving repeat interviews of the same respondents are used, excess response may be an artifact of repeat interviewing. Repeat questioning may lead to demand-response bias or response set.

There is experimental evidence that demand-response bias does not occur in repeat measures of annoyance responses. Fields *et al.* (2000) analyzed a range of panel studies and concluded that they do not appear to introduce survey-resurvey bias in noise response, particularly if repeat surveys are at least 1 month apart. Jonsson and Sörensen (1973) noted work in which measurements on the same individuals on two occasions, and also on a control group on the second occasion, show no demand-response bias. Fidell *et al.* (1985) reported that after-change annoyance responses (for short-term annoyance) were similar over three repeat rounds of interviews (in what they regarded as being tantamount to a panel study) conducted over 3 months. They concluded that personal decision criteria for selecting annoyance response categories are stable over time and rounds of interviews (for relatively constant noise exposures)—a finding of no demand-response bias. These were the same data in which a reanalysis by Raw and Griffiths (1985) found excess-response change effects—hence the latter was not an outcome of demand-response bias.

Job (1988b) noted that demand-response bias could be a particular problem where a panel is interviewed before and after the changes. Respondents are likely to feel that the interviewer is expecting (demanding) a changed reaction—suggesting any observed change effect would not be "genuine." However, Weinstein (1982) used a mixed panel design with repeat measures on the same respondents, supplemented by "one occasion" respondents—controlling for the possibility that initial contact with the panel group changes how it reacts. At 16 months after the change, the control group reported the same noise disturbance as the panel group. Fidell and Jones (1975) also found no difference in annoyance responses between a panel (interviewed three times by telephone before and after a change in flight paths at the Los Angeles airport) and independent control samples at the first and third interviews. The evidence suggests that demand-response bias generated by repeated questioning is unlikely to be the cause of observed excess-response change effects.

F. Adaptation-level theory

The adaptation-level explanation hypothesizes that people who have been living in an area with high noise ex-

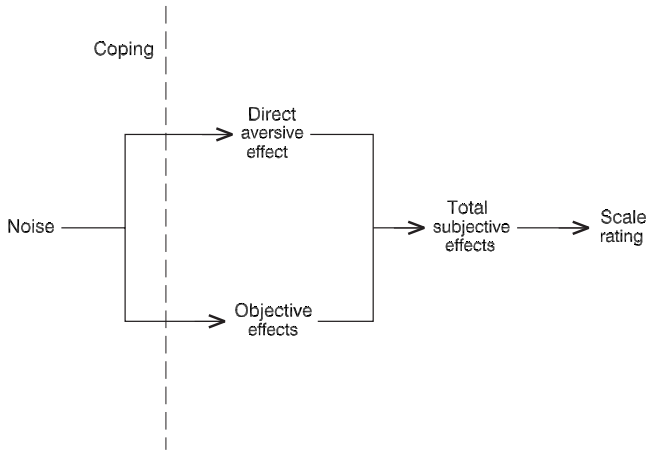


FIG. 2. The model for the retention of behavioral coping strategy explanation. Total subjective effects include direct aversive effect (annoyance) as well as activity interferences (objective effects), both filtered by coping strategies. (Source: Raw and Griffiths, 1990).

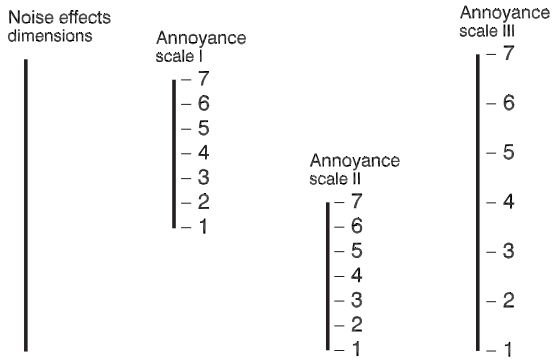


FIG. 3. The differential response criteria model suggests a different scaling of noise effects by respondents chronically exposed to high noise levels (annoyance scale I) and to low noise levels (annoyance scale II). After Brown (1987).

posures for some time change their expectation about the noise environment. Weinstein (1982), citing Helson (1964) and others, noted that there is a large body of research that people making perceptual judgments are influenced by the conditions to which they have been exposed previously, with their response to a stimulus a measure of the deviation of that stimulus from an optimal level of stimulation. This optimal level of stimulation is termed adaptation level. Respondents who are chronically exposed to high levels of noise would become desensitized to the exposure, experiencing reduced effects. While it could explain excess response in a single-site study using retrospective assessments (Brown *et al.*, 1985), it did not in studies where before and after measurements were used (Raw and Griffiths, 1990; Brown, 1987). Adaptation-level theory can be discarded as an explanation for excess response.

G. Partial retention of behavioral coping strategies

Raw and Griffiths (1990) proposed this explanation, observing that excess response is much less evident (see also Babisch and Gebhardt, 1986) in objective measurements of activity interference than in the subjective measurements of annoyance. To explain this, they formulated a set of hypotheses in which coping plays a key part. Figure 2 shows the total subjective effect defined to include direct aversive effect (annoyance) as well as activity interferences (objective effects), both filtered by coping strategies.

Raw and Griffiths (1990) provided evidence that when respondents experience a change in exposure, they change some of their coping strategies, in particular those "noise mitigating behaviors" such as closing windows or changing use of rooms, but partially retain them after the change. After a decrease in exposure, fewer people might keep their windows closed, but partial retention of this strategy would result in an excess effect in total subjective effects—coping appropriate to a previous higher exposure would result in lower than expected annoyance (compared to respondents who had always been exposed to the same low levels and who would not have applied coping strategies to the same extent). The reverse would occur after an increase in exposure. Partial retention of behavioral coping strategies after a change would explain excess response in the annoyance scores, but Raw and Griffiths (1990) questioned why, over time, coping strategies would not continue to change to those appropriate to the new exposure, and the excess-response effect eventually disappears. They suggested, based on their evidence, that the relationship between objective effects of noise and total subjective effects is not constant before and after a change—objective effects (interferences) changed less markedly than subjective effects—and assumed that a relatively rapid sensitization to the increased noise stimulus (desensitization for a decrease in exposure) must occur. The retention of behavioral coping strategy explanation has not been subjected to further testing but, based on the limited evidence, remains a plausible explanation of the change effect.

H. Differential response criteria for the annoyance scale (response bias)

The differential response criteria explanation invokes a very different mechanism to that of the demand-response bias explanation—though often the literature on change does not differentiate these two bias mechanisms. This explanation was originally referred to as response bias (Brown, 1987) but differential response criteria is a better term. The differential response criteria explanation is a measurement error, distinct from the adaptation-level explanation, which required a real difference in sensory response to the stimulus.

The explanation requires the introduction of a "noise effect" dimension (as in the "total subjective effect" dimension of Raw and Griffiths (1990), Fig. 2). It suggests that annoyance rating scales might be used differently by respondents who are chronically exposed to different levels of steady-state noise exposure (Fig. 3)—scale I if they are chronically exposed to high levels and scale II if they are chronically exposed to low levels. Researchers invariably assume that response criteria for annoyance scales, or personal decision criteria (Fidell *et al.*, 1985), are the same across all respondents irrespective of their exposure, but Berglund *et al.* (1975) demonstrated that response criteria for annoyance scales are not independent of the noise condition.

The explanation suggests that when chronically exposed respondents experience a change, they may expand the annoyance scale to cover more of the range of the noise effect dimension, adopting postchange response criteria for the annoyance scale (say, scale III). After a reduction in exposure, the model suggests that change respondents might use scale III to scale their noise effects. They would thus report lower annoyance scores than would be reported by people chronically exposed to equivalent after-change levels. Both change respondents and chronically exposed respondents would be experiencing the same distribution of noise effects (as they are exposed to equivalent levels of noise), but the latter would scale these effects using scale II. A corresponding explanation applies to an increase in noise exposure—the model satisfactorily explaining change effects for both decrements and increments.

The logic of an ability to adjust scales according to experience is illustrated by a simple, but not trivial, example. Say, a person considered him/herself adversely affected by current long-term exposure (though only subject to moderate levels of noise) and, accordingly, responded to a presented annoyance scale with the maximum score. If this person were then to be subject to a significant noise increase (with consequent increase in noise effects), he/she would have no alternative, when presented with the same annoyance scale, but to respond again with a maximum score. The person's score has not changed, and this same score now represents greater noise effects on the individual (that is, noise effects have been scaled differently). A reviewer of this paper suggested that the person in this example illustrates an underreaction, not an excess-response, change effect. But such a conclusion is based on a different comparison—comparing the person's annoyance score after the change to that from before the change. The germane comparison for the differential response criteria explanation is of annoyance scores (after the change) of those who have experienced the change and annoyance scores (in the steady state) of people chronically exposed to equivalent after-change exposure levels. The latter yields an excess-response change effect.

This model was developed, *post hoc*, as a possible explanation of observed change effect in small studies of change in exposure (Brown, 1987; Brown *et al.*, 1985). Its only subsequent testing was by Raw and Griffiths (1985), who suggested that the model was wanting in terms of explaining their own excess-response results. However, the mathematical analysis they adopted was flawed in their assumption of a constant and small range for the steady-state and postchange annoyance scales, not an expanded annoyance scale postchange (scale III in Fig. 3). Had their analysis not been limited in this way, the model would have fitted all of their empirical results.

The model may also explain why activity interferences may not display excess response as do annoyance scores. In contrast to the differential response criteria bias in the annoyance scores, self-reports of activity interferences can be expected to be the same for change respondents and chronically exposed respondents at the same level of exposure.

I. Memory distortion

A small number of change studies (Lambert, 1978; Brown *et al.*, 1985; Mehra and Lutz, 2000) used retrospective assessment of before-change response to examine the effects of change. Raw and Griffiths (1990) noted that retrospective assessments cannot substitute for studies using carefully matched pairs, or experimental and control sites, in true before and after designs.

Retrospective assessments clearly have the disadvantage of possible memory distortion in reported effects. Two studies utilized both before and after measurements of response as well as retrospective assessment of before-change conditions. Brown (1987) found that respondents reported a very different distribution of annoyance scores before a change (increase) to what they reported, retrospectively, of their prechange condition. Lambert (1978) found that the mean retrospective-score of before-change conditions was slightly higher than the mean before-score, though only marginally so. In both studies, the sample size was small.

While retrospective data on before levels and *post hoc* self-reported adaptation can be considered as weak measures [self-reports of adaptation also potentially include memory distortion (see Sec. II A)], as most studies in which a change effect has been observed have not used retrospective assessments, they can be discarded as an explanation of the change effect.

J. Self-selection

This explanation suggests that it is only high noise level situations that receive interventions to reduce noise exposure (Baughan and Huddart, 1993), and studies of change that involve a decrease in noise are thus a biased sample of sites in the community. Fidell *et al.* (2000) and Mackie and Davies (1981) argued that people with high noise sensitivity will have already self-selected themselves out of such highly exposed sites. The self-selection explanation for excess response suggests that such movement will result in sites where respondents have lower average noise sensitivity. After a change to lower noise levels, these less vulnerable (that is, less noise sensitive) respondents will report much lower annoyance scores than predicted by exposure-response curves, resulting in a change effect (Weinstein, 1978; Evans *et al.*, 1998).

However, a self-selection hypothesis requires that for a situation of unchanging high noise exposure, there should be a negative correlation between annoyance and length of residence. There is no evidence of this correlation. Weinstein (1982) indicated that the majority of noise studies have found no appreciable relationship between noise disturbance and length of residence, though the study of Griffiths and Langdon (1968) was an exception. Recent evidence of no self-selection effect can be found in Nijland *et al.* (2007). There is no convincing evidence supporting this explanation (Baughan and Huddart, 1993).

K. Perceptual constancy

Perceptual constancy or loudness constancy (Robinson *et al.*, 1963; Langdon and Griffiths, 1982) suggests that

TABLE I. Variables in the various change effect explanations. Subscripts B and A refer to conditions before and after the change in noise exposure. The total subjective effects of noise (AE) are not able to be measured directly.

				Noise effects		
	Exposure	Moderator variables	Behavioral coping strategies	Interference effects (e.g., sleep disturbance)	Total subjective effects	Reported scores (Annoyance scale scores)
Before change	EXP_B	MOD_B	COP_B	IE_B	AE_B	$ANNOY_B$
After change	EXP_A	MOD_A	COP_A	IE_A	AE_A	$ANNOY_A$

people respond according to their perception of the source levels of the noise rather than to what they experience after the noise levels have been reduced through some path attenuation. For example, it suggests that they respond to levels of outdoor aircraft noise rather than to the lower levels that they actually experience indoors as a result of attenuation of the building envelope.

Perceptual constancy could explain why type 2 changes (noise levels reduced by building insulation or by barriers) may be different from type 1 changes (noise levels reduced by changes in the source) in terms of evidence of excess-response change effects. Some authors (Langdon and Griffiths, 1982) refer specifically to this concept. Others, without using the term, have made related observations. Nilsson and Berglund (2006) noted that indoor annoyance was significantly reduced by the erection of a barrier though the effect of the barriers on the indoor sound levels was small, and Öhrström (2004) suggested that perception of noise and its disturbances *outdoors* contributes more to general annoyance reaction than indoor disturbances. Perceptual constancy remains a possible, though speculative, explanation for differences in change effects reported between type 1 and type 2 changes.

III. DISCUSSION

There is insufficient evidence to support a single theory, or explanation, of the change effect. Several of the explanations can be rejected based on a lack of supporting evidence, and several have been identified that warrant further consideration.

A. Explanations rejected

Adaptation/habituation is a convenient descriptor of a trajectory of response to change but has no value in the explication of a change effect. Even where described more formally in terms of adaptation-level theory, evidence that annoyance reactions do not diminish over time counts against this as an explanation for a change effect. There is also evidence that demand response resulting from repeated questioning of the same respondents in longitudinal panel surveys is unlikely to be the cause. Expectation has little support as a direct explanation of the change effect, though potentially may be associated with changing attitudes (if such attitudes actually change) before the noise exposure changes. There is no evidence for the self-selection hypothesis, and memory distortion can play no role where no retrospective judgments of response are utilized. Perceptual constancy may be a factor in explaining the difference in response effects resulting from type I and type II changes but not the excess response itself.

B. Categorization of the remaining explanations

The remaining four explanations fall into three categories based on distinctly different, but plausible, mechanisms:

- change effects resulting from a change in variables modifying the exposure-response relationship before and after the change,
- change effects resulting from measurement error in annoyance scales, involving a variable relationship between total subjective effects of noise and annoyance scores, and
- change effects resulting from the retention of behavioral coping strategies after a change.

While we distinguish between the different mechanisms, we do not discount the possibility, suggested by Baughan and Huddart (1993), that change effects may result from a combination of mechanisms.

1. Change in variables modifying the exposure-response relationship between the before and after exposure conditions

A range of nonacoustic variables that intervene in the exposure relationship for noise has been extensively measured, tested, and discussed in the literature (Schultz, 1978; Fields, 1993, 1994; Job, 1988a; Miedema and Vos, 1999, 2003; van Kamp *et al.*, 2004). Under steady-state noise exposure conditions, these variables are regarded as constant. Or at least it is assumed that they are constant, given little suggestion to the contrary. However, two of the explanations suggest that this assumption needs to be revisited in situations where noise levels change.

For each of the attitudes to the source/authorities explanation and the surrogate effect explanation, variables that are known to intervene in the exposure-response relationship, *attitudes toward the source* and *overall opinion of the neighborhood*, respectively, could change in a change situation. As suggested in Table I, where EXP_B and EXP_A are the noise exposure levels before and after a change, the values of

MOD_B may be very different from those of MOD_A, affecting the exposure-effect relationship differentially in the before-change and after-change conditions and hence in the reported scores ANNOY.

2. Changes in the relationship between effect of noise and self-reports of those effects

The differential response criteria explanation suggests that annoyance rating scales might be scaled differently by respondents who have been chronically exposed to different levels of noise. Differences in response criteria represent a measurement error in steady-state responses that only becomes apparent where people have experienced a change in exposure. In Table I, this would be represented by a relationship between AE_B, IE_B, and $ANNOY_B$ different from that between AE_A, IE_A, and $ANNOY_A$. This explanation also suggests why change effects may appear in the reported annoyance scores (ANNOYs) but not in the interference effects (IEs).

3. Retained coping strategies following a change

In the terms of Table I, the *retention of coping* explanation suggests that the coping strategies that people develop at different levels of noise exposure (COP) may be retained to some extent after a change in exposure. The after-change coping (COP_A) of those who have experienced the change will be different from the coping strategies of people chronically exposed to noise levels equivalent to the after-change levels.

C. Newcomers versus long-term residents

There is one matter still unresolved with respect to all of these explanations. If people change their attitudes and their coping behavior or adjust annoyance scales when noise levels change and if this is not a short-term phenomenon (and the evidence of persistence of change effects months and years after a change suggests that it is not), newcomers to an area following a change should report different annoyance responses from those who had lived there throughout the change. There is conflicting evidence. Griffiths and Raw (1989) reported that newcomers after a noise decrement were more annoyed than long-term residents who had experienced the change, and van Dongen and van den Berg (1983) found that a small group of newcomers after a noise increment was less annoyed than long-term residents. However, Klæboe *et al.* (1998) found that newcomers after a noise decrement were less annoyed than long-term residents—the opposite direction to that required by the explanations. Further longitudinal studies are necessary, which carefully compare the response of long-term residents who have experienced a change with those of newcomers.

IV. IMPLICATIONS OF THE EXPLANATIONS

Our examination of different explanations for the change-effect phenomenon raises two larger but interrelated questions. First, should excess response to change be of concern to policy makers? And if so should it be addressed in environmental assessments of infrastructure projects? Second, are there implications of the different potential explanations of change effects to the interpretation of existing exposure-response relationships for transportation noise?

A. Assessing change in exposure

The evidence of the magnitude and the persistence, over time, of the change effect (Brown, 2009) and the existence of plausible explanations for it suggest that it is a real effect and needs to be taken into account in assessing the response of communities in situations where noise levels change. Within the limitations of existing evidence on change, communities that experience an increase in noise exposure are likely to experience greater annoyance than is predicted from existing exposure-response relationships, and communities that experience a decrease in exposure experience greater benefit than predicted. Policy makers need to be informed of these potential change effects, particularly as situations in which noise levels increase as a result of infrastructure changes are always likely to be contentious. To do otherwise would be to deny them important information regarding potential community response in these contexts. There is already one practical example of this (The Highways Agency, 2008). The change effect is not transient and is likely to be present until the normal turnover of residents in any particular community results in newcomers replacing those who experienced the change. This policy implication applies irrespective of which of the potential mechanisms—change in nonacoustical variables, differential response criteria, or retained coping—might explain the change effect. However, some of the explanations have additional implications.

B. Implications for the interpretation of exposure-response relationships obtained in the steady state

If changing attitudes to the source/authorities prove to be the explanation of the change effect, there is the potential for considered interventions to be used as an instrument to reduce noise annoyance of affected populations in situations of change. Transparent information/communication about the noise changes could positively affect attitudes and expectations of the community. Evidence of the existence of a change effect demonstrates that this should not be perceived merely as manipulative public relations, but a bona fide and positive contribution to managing the magnitude of the annoyance responses of the community subject to the change.

A differential response criteria explanation has much wider implications. It raises the question that there may be a measurement error across the generalized exposure-response curves. These are based on responses of people who have been exposed in the steady-state to particular noise levels. The explanation suggests that measurement error may be operational in all steady-state situations but is revealed only in situations of change. The consequence, from the direction of the change effect, is that the gradient of an exposure-response curve adjusted for this purported error would be much steeper than that of currently used steady-state curves. There is a similar consequence arising from the explanation of changing attitudes. If attitudes of a group of respondents can collectively shift, in the manner required by this expla-

nation in change situations, one cannot discount that the groups of respondents included in the surveys from which steady-state exposure-response curves have been derived might also have collectively shifted their attitudes, particularly those groups chronically exposed to high levels of noise. This is confounding as it would mean that exposure-response curve baselines may already include in them, in part, the effect of these attitudes.

C. Future studies of change

This clarification of potential mechanisms provides a structure for the design of future studies of change. It provides guidance (Table I) as to what needs to be measured in longitudinal studies to overcome weaknesses in the existing set of studies/data in testing not only for the existence and durability of an excess-response change effect, but also for the various hypotheses to explain it.

The design of such studies will need to take into account that changes in nonacoustic factors may occur out of phase with the change in noise levels itself, potentially shaped by earlier announcements of infrastructure change (Job *et al.*, 1996), public information and consultative/nonconsultative processes, and construction activities (Schuemer and Schreckenberg, 2000). All survey waves in longitudinal studies need to include adequate measurements of exposure, changes in exposure, responses, and changes in the appropriate nonacoustical variables. Babisch and Gebhardt (1986), as did Fields *et al.* (2000), suggested that panel studies would be strengthened by the addition of nonpanel respondents in the repeat surveys.

V. CONCLUSIONS

A wide range of explanations for excess response has been proposed. This review suggests that, while there is still no accepted and evidence-based view on a single mechanism that can explain the change effect, several of the explanations can be discarded on the basis of insufficient evidence, inconsistencies, or the inability to fit the limited empirical data available on response to change. The residual plausible explanations are grouped into three categories, each relying on a different mechanism to explain the change effect: changes in modifiers of exposure-response relationships in the context of change in exposure, differential scaling criterion for the annoyance scale at different levels of exposure, and retention of coping strategies following a change. There are significant policy implications of the change effect in terms of assessing human response in situations of change, irrespective of mechanism, and several of the mechanisms raise important questions regarding the interpretation of the exposure-response relationships based on steady-state surveys.

[1]Not to be confused with self-reported "noise sensitivity," which is generally regarded as an unvarying personal characteristic.

Babisch, W., and Gebhardt, S. (**1986**). "Gestortheitsreaktionen durch Verkehrslamr—Eine 'vorher/nachher'-untersuchung (Annoyance reactions caused by traffic noise—A 'before/after'-study)," Z. Lärmbek. **33**, 38–45.

Baughan, C., and Huddart, L. (**1993**). "Effects of traffic noise changes on residents' nuisance ratings," Proceedings of the Sixth International Congress on Noise as a Public Health Problem, Noise and Man '93, Nice, July, Vol. **2**, pp. 585–588.

Bell, I. R., Hardin, E. E., Baldwin, C. M., and Schwartz, G. E. (**1995**). "Increased limbic system symptomatology and sensitizability of young adults with chemical and noise sensitivities," Environ. Res. **70**, 84–97.

Berglund, B., Berglund, U., and Lindvall, Y. (**1975**). "A study of response criteria in populations exposed to aircraft noise," J. Sound Vib. **41**, 33–39.

Breugelmans, O., Houthuijs, D., van Kamp, I., Stellato, R., van Wiechen, C., and Doornbos, G. (**2007**). "Longitudinal effects of a sudden change in aircraft noise exposure on annoyance and sleep disturbance around Amsterdam Airport," Proceedings of ICA, Madrid, Paper No. ENV-04-002-IP.

Brooker, P. (**2008**). "ANASE: Measuring aircraft noise annoyance very unreliably," Significance **5**, 18–24.

Brown, A. L. (**1987**). "Responses to an increase in road traffic noise," J. Sound Vib. **117**, 69–80.

Brown, A. L., Hall, A., and Kyle-Little, J. (**1985**). "Response to a reduction in traffic noise exposure," J. Sound Vib. **98**, 235–246.

Brown, A. L. (**2009**). "Response to a change in transport noise exposure: A review of evidence of a change effect," J. Acoust. Soc. Am. (in press).

Djokvucic, I., Hatfield, J., and Job, R. F. S. (**2004**). "Experimental examination of the effects of attitude to the noise source on reaction and on reaction to performance," Proceedings of Internoise 2004, Prague, Paper No. 325.

Evans, G. W., Bullinger, M., and Hygge, S. (**1998**). "Chronic noise exposure and physiological response: A prospective study of children living under environmental stress," Psychol. Sci. **9**, 75–77.

Fidell, S., Horonjeff, R., Mills, J., Baldwin, E., Teffeteller, S., and Pearsons, K. (**1985**). "Aircraft annoyance at three joint air carrier and general aviation airports," J. Acoust. Soc. Am. **77**, 1054–1068.

Fidell, S., and Jones, G. (**1975**). "Effects of cessation of late-night flights on an airport community," J. Sound Vib. **42**, 411–427.

Fidell, S., Pearsons, K., Tabachnik, B. G., and Howe, R. (**2000**). "Effects on sleep disturbance of changes in aircraft noise near three airports," J. Acoust. Soc. Am. **107**, 2535–2547.

Fidell, S., and Silvati, L. (**2004**). "Parsimonious alternatives to regression analysis for characterizing prevalence rates of aircraft noise annoyance," Noise Control Eng. J. **52**, 56–68.

Fields, J. M. (**1993**). "Effect of personal and situational variables on noise annoyance in residential areas," J. Acoust. Soc. Am. **93**, 2753–2763.

Fields, J. M. (**1994**). "A review of an updated synthesis of noise/annoyance relationships," NASA Report No. CR-194950, NASA Langley Research Center, Hampton, VA.

Fields, J. M., Ehrlich, G. E., and Zador, P. (**2000**). "Theory and design tools for studies of reactions to abrupt changes in noise exposure," NASA Report No. CR-2000-210280, NASA Langley Research Center, Hampton, VA.

Fields, J. M., and Hall, F. L. (**1987**). "Community effects of noise," in *Transportation Noise Reference Book*, edited by P. Nelson (Butterworth, Washington, DC).

Flindell, I. H., and Witter, I. J. (**1999**). "Non-acoustical factors in noise management at Heathrow airport," Noise Health **1**, 27–44.

Griffiths, I. D., and Langdon, F. J. (**1968**). "Subjective response to road traffic noise," J. Sound Vib. **8**, 16–32.

Griffiths, I. D., and Raw, G. J. (**1989**). "Adaptation to changes in traffic noise exposure," J. Sound Vib. **132**, 331–336.

Guski, R. (**1999**). "Personal and social variables as co-determinants of noise annoyance," Noise Health **3**, 45–56.

Guski, R. (**2004**). "How to forecast community annoyance in planning noisy facilities," Noise Health **6**, 59–64.

Hatfield, J., Job, R. F. S., Carter, N. L., Peploe, P., Taylor, R., and Morell, S. (**2001**). "The role of adaptation in responses to noise exposure: Comparison of steady state with newly high noise areas," Proceedings of the Fourth European Conference on Noise Control, Euronoise PATRA, January.

Helson, H. (**1964**). *Adaptation-Level Theory* (Harper & Row, New York).

Horonjeff, R. D., and Robert, W. E. (**1997**). "Attitudinal response to changes in noise exposure in residential communities," NASA Report No. CR-97-205813, National Aeronautics and Space Administration, Washington DC, p. 150.

Houthuijs, D., Breugemans, O., van Kamp, I., and van Wiechen, C. (**2007**). "Burden of annoyance dues to aircraft noise and non-acoustical factors," Proceedings of Internoise 2007, Instanbul, Paper No. 838472.

Job, R. F. S. (**1988b**). "Over-reaction to changes in noise exposure: The

possible effect of attitude," J. Sound Vib. **126**, 550–552.
Job, R. F. S. (**1988a**). "Community response to noise: A review of factors influencing the relationship between noise exposure and reaction," J. Acoust. Soc. Am. **83**, 991–1001.
Job, R. F. S., Topple, A., Carter, N. L., Peploe, P., Taylor, R., and Morell, S. (**1996**). "Public reactions to changes in noise levels around Sydney Airport," Proceedings of Internoise 1996, Liverpool, UK, Vol. **5**, pp. 2419–2424.
Jonsson, E., and Sörensen, S. (**1973**). "Adaptation to community noise—A case study," J. Sound Vib. **26**, 571–575.
Kastka, J. (**1981**). "Zum Einfluss verkehrsberuhigender Maßnahmen auf Lärmbelastung und Lärmbelästigung (About the impact of traffic calming measures on noise levels and annoyance)," Z. Lärmbek. **28**, 25–30.
Klæboe, R., Kolbenstvedt, M., Clench-Aas, J., and Bartonova, A. (**2000**). "Oslo traffic study—Part 1: An integrated approach to assess the combined effects of noise and air pollution on annoyance," Atmos. Environ. **34**, 4727–4736.
Klæboe, R., Kolbenstvedt, M., Lercher, P., and Solberg, S. (**1998**). "Changes in noise reactions—Evidence for an area effect?," Proceedings of Internoise 1998, Christchurch, New Zealand, pp. 16–18 (CD-ROM).
Lambert, R. F. (**1978**). "Experimental evaluation of a freeway noise barrier," Noise Control Eng. **11**, 86–94.
Lambert, J., Champelovier, P., and Vernet, I. (**1996**). "Annoyance from high speed train noise: A social survey," J. Sound Vib. **193**, 21–28.
Langdon, J. (**1976**). "Noise nuisance caused by road traffic noise in residential areas: Part I," J. Sound Vib. **111**, 243–263.
Langdon, F. J., and Griffiths, I. D. (**1982**). "Subjective effects of traffic noise exposure: II. Comparisons of noise indices," J. Sound Vib. **83**, 171–180.
Lawson, B. R., and Walters, D. (**1973**). "The effects of a new motorway on an established /residential area," Proceedings of the International Conference on Environmental Psychology, University of Surrey, Guildford.
Mackie, M., and Davies, C. H. (**1981**). "Environmental effects of traffic change," TRRL Laboratory Report No. 1015, Transport and Road Rersearch Laboratory, Crowthorne, UK.
Maris, E., Stallen, P. J., Vermunt, R., and Steensma, H. (**2007**). "Noise within the social context: Annoyance reduction through fair procedures," J. Acoust. Soc. Am. **121**, 2000–2010.
Mehra, S. R., and Lutz, C. (**2000**). "Berechnung und subjective Wahrnehmung der Lärmpegeländerung aufgrund einer neu erstellten Umgehungsstraße (Measurement and subjective perception of noise level changes due to a new roadway)," Z. Lärmbek. **47**, 58–67.
Miedema, H. M. E., and Vos, H. (**1999**). "Demographic and attitudinal factors that modify annoyance from transportation noise," J. Acoust. Soc. Am. **105**, 3336–3344.
Miedema, H. M. E., and Vos, H. (**2003**). "Noise sensitivity and reactions to noise and other environmental conditions," J. Acoust. Soc. Am. **113**, 1492–1504.
Ministry of Transport, Public Works and Water Management (**2005**). "Evaluatie Schipholbeleid: Schiphol beleefd door omwonenden (Evaluation Schiphol: Schiphol perceived by its residents)," Directorate-General Transport and Aviation.
Nijland, H. A., Hartemink, A., van Kamp, I., and van Wee, B. (**2007**). "The influence of sensitivity for road traffic noise on residential location: Does it trigger a process of spatial selection?," J. Acoust. Soc. Am. **122**, 1595–1601.
Nilsson, M. E., and Berglund, B. (**2006**). "Noise annoyance and activity disturbance before and after the erection of a roadside noise barrier," J. Acoust. Soc. Am. **119**, 2178–2188.
Nimura, T., Sone, T., and Kono, S. (**1973**). "Some considerations on noise problem of high-speed railway in Japan," Proceedings of Internoise 73, Copenhagen, pp. 298–307.
Öhrström, E. (**1997**). "Community reactions to railway traffic—Effects of countermeasures against noise and vibration," Proceedings Internoise 97, Budapest, pp. 1065–1070.
Öhrström, E. (**2004**). "Longitudinal surveys on effects of changes in road traffic noise-annoyance, activity disturbances, and psycho-social well-being,"J. Acoust. Soc. Am. **115**, 719–729.
Quehl, J., and Basner, M. (**2006**). "Annoyance from nocturnal aircraft noise exposure: Laboratory and field-specific dose-response curves," J. Environ. Psychol. **26**, 127–140.
Raw, G. J., and Griffiths, I. D. (**1985**). "The effect of changes in aircraft noise exposure (Letter to the editor)," J. Sound Vib. **101**, 273–275.
Raw, G. J., and Griffiths, I. D. (**1990**). "Subjective response to changes in road traffic noise: A model," J. Sound Vib. **141**, 43–54.
Robinson, D. W., Bowsher, J. M., and Copeland, W. C. (**1963**). "On judging the noise from aircraft in flight," Acustica **13**, 324–336.
Scholes, W. E. (**1977**). "The physical and subjective evaluation of roadside barriers," Proceedings of Internoise, pp. 144–153.
Schreckenberg, D., and Meis, M. (**2007**). "Noise annoyance around an international airport planned to be extended," Proceedings of Internoise 2007, Istanbul, Turkey.
Schreckenberg, D., Schuemer, R., and Moehler, U. (**2001**). "Railway-noise annoyance and 'misfeasance' under conditions of change," Proceedings of Internoise 2001, The Hague, Netherlands, Paper No. 344 (CD-ROM).
Schuemer, R., and Schreckenberg, D. (**2000**). "Anderung der Larmbelastung bei massnahme bedingter stufenweise veranderter gerauschbelastung—Hinweise auf einige Befunde und Interpretationsansatze (The effect of stepwise change of noise exposure on annoyance)," Z. Lärmbek. **47**, 134–143.
Schultz, T. J. (**1978**). "Synthesis of social surveys on noise annoyance," J. Acoust. Soc. Am. **64**, 377–405.
The Highways Agency (**2008**). Design Manual for Roads and Bridges, Volume **11**, Section 3, Part 7, Traffic Noise and Vibration, http://www.standardsforhighways.co.uk/dmrb/ vol11/section3/11s3p07.pdf (Last viewed April 7, 2008).
van Dongen, J. E. F., and van den Berg, R. (**1983**). "De gewenning aan het geluid van een nieuwe spoorlijn (Getting used to noise from a new railway line)," Report No. RL-HR-03-02, IMG-TNO, Delft.
van Kamp, I., Job, R. F. S., Hatfield, J., Haines, M., Stellato, R. K., and Stansfeld, S. A. (**2004**). "The role of noise sensitivity in the noise-response relation: A comparison of three International Airport Studies," J. Acoust. Soc. Am. **116**, 3471–3479.
van Kempen, E. E. M. M., and van Kamp, I. (**2005**). "Annoyance from air traffic noise. Possible trends in exposure-response relationships," Report No. 01/2005, MGO Evk, RIVM, Bilthoven.
Vincent, B., and Champelovier, P. (**1993**). "Changes in the acoustic environment: Need for an extensive evaluation of annoyance," Proceedings Noise and Man '93, Sixth International Congress on Noise as a Public Health Problem, Vol. **2**, pp. 425–428.
Weinstein, N. D. (**1978**). "Individual differences in reaction to noise: A longitudinal study in a college dormitory," J. Appl. Psychol. **2**, 87–97.
Weinstein, N. D. (**1982**). "Community noise problems: Evidence against adaptation," J. Environ. Psychol. **2**, 87–97.

A description of transversely isotropic sound absorbing porous materials by transfer matrices

P. Khurana, L. Boeckx, and W. Lauriks
Laboratorium voor Akoestiek en Thermische Fysica, Katholieke Universiteit Leuven, Celestijnenlaan 200D, B-3001 Heverlee, Belgium

P. Leclaire
Laboratoire de Recherche en Mécanique et Acoustique, Université de Bourgogne, 49 rue Mademoiselle Bourgeois, B.P. 31, 58027 Nevers Cedex, France

O. Dazel and J. F. Allard
Laboratoire d'Acoustique de l'Université du Maine, UMR CNRS 6613, Avenue Olivier Messiaen, F-72085 Le Mans Cedex, France

(Received 2 March 2008; revised 31 October 2008; accepted 3 November 2008)

A description of wave propagation in transversely isotropic porous materials saturated by air with a recent reformulation of the Biot theory is carried out. The description is performed in terms of a transfer matrix method (TMM). The anisotropy is taken into account in the mechanical parameters (elastic constants) and in the acoustical parameters (flow resistivity, tortuosity, and characteristic lengths). As an illustration, the normal surface impedance at normal and oblique incidences of transversely isotropic porous layers is predicted. Comparisons are performed with experimental results. © *2009 Acoustical Society of America.* [DOI: 10.1121/1.3035840]

PACS number(s): 43.55.Ev, 43.20.Gp, 43.20.Jr [SFW] Pages: 915–921

I. INTRODUCTION

A full description of wave propagation through an anisotropic poroelastic material is seldom used for the study of the acoustical behavior of soft highly porous absorbers, which are used for acoustic noise reduction. These materials, such as foams and fibrous materials, may be orthotropic or transversely isotropic, due to their manufacturing process. A more precise optimization of the acoustic performance of these materials, either isolated or as a part of a layered structure, can be achieved by modeling the effects of anisotropy. The theory of wave propagation in anisotropic poroelastic solid was achieved by Biot.[1,2] Since then, many authors, mostly in the field of geophysics, studied the aspects of the wave propagation through anisotropic poroelastic solids and made adaptations on the original formulation of Biot. Carcione[3] analyzed the anisotropic poroelastic media and numerically solved Biot's anisotropic equations. Vashishth and Khurana[4] studied the wave propagation in stratified anisotropic materials taking into account the anisotropy in the elastic constants but neglected the anisotropy in the dynamic permeability. Liu and Liu[5] studied the wave fronts and velocity surfaces of Rayleigh waves in water-saturated orthotropic porous media and indicated the differences with the isotropic and transversely isotropic cases. The wave propagation through highly porous soft absorbers was often studied by using the rigid frame approximation[6–8] and restricted to isotropic media. The use and development of more complete models are mainly limited by the lack of measurement data on the anisotropy of porous sound absorbers. Allard *et al.*[6] measured the effect of the anisotropy in glass wool on the normal surface impedance. They observed deviations in the real part of the surface impedance, which they were able to model using the laws of Delany and Bazley.[9] Several authors investigated the directional differences in material parameters as defined in the Johnson–Allard *equivalent fluid* model. The tortuosity, the viscous characteristic length, and the flow resistivity of open cell foams along the principal axes of the material were measured by Melon *et al.*[10,11] It was shown that there were differences depending on the measurement direction in the mechanical parameters and in the parameters of the rigid frame model. However, there was no description of the acoustical behavior according to the measured material properties. Tarnow[12] proposed an experimental setup allowing the measurement of the elastic constants in a frequency range of 20–160 Hz along the principal axes of symmetry. He gave a complete set of elastic constants for a glass wool considered as a transversely isotropic medium. This allows a modeling of wave propagation in transversely isotropic porous media in terms of a transfer matrix method (TMM) as described in the present work. The anisotropy in mechanical parameters and in the flow resistivity σ, the tortuosity α_∞, and the viscous characteristic length Λ as defined in the Johnson–Allard–Champoux[13–15] model is taken into account. Moreover, a recent formulation[16] of the Biot theory which allows important simplifications in the calculations is used. As an illustration, the TMM is used to predict the surface impedance at normal and oblique incidences of a transversely isotropic soft porous medium of high porosity. Measurement data on the anisotropy factors for the studied material are provided and the influence of these parameters on surface impedance of the material is discussed. Predictions and measurements of the surface impedance are compared.

II. EXPRESSION OF THE SURFACE IMPEDANCE OF A TRANSVERSELY ISOTROPIC POROUS MATERIAL

The surface impedance of a transversely isotropic porous material is predicted from a TMM using the $\{\mathbf{u}^s,\mathbf{u}^t\}$ representation of the Biot model. This representation is based on the use of the displacement vector $\mathbf{u}^s$ of the solid phase and the total displacement $\mathbf{u}^t$ which is defined in what follows. This representation provides a description simpler than the one obtained with both classical representations of the Biot theory. This section is organized as follows: the motion equations are derived, the properties (slowness and polarization) of plane waves are then studied, and the expression of the transfer matrix of a transversely isotropic porous material is obtained. This transfer matrix is then used to obtain the surface impedance of a layer with the axis of symmetry z perpendicular to the faces.

A. $\{\mathbf{u}^s,\mathbf{u}^t\}$ motion equations for a transversely isotropic porous material

Biot[2] extended the isotropic theory for poroelastic materials to the anisotropic case. He took into account the anisotropy in mechanical parameters only but his equations can easily be modified to consider the anisotropy of flow resistivity, tortuosity, and the viscous characteristic length. Let $\mathbf{u}^s$ be the frame displacement and $\mathbf{u}^f$ be the air displacement. Under harmonic excitation at circular frequency ω, the Biot motion equations become

$$\begin{aligned}\nabla\cdot\sigma^s(\mathbf{u}^s,\mathbf{u}^f) &= -\omega^2[\tilde{\rho}_{11}]\mathbf{u}^s-\omega^2[\tilde{\rho}_{12}]\mathbf{u}^f,\\ \nabla\cdot\sigma^f(\mathbf{u}^s,\mathbf{u}^f) &= -\omega^2[\tilde{\rho}_{12}]\mathbf{u}^s-\omega^2[\tilde{\rho}_{22}]\mathbf{u}^f,\end{aligned} \tag{1}$$

where $\mathbf{u}^s$ is the solid phase displacement, $\mathbf{u}^f$ is the fluid phase displacement, $\sigma^s(\mathbf{u}^s,\mathbf{u}^f)$ [respectively, $\sigma^f(\mathbf{u}^s,\mathbf{u}^f)$] is the stress tensor of the solid (respectively, of the fluid) phase, and the $[\tilde{\rho}_{ij}]$ $\{i,j\}\in\{1,2\}$ are diagonal matrices defined by

$$[\tilde{\rho}_{ij}]=\mathrm{diag}(\tilde{\rho}^x_{ij},\tilde{\rho}^x_{ij},\tilde{\rho}^z_{ij}). \tag{2}$$

In this equation, the Biot densities $\tilde{\rho}^i_{ij}$, with i replaced by x or z, are given by

$$\tilde{\rho}^i_{12}=\phi\rho_0(1-\tilde{\alpha}^i),\quad \tilde{\rho}^i_{22}=\phi\rho_0-\tilde{\rho}^i_{12},\quad \tilde{\rho}^i_{11}=(1-\phi)\rho_s-\tilde{\rho}^i_{12}, \tag{3}$$

where ρ_0 is the density of air, ρ_s is the density of the frame, ϕ is the porosity, and $\tilde{\alpha}^i$ is the dynamic tortuosity in the direction x or in the direction z. In the work by Johnson *et al.*,[13] the dynamic tortuosity is given by

$$\tilde{\alpha}^i_\infty=1-\frac{i\phi\sigma^i}{\alpha^i_\infty\rho_0\omega}\sqrt{1-\frac{4i\alpha^{i2}_\infty\gamma_a\rho_0\omega}{(\sigma^i\Lambda^i\sigma)^2}}, \tag{4}$$

where η_a is the dynamic viscosity of air. With the total displacement formulation,[16] the motion is described in terms of the frame displacement $\mathbf{u}^s$ and the total displacement $\mathbf{u}^t$ given by

$$\mathbf{u}^t=(1-\phi)\mathbf{u}^s+\phi\mathbf{u}^f. \tag{5}$$

The total displacement formulation simplifies the formalism of the Biot theory. This formulation can be extended to the case of transversely isotropic porous media. The strain-stress relations can be written, under the hypothesis that the medium, the frame is made of, is not compressible:

$$-p=\tilde{K}_{\mathrm{eq}}\nabla\cdot\mathbf{u}^t,\quad \sigma^s_{ij}=\hat{\sigma}^s_{ij}-(1-\phi)p\,\delta_{ij}, \tag{6}$$

where p is the interstitial pressure, $\hat{\sigma}^s_{ij}$ is the *in vacuo* stress tensor of the frame which only depends on $\mathbf{u}^s$, and $\tilde{K}_{\mathrm{eq}}$ is the bulk modulus of the fluid modified by the thermal exchanges with the frame,[16] which is given with the Champoux–Allard model by

$$\tilde{K}_{\mathrm{eq}}=(\gamma P_0/\phi)\left\{\gamma-(\gamma-1)\times\left[1+\frac{8\eta_a}{i\Lambda'\mathrm{Pr}\omega\rho_0}\sqrt{1+\frac{i\rho_0\omega\mathrm{Pr}\Lambda'^2}{16\eta_a}}\right]\right\}^{-1}. \tag{7}$$

In this equation, P_0 is the atmospheric static pressure, γ is the ratio of the specific heats, Λ' is the thermal characteristic length, and Pr is the Prandtl number. The *in vacuo* stress-strain relations can be written as

$$\begin{aligned}\hat{\sigma}_{xx}&=(2N+\hat{A})\varepsilon_{xx}+\hat{A}\varepsilon_{yy}+\hat{F}\varepsilon_{zz},\\ \hat{\sigma}_{yy}&=\hat{A}\varepsilon_{xx}+(2N+\hat{A})\varepsilon_{yy}+\hat{F}\varepsilon_{zz},\\ \hat{\sigma}_{zz}&=\hat{F}\varepsilon_{xx}+\hat{F}\varepsilon_{yy}+\hat{C}\varepsilon_{zz},\\ \hat{\sigma}_{yz}&=2L\varepsilon_{yz},\quad \hat{\sigma}_{xz}=2L\varepsilon_{xz},\quad \hat{\sigma}_{xy}=2N\varepsilon_{xy}.\end{aligned} \tag{8}$$

The equations of motion for the $\{\mathbf{u}^s,\mathbf{u}^t\}$ formulation are

$$\begin{aligned}\nabla\cdot\hat{\sigma}^s&=-\omega^2[\tilde{\rho}_s]\mathbf{u}^s-\omega^2[\tilde{\gamma}][\tilde{\rho}_{\mathrm{eq}}]\mathbf{u}^t,\\ \tilde{K}_{\mathrm{eq}}\nabla\cdot(\nabla\cdot\mathbf{u}^t[I])&=-\omega^2[\tilde{\gamma}][\tilde{\rho}_{\mathrm{eq}}]\mathbf{u}^s-\omega^2[\tilde{\rho}_{\mathrm{eq}}]\mathbf{u}^t,\end{aligned} \tag{9}$$

where $[\tilde{\gamma}]$, $[\tilde{\rho}_{\mathrm{eq}}]$, and $[\tilde{\rho}_s]$ are diagonal matrices given by

$$\begin{aligned}[\tilde{\gamma}]&=\phi\left([\tilde{\rho}_{22}]^{-1}[\tilde{\rho}_{12}]-\frac{1-\phi}{\phi}[I]\right),\\ [\tilde{\rho}_{\mathrm{eq}}]&=[\tilde{\rho}_{22}]/\phi^2,\\ [\tilde{\rho}_s]&=[\tilde{\rho}]+[\tilde{\gamma}]^2[\tilde{\rho}_{\mathrm{eq}}].\end{aligned}$$

In these equations, the matrix $[I]$ is the identity matrix of size 3 and the matrix $[\tilde{\rho}]=[\tilde{\rho}_{11}]-[\tilde{\rho}_{12}]^2[\tilde{\rho}_{22}]^{-1}$. The equations of motion obtained with the new $\mathbf{u}^s,\mathbf{u}^t$ formulation[16] are simpler than the ones in the previous representations.

B. Plane waves propagating in TIPM

This section is related to poroelastic plane waves.[1] Regardless of the formulation $\{u^s,u^f\}$, $\{u^s,u^t\}$, or $\{u^s,w\}$, for poroelastic medium the solid and fluid displacements follow the same dispersion curve. The methodology of Vashishth and Khurana[4] can then be extended to our formulation.

The acoustic field is created in a transversely isotropic medium by an incident air wave. Without loss of generality, the incidence plane is the xz plane. The angle of incidence is θ. The time dependence is $\exp(i\omega\tau)$. The space dependence for a plane wave can be written as

$$\mathbf{u}^s = \mathbf{a}\exp(-\mathbf{qx}), \quad \mathbf{u}^t = \mathbf{b}\exp(-\mathbf{qx}), \tag{10}$$

where $\mathbf{a}=\{a_x\ a_y\ a_z\}$ and $\mathbf{b}=\{b_x\ b_y\ b_z\}$ are the polarization vectors and $\mathbf{q}=\{q_x=\sin\theta/c_0, q_y=0, q_z\}$ is the slowness vector. The x and y slowness components are those of the incident field. Substituting the expressions for displacement given by Eq. (10) in Eq. (9) provide a homogeneous linear system of six equations which can be split in two sets. One set corresponds to the two y direction equations for the solid displacement and the total displacement and concerns the quasishear horizontal (qSH) waves. The following relations are obtained for these waves:

$$b_y = -\tilde{\gamma}_y a_y, \quad q_z^2 = \frac{1}{L}(\hat{\rho}^x - Nq_x^2). \tag{11}$$

Hence two qSH waves are obtained by taking the square root of q_z^2, a downgoing ($R_{\mathrm{eq}_z}>0$) wave and an upgoing ($R_{\mathrm{eq}_z}>0$) wave. These two waves are not investigated as they are not excited by the incident field. The four remaining equations from Eq. (10) relate the polarizations in the x and the z directions and can be written in the following form:

$$[\mathbf{A}]\{a_x \quad a_z \quad b_x \quad b_z\}^T = \{\mathbf{0}\}, \tag{12}$$

with

$$[\mathbf{A}] = \begin{bmatrix} Lq_x^2 - \tilde{\rho}_s^x + q_x^2\hat{P} & (\hat{F}+L)q_xq_z & -\gamma_x\tilde{\rho}_{\mathrm{eq}}^x & 0 \\ (L+\hat{F})q_xq_z & (\hat{C}q_z^2 + Lq_x^2) - \tilde{\rho}_s^z & 0 & -\gamma_z\tilde{\rho}_{\mathrm{eq}}^z \\ -\gamma_x\tilde{\rho}_{\mathrm{eq}}^x & 0 & -\tilde{\rho}_{\mathrm{eq}}^x + \tilde{K}_{\mathrm{eq}}q_x^2 & \tilde{K}_{\mathrm{eq}}q_xq_z \\ 0 & -\gamma_z\rho_{\mathrm{eq}}^z & \tilde{K}_{\mathrm{eq}}q_x & \tilde{K}_{\mathrm{eq}}q_z^2 - \tilde{\rho}_{\mathrm{eq}}^z \end{bmatrix}. \tag{13}$$

The researched values of q_z correspond to $|\mathbf{A}|=0$ whose expression leads to

$$T_3q_z^6 + T_2q_z^4 + T_1q_z^2 + T_0 = 0, \tag{14}$$

with

$$T_3 = -L\hat{C}\tilde{K}_{\mathrm{eq}}\tilde{\rho}_{\mathrm{eq}}^x, \tag{15}$$

$$T_2 = T_{2,2}q_x^2 + T_{2,0}, \tag{16}$$

$$T_{2,2} = -\tilde{K}_{\mathrm{eq}}[L\hat{C}\tilde{\rho}_{\mathrm{eq}}^z + \tilde{\rho}_{\mathrm{eq}}^x(\hat{P}\hat{C} + L^2 - (\hat{F}+L)^2)], \tag{17}$$

$$T_{2,0} = \tilde{\rho}_{\mathrm{eq}}^x[\tilde{\rho}^x\hat{C}\tilde{K}_{\mathrm{eq}} + L(\hat{C}\tilde{\rho}_{\mathrm{eq}}^z + \tilde{\rho}_{s,z}\tilde{K}_{\mathrm{eq}})], \tag{18}$$

$$T_1 = T_{1,4}q_x^4 + T_{1,2}q_x^2 + T_{1,0}, \tag{19}$$

$$T_{1,4} = -\tilde{K}_{\mathrm{eq}}[L\hat{P}\tilde{\rho}_{\mathrm{eq}}^x + \tilde{\rho}_{\mathrm{eq}}^z(\hat{P}\hat{C} + L^2 - (\hat{F}+L)^2\ ], \tag{20}$$

$$T_{1,2} = \tilde{K}_{\mathrm{eq}}[L(\tilde{\rho}_{\mathrm{eq}}^z\tilde{\rho}^z + \tilde{\rho}_{\mathrm{eq}}^x\tilde{\rho}^x) + \hat{P}\tilde{\rho}_{\mathrm{eq}}^x\tilde{\rho}_s^z + \hat{C}\tilde{\rho}_{\mathrm{eq}}^z\tilde{\rho}_s^x] + \tilde{\rho}_{\mathrm{eq}}^z\tilde{\rho}_{\mathrm{eq}}^x[L^2 + \hat{P}\hat{C} - (F+L)^2 - 2(F+L)\tilde{K}_{\mathrm{eq}}\tilde{\gamma}_x\tilde{\gamma}_z\ , \tag{21}$$

$$T_{1,0} = -\tilde{\rho}_{\mathrm{eq}}^x[L\tilde{\rho}_{\mathrm{eq}}^x\tilde{\rho}^z + \hat{C}\tilde{\rho}_{\mathrm{eq}}^z\tilde{\rho}_x + \tilde{K}_{\mathrm{eq}}\tilde{\rho}^x\tilde{\rho}_s^z], \tag{22}$$

$$T_0 = T_{0,6}q_x^6 + T_{0,4}q_x^4 + T_{0,2}q_x^2 + T_{0,0}, \tag{23}$$

$$T_{0,6} = -\tilde{\rho}_{\mathrm{eq}}^z L\hat{P}\tilde{K}_{\mathrm{eq}}, \tag{24}$$

$$T_{0,4} = \tilde{\rho}_{\mathrm{eq}}^z[L(\tilde{K}_{\mathrm{eq}}\tilde{\rho}_s^x + \hat{P}\tilde{\rho}_{\mathrm{eq}}^x) + \tilde{\rho}^x\hat{P}\tilde{K}_{\mathrm{eq}}], \tag{25}$$

$$T_{0,2} = -\tilde{\rho}_{\mathrm{eq}}^z[L\tilde{\rho}^x\tilde{\rho}_{\mathrm{eq}}^x + \tilde{\rho}_z(\tilde{K}_{\mathrm{eq}}\tilde{\rho}_s^x + \hat{P}\tilde{\rho}_{\mathrm{eq}}^x)], \tag{26}$$

$$T_{0,0} = \tilde{\rho}_{\mathrm{eq}}^z\tilde{\rho}^z\tilde{\rho}^x\tilde{\rho}_{\mathrm{eq}}^x. \tag{27}$$

There are no odd terms in this cubic polynomial in q_z^2. Each root provides two square roots. These can be numbered with the index $k=1,2,3$ for the downgoing waves and $k+3$ for the upgoing waves. For each q_z, the polarization of the wave can be normalized so that $b_z=1$

$$\{a_x, a_z, b_x, b_z\} \propto \{\mu_{x,s}, \mu_{z,s}, \mu_{x,t}, 1\}. \tag{28}$$

This normalization does not allow a null z total displacement component, but there is no restriction to use it in our context. The coefficients μ can be written as

$$\mu_{x,t} = \frac{-(\gamma_z q_{c,z}^2 q_x q_z)[(P_0 q_x^2) + L_0 q_z^2 - q_{s,x}^2] - (\gamma_x q_{c,x}^2 q_x q_z)(q_z^2 - q_{\mathrm{eq},z}^2)}{\gamma_z q_{c,z}^2[(P_0 q_x^2 + L_0 q_z^2 - q_{s,x}^2)(q_x^2 - q_{\mathrm{eq},x}^2) - \gamma_x q_{c,x}^2 q_{e,x}^2] + \gamma_z q_{c,z}^2 q_z^2 q_x^2}, \tag{29}$$

where $P_0=\hat{P}/(L+\hat{F})$ and $L_0=L/(L+\hat{F})$,

$$\mu_{z,s}=\frac{\tilde{K}_{\mathrm{eq}}[(q_z^2-q_{\mathrm{eq},z}^2)+q_zq_x\mu_{x,t}]}{(L+\hat{F})\tilde{\gamma}_zq_{c,z}^2}, \tag{30}$$

$$\mu_{x,s}=\frac{\tilde{K}_{\mathrm{eq}}(q_zq_x+(q_x^2-q_{\mathrm{eq},x}^2)\mu_{x,t})}{(L+\hat{F})\tilde{\gamma}_zq_{c,z}^2}, \tag{31}$$

with the slownesses of the material

$$q_{\mathrm{eq},i}=\sqrt{\frac{\tilde{\rho}_{\mathrm{eq}}^i}{\tilde{K}_{\mathrm{eq}}}},\quad q_{c,i}=\sqrt{\frac{\tilde{\rho}_{\mathrm{eq}}^i}{L+\hat{F}}},\quad q_{s,i}=\sqrt{\frac{\tilde{\rho}_s^i}{L+\hat{F}}}. \tag{32}$$

The following parity/imparity relations are used for the upgoing waves:

$$\begin{aligned}&\mu_{x,t}(k+3)=-\mu_{x,t}(k),\quad \mu_{z,s}(k+3)=\mu_{z,s}(k),\\&\mu_{z,s}(k+3)=-\mu_{x,s}(k).\end{aligned} \tag{33}$$

C. Expression of the transfer matrix of a transversely isotropic layer in the $\{\mathbf{u}^s,\mathbf{u}^t\}$ formulation

A transfer matrix connecting state vectors describing the mechanical field at each plane boundary of an anisotropic elastic medium was used by Brekhovskikh.[17] In this case, the mechanical field can be described by two waves which propagate toward increasing z and two waves which propagate toward decreasing z. The field in the medium is completely described if four amplitudes of these waves or for independent quantities describing the field are known. In this case, a state vector with four components is used. The first use of a transfer matrix for isotropic porous media was performed by Depollier.[18] Three kinds of waves, in context of the Biot theory, can propagate in an isotropic porous medium and the state vector has six components. The number of components is also 6 for a trans (TIPM)] if only the waves polarized in the meridian plane are present.

The transfer matrix provides a relation between state vectors of the medium at two different z. The state vector at a generic position $z=l$ is defined by

$$V(l)=[\dot{u}_z^s(l)\quad \dot{u}_x^s(l)\quad \dot{u}_z^t(l)\quad p(l)\quad \hat{\sigma}_{xz}(l)\quad \hat{\sigma}_{zz}(l)]^t. \tag{34}$$

This state vector is conserved at the interface between two porous media and it allows a simple coupling with air.[16] The transfer matrix is obtained with the method described in Refs. 4, 15, 19, and 20. Each component of the state vector can be expressed as a sum of the contributions of the six waves. For example, the z solid phase velocity components at $z=0$ and at $z=H$ are linked to the amplitudes f_k of the waves by

$$\dot{u}_z^s(0)=\sum_{k=1}^{6}r_1(k)f_k,\quad \dot{u}_z^s(H)=\sum_{k=1}^{6}r_1(k)e_kf_k, \tag{35}$$

where

$$\begin{aligned}&r_1(k)=i\omega\mu_{z,s}(k),\quad e_k=\exp(i\omega q_z(k)H),\\&k=1,2,3,\quad e_{k+3}=\frac{1}{e_k},\end{aligned} \tag{36}$$

and f_k denotes the amplitude of the waves. Similar equations can be obtained from the five remaining components of the state vector which involve the functions r_i, $i=2,3,4,5,6$:

$$r_2(k)=i\omega\mu_{x,s}(k),\quad r_3(k)=i\omega, \tag{37}$$

$$r_4(k)=\lambda_p(k),\quad r_5(k)=\lambda_x(k),\quad r_6(k)=\lambda_z(k), \tag{38}$$

where the λ are given by

$$\lambda_x(k)=-iL\omega[q_z(k)\mu_{x,s}+q_x\mu_{z,s}],\quad \lambda_x(k+3)=\lambda_x(k), \tag{39}$$

$$\lambda_z(k)=-i\omega[\hat{F}q_x\mu_{x,s}+\hat{C}q_z(k)\mu_{z,s}],\quad \lambda_z(k+3)=-\lambda_z(k), \tag{40}$$

$$\lambda_p(k)=i\omega\hat{K}_{\mathrm{eq}}[q_x\mu_{x,t}+q_z(k)],\quad \lambda_p(k+3)=-\lambda_p(k). \tag{41}$$

The transfer matrix $[\mathbf{T}]$ can be defined by

$$\mathbf{V}(H)=[\mathbf{T}]\mathbf{V}(0). \tag{42}$$

The matrix elements are given by

$$T_{ij}=\sum_{k=1}^{3}\left(e_k+\frac{(-1)^{i+j}}{e_k}\right)r_i(k)c_j(k), \tag{43}$$

where

$$c_1(k)=\frac{\lambda_x(k^+)-\lambda_x(k^{++})}{i\omega\Delta_2}, \tag{44}$$

$$c_2(k)=\frac{\lambda_p(k^+)\lambda_z(k^{++})-\lambda_p(k^{++})\lambda_z(k^+)}{i\omega\Delta/\Delta_1}, \tag{45}$$

$$c_3(k)=\frac{\mu_{z,s}(k^+)\lambda_x(k^{++})-\mu_{z,s}(k^{++})\lambda_x(k^+)}{i\omega\Delta_2}, \tag{46}$$

$$c_4(k)=\frac{\mu_{x,s}(k^{++})\lambda_z(k^+)-\mu_{x,s}(k^+)\lambda_z(k^{++})}{\Delta/\Delta_1}, \tag{47}$$

$$c_5(k)=\frac{\mu_{z,s}(k^{++})-\mu_{z,s}(k^+)}{\Delta_2}, \tag{48}$$

$$c_6(k)=\frac{\lambda_p(k^{++})\mu_{x,s}(k^+)-\lambda_p(k^+)\mu_{x,s}(k^{++})}{\Delta/\Delta_1}, \tag{49}$$

$$\Delta_1=4\frac{\Sigma_{k=1}^3\mu_{z,s}(k)(\lambda_x(k^+)-\lambda_x(k^{++}))}{e_1e_2e_3}, \tag{50}$$

$$\Delta_2=2e_1e_2e_3\sum_{k=1}^{3}\mu_{x,s}(k)(\lambda_p(k^+)-\lambda_p(k^{++})), \tag{51}$$

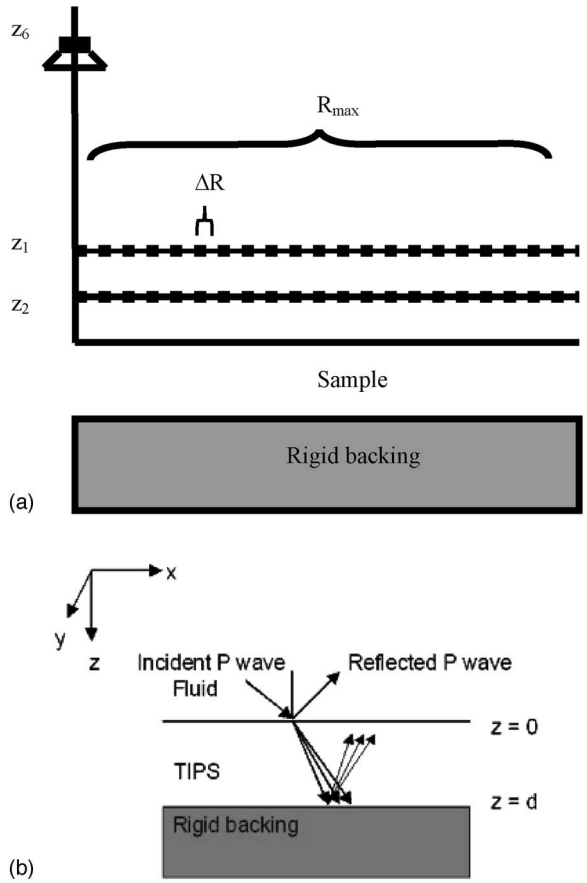


FIG. 1. (a) Experimental setup and indication of the respective geometrical parameters. (b) Geometry of model.

$$\Delta = -8\left[\sum_{k=1}^{3}\lambda_p(k^+)(\mu_{x,s}(k^+)\lambda_z(k^{++}) - \mu_{x,s}(k^{++})\lambda_z(k^+))\right] \times \left[\sum_{k=1}^{3}\lambda_x(k)(\mu_{z,s}(k^{++}) - \mu_{z,s}(k^+))\right]. \quad (52)$$

In all the preceding expressions, the + (respectively, ++) superscript corresponds to a first (respectively, second) following index in circular permutation of the {1,2,3} set (for instance, $2^+=3$, $2^{++}=1$). The new expressions of the $T_{i,j}$ are simpler than the previous expressions obtained in Refs. 4, 15, and 20. They can be more easily implemented in scientific programs.

D. Expression of the surface impedance

The porous layer of thickness H is bonded onto a rigid impervious backing (see Fig. 1). At $z=H$ on the rigid backing, the displacement components are equal to 0

$$\mathbf{V}(H) = [0 \quad 0 \quad 0 \quad p(H) \quad \hat{\sigma}_{xz}(H) \quad \hat{\sigma}_{zz}(H)]^t. \quad (53)$$

At $z=0$, the following conditions must be satisfied:

$$\mathbf{V}(0) = [\dot{u}_z^s(0) \quad \dot{u}_x^s(0) \quad \dot{u}_z^t(0) = v_z^{\text{air}} \quad p(0) = p^{\text{air}} \quad 0 \quad 0]^t, \quad (54)$$

where p^{air} and v^{air} are the pressure and the normal velocity in the free air at the interface with the porous material. The surface impedance is defined by $Z=p^{\text{air}}/v_z^{\text{air}}$ and Zv^{air} can be substituted for p^{air} in the preceding equations. The three displacement components at $z=H$ can be obtained from the components of $\mathbf{V}(0)$, leading to the following system of three equations.

$$T_{11}\dot{u}_z^s + T_{12}\dot{u}_x^s + (T_{13} + ZT_{14})v_z^{\text{air}} = 0, \quad (55)$$

$$T_{21}\dot{u}_z^s + T_{22}\dot{u}_z^s + (T_{23} + ZT_{24})v_z^{\text{air}} = 0, \quad (56)$$

$$T_{31}\dot{u}_z^s + T_{32}\dot{u}_x^s + (T_{33} + ZT_{34})v_z^{\text{air}} = 0. \quad (57)$$

The determinant of the system must be equal to 0 and Z is given by

$$Z = -\begin{vmatrix} T_{11} & T_{12} & T_{13} \\ T_{21} & T_{22} & T_{23} \\ T_{31} & T_{32} & T_{33} \end{vmatrix} \Bigg/ \begin{vmatrix} T_{11} & T_{12} & T_{14} \\ T_{21} & T_{22} & T_{24} \\ T_{31} & T_{32} & T_{34} \end{vmatrix}. \quad (58)$$

TABLE I. Acoustical and mechanical parameters of the porous material.

Thickness	H	cm	6
Frame density	ρ_s	kg/m^3	60
Porosity	ϕ		0.99
Flow resistivity (perpendicular)	σ^z	N m^{-4} s	17 000
Flow resistivity (parallel)	$\sigma^{x'y}$	N m^{-4} s	5000
Viscous dimension (perpendicular)	Λ^z	μm	140
Viscous dimension (parallel)	$\Lambda^{x,y}$	μm	126
Thermal dimension	Λ'	μm	150
Tortuosity (perpendicular)	α_∞^z		1.01
Tortuosity (parallel)	$\alpha_\infty^{x,y}$		1.01
Shear modulus (perpendicular)	L	kPa	50+i7
Shear modulus (parallel)	N	kPa	120+i22

III. PREDICTIONS AND MEASUREMENT OF THE SURFACE IMPEDANCE

A. Acoustical and mechanical parameters

The material is a layer of glass wool of thickness 6 cm. The layer is transversely isotropic with the symmetry axis, the z axis in Fig. 1, perpendicular to the surface. The acoustic parameters and rigidity coefficients that are given in Table I were all measured. Standard methods exist for measuring the flow resistivity and porosity. Tortuosity, viscous characteristic, and thermal characteristic length were measured using ultrasonic transmission methods (see Refs. 10, 11, and 21). These methods were originally proposed for isotropic porous materials. For the fibrous material under investigation, cylindrical samples were cut according to the principal axes. The acoustical parameters were determined for these cylindrical samples, on which also impedance tube measurements were performed. It was verified that the results of the impedance tube could be modeled by a numerically calculated absorption coefficient based on the individually measured acoustical parameters given in Table I. The shear modulus L in a plane perpendicular to the surface and the shear modulus N in a plane parallel to the surface have been measured at low frequencies using a similar technique as the one developed by Etchessahar *et al.*[20] The Poisson ratios are negligible for glass wools[12] and the rigidity coefficients $\hat{F}$ and $\hat{A}$ are equal

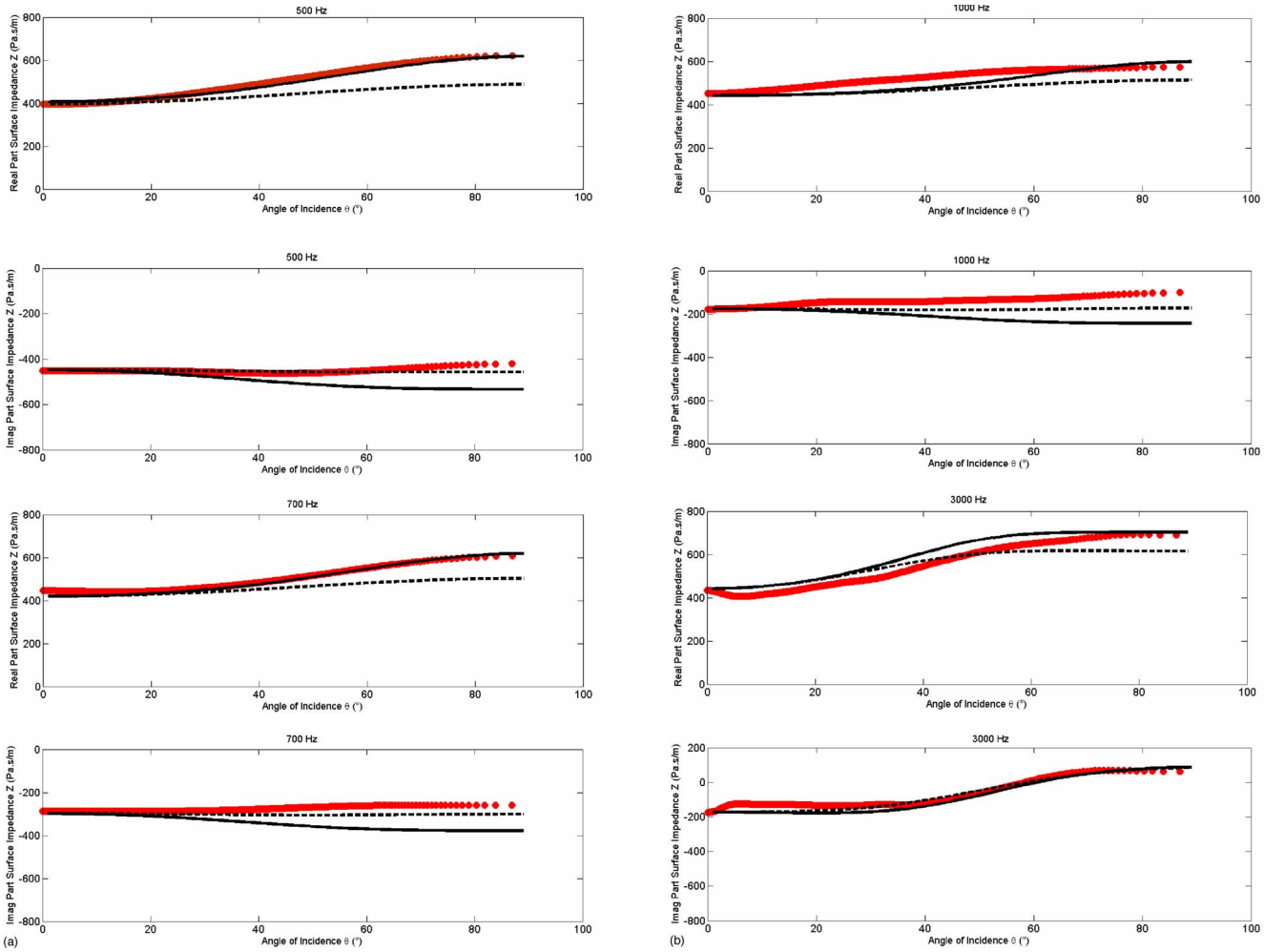


FIG. 2. (Color online) (a) Real and imaginary parts of the surface impedance for 0.5 kHz [(i) and (ii)] and 0.7 kHz [(iii) and (iv)] as a function of angle of incidence. Measurements are indicated by circles, dashed lines indicate calculated impedances for the isotropic case, and solid lines indicate the calculated impedances taking the anisotropy into account. (b) Real and imaginary parts of the surface impedance for 1 kHz [(i) and (ii)] and 3 kHz [(iii) and (iv)] as a function of angle of incidence. Measurements are indicated by circles, dashed lines indicate calculated impedances for the isotropic case, and solid lines indicate the calculated impedances taking the anisotropy into account.

to zero. Due to the large loss angle of the coefficients and the fact that the frame density is much larger than the air density, the frame displacement induced by a pressure field in the free air is very small compared to the displacement of the saturating air. The surface impedance will not strongly dependent on the rigidity coefficients, and $\hat{C}$ is arbitrarily set equal to $2L$, like if the meridian plane were an isotropic plane. A sensitivity analysis also showed that, for the fibrous material under investigation, the calculated surface impedance was only minor dependent on the measured values of tortuosity, viscous, and thermal characteristic lengths. The ratio of flow resistivities proved to have the largest influence on the calculated surface impedances of fibrous materials, which was also pointed out by Ref. 6.

B. Surface impedance measurement

The measurement of the surface impedance was performed by using the near-field holographic method introduced by Tamura *et al.*[22,23] The method initially relied on the decomposition of the wave field into its plane wave components by means of spatial Fourier transformation. In the present version of the Tamura method, the acoustic field created by the source is axisymmetric and the Fourier transform is replaced by the Hankel transform. A sketch of the experimental setup is represented in Fig. 1. The fibrous material is glued to a rigid backing and an unbaffled loudspeaker which is a dipole source with a good approximation that is placed at $z_s = 10$ cm above the sample. The sound pressure is measured at $z_1 = 10$ mm and $z_2 = 16$ mm. The pressure is measured at radial distances of the source ranging from $R = 0$ m up to $R_{\max} = 1.4$ m; the interval ΔR between two measurements is equal to 2 mm. The input signal is a sine sweep ranging from 100 Hz to 7 kHz. At each measurement, the transfer function between the measured and the input signal is evaluated. A time window is used to avoid unwanted reflections and a Hanning window is applied, as a function of the radial distance, to the amplitude of the measured pressure. The real and imaginary parts of the surface impedance of the fibrous material were measured for an angle of incidence varying

from 0° up to 81° and for frequencies from 300 Hz up to 6 kHz. The surface impedance as a function of the angle of incidence at 500 Hz, 700 Hz, 1 kHz, and 3 kHz, is presented in Fig. 2.

C. Comparison between predictions and measurements

The main effect of the anisotropy is that the real part of the surface impedance increases with the angle of incidence. This was already noted by Allard *et al.*[6] This effect is observed at 500 Hz, 700 Hz, 1 kHz, and 3 kHz. This is due to the relatively smaller value of ratio σ^{xy}/σ^{z} than one which is the isotropic case. It is more pronounced at medium frequencies of 500 and 700 Hz. The measurements in Fig. 2 are indicated by full circles, the dashed lines represent simulated surface impedances for the isotropic case, and the solid lines are the predicted surface impedances for the anisotropic case. The material data used in the calculation of the impedance of the isotropic material are the parameters measured in the z direction (indicated by the superscript z in Table I). Near normal incidence, the difference between measurement and the calculations, is negligible. With increasing angle of incidence, the difference between the isotropic case and the measured values increases due to the anisotropy. The influence of the anisotropy is most pronounced in the midfrequency range (500–700 Hz). The difference between the surface impedance calculated for the isotropic case, and the surface impedance measured and predicted when the anisotropy is taken into account, decreases with increasing frequency. Even if our model is more general than a transversely isotropic rigid frame model, it should be noticed that for the proposed example, the results are nearly similar thereby reducing the sensitivity to mechanical properties. A decrease in the imaginary part of the surface impedance with the angle of incidence, smaller than the increase in the real part, is predicted at 0.5, 0.7, and 1 kHz. This decrease does not appear in the measurements. This small discrepancy is probably due to a systematic error in the measurements mainly due to the difficulty of keeping constant the height of the microphones.

IV. CONCLUSION

A description of wave propagation in transversely isotropic porous materials was performed in terms of a TMM developed in the context of a recent formulation of the Biot theory. With the new formulation, the expressions of the matrix elements are simplified. As an illustration of the method, the surface impedance of a highly porous material was measured as a function of frequency and of the angle of incidence, and comparisons were performed with predictions obtained with the TMM. It was shown that the anisotropy can have a significant influence on the acoustical behavior of the material. A good agreement was found between theoretical and experimental results.

[1] M. A. Biot, "Theory of propagation of elastic waves in a fluid-filled saturated porous solid," J. Acoust. Soc. Am. **28**, 168–191 (1956).
[2] M. A. Biot, "Mechanics of deformation and acoustic propagation in porous media," J. Appl. Phys. **33**, 1482–1484 (1962).
[3] J. Carcione, "Wave propagation in anisotropic, saturated porous media: Plane-wave theory and numerical simulation," J. Acoust. Soc. Am. **99**, 2655–2666 (1996).
[4] A. K. Vashishth and P. Khurana, "Waves in stratified anisotropic poroelastic media: A transfer matrix approach," J. Sound Vib. **277**, 239–275 (2004).
[5] K. Liu and Y. Liu, "Propagation characteristic of Rayleigh waves in orthotropic fluid-saturated porous media," J. Sound Vib. **271**, 1–13 (2004).
[6] J. F. Allard, R. Bourdier, and A. L'Esperance, "Anisotropy effect in glass wool on normal impedance in oblique incidence," J. Sound Vib. **114**, 233–238 (1987).
[7] K. Attenborough, "Acoustical characteristics of porous materials," Phys. Rep. **82**, 179–177 (1982).
[8] D. Wilson, "Relaxation-matched modeling of propagation through porous media, including fractal pore structure," J. Acoust. Soc. Am. **94**, 1136–1145 (1993).
[9] M. Delany and E. Bazley, "Acoustical properties of fibrous absorbent materials," Appl. Acoust. **3**, 105–116 (1970).
[10] M. Melon, D. Lafarge, B. Castagnede, and N. Brown, "Measurement of tortuosity of anisotropic acoustic materials," J. Appl. Phys. **78**, 4929–4932 (1995).
[11] M. Melon, E. Mariez, C. Ayrault, and S. Sahraoui, "Acoustical and mechanical characterization of anisotropic open-cell foams," J. Acoust. Soc. Am. **104**, 2622–2627 (1988).
[12] V. Tarnow, "Dynamic measurements of the elastic constants of glass wool," J. Acoust. Soc. Am. **118**, 3672–3678 (2005).
[13] D. Johnson, J. Koplik, and R. Dashen, "Theory of dynamic permeability and tortuosity in fluid-saturated porous media," J. Fluid Mech. **176**, 379–403 (1987).
[14] Y. Champoux and J. F. Allard, "Dynamic tortuosity and bulk modulus in air-saturated porous media," J. Appl. Phys. **70**, 1975–1979 (1991).
[15] J. F. Allard, *Propagation of Sound in Porous Media: Modelling Sound Absorbing Materials* (Elsevier Aplied Science, New York, 1993).
[16] O. Dazel, B. Brouard, C. Depollier, and S. Griffith, "A alternative Biot's displacement formulation for porous materials," J. Acoust. Soc. Am. **121**, 3509–3516 (2007).
[17] L. M. Brekhovskikh, *Waves in Layered Media* (Academic, New York, 1960).
[18] C. Depollier, "Theorie de Biot et Prdiction des Proprits Acoustiques des Matriaux Poreux, Propagation dans les milieux Acoustiques Dsordonns (Biot's theory and properties of sound absorbing materials. Propagation in disorders porous materials)," thesis, Universit du Maine, France (1989).
[19] B. Brouard, D. Lafarge, and J. F. Allard, "A general method of modelling the acoustical properties of layered materials including fluid, elastic, and porous layers," in *ICA* (Trondheim, Norway, 1995).
[20] M. Etchessahar, S. Sarhraoui, L. Benyahia, and J. Tassin, "Frequency dependence of the elastic properties of acoustic foams," J. Acoust. Soc. Am. **117**, 1114–1121 (2005).
[21] P. Leclaire, L. Kelders, W. Lauriks, M. Melon, N. Brown, and B. Castagnede, "Determination of the viscous and thermal characteristic lengths by ultrasonic measurements in helium and air," J. Appl. Phys. **80**, 2009–2012 (1996).
[22] M. Tamura, "Spatial Fourier transform method of measuring reflection coefficients at oblique incidence. I Theory and numerical examples," J. Acoust. Soc. Am. **88**, 2259–2264 (1990).
[23] M. Tamura, J. F. Allard, and D. Lafarge, "Spatial Fourier transform method of measuring reflection coefficients at oblique incidence. II Experimental results," J. Acoust. Soc. Am. **97**, 2255–2262 (1995).

Effects of room acoustics on the intelligibility of speech in classrooms for young children

W. Yang and J. S. Bradley[a)]
National Research Council, Montreal Road, Ottawa K1A 0R6, Canada

(Received 10 September 2008; revised 8 December 2008; accepted 8 December 2008)

This paper reports new measurements of the intelligibility of speech in conditions representative of elementary school classrooms. The speech test material was binaurally recorded in simulated classroom conditions and played back to subjects over headphones. Subjects included grade 1, 3, and 6 students (6, 8, and 11 year olds) as well as adults. Recognizing that reverberation time is not a complete descriptor of room acoustics conditions, simulated conditions included realistic early-to-late arriving sound ratios as well as varied reverberation time. For conditions of constant signal-to-noise ratio, intelligibility scores increased with decreasing reverberation time. However, for conditions including realistic increases in speech level with varied reverberation time for constant noise level, intelligibility scores were near maximum for a range of reverberation times. Young children's intelligibility scores benefited from added early reflections of speech sounds similar to adult listeners. The effect of varied reverberation time on the intelligibility of speech for young children was much less than the effect of varied signal-to-noise ratio. The results can be used to help to determine ideal conditions for speech communication in classrooms for younger listeners. © *2009 Acoustical Society of America.* [DOI: 10.1121/1.3058900]

PACS number(s): 43.55.Hy [NX] Pages: 922–933

I. INTRODUCTION

Most classroom learning involves oral communication and the intelligibility of spoken words is obviously very important for a successful learning environment. The intelligibility of speech in classrooms is influenced most by the speech-to-noise ratio (S/N) at the listener's position and also by reflected sounds and the age of the listener. All three factors must be considered when determining optimum conditions for speech communication in classrooms.

The effects of S/N and the age of the listener were recently investigated in classrooms of grade 1, 3, and 6 students (6, 8, and 11 year olds).[1,2] In this previous work, speech intelligibility tests were performed by children listening naturally (binaurally) in their own classrooms with the natural ambient noises. The results gave a clear indication of the effects of both S/N and listener age on the resulting speech intelligibility scores and can contribute in determining optimum acoustical conditions for younger children.

Although the study tried to also examine the effects of varied room reverberation times, this was not successful because the 41 classrooms tested had similar and quite acceptable reverberation times. As a result, the current work was planned to consider the effect of varied room acoustics on the intelligibility of speech for children in school classrooms.

A number of previous studies have considered issues related to the effect of room reverberation on the intelligibility of speech in classrooms. However, the results of the various studies have some serious limitations.

Nábělek and Pickett[3] used a modified rhyme test with the speech and noise played back from two separate loudspeakers to investigate the effects of reverberation in classrooms. The test room had adjustable absorption making it possible to obtain conditions of 0.3 and 0.6 s reverberation times. Although increasing the reverberation time also increased the sound levels by about 2 dB (page 630 of Ref. 3), this effect was removed by adjusting the amplifier gains to create conditions with constant S/N. For the constant S/N conditions, the intelligibility scores increased for decreased reverberation time. However, if the natural increase in speech levels of 2 dB had been maintained for the 0.6 s reverberation time case, different results would have occurred with a reduced effect of varied reverberation time. The subjects were located approximately one critical distance from the loudspeakers and hence would have experienced approximately equal amounts of direct and reflected sound for an omnidirectional source. Because the loudspeakers used would be more directional than a human talker, subjects may have actually experienced predominantly direct sound. The study can be criticized as providing conditions that would not accurately reflect the effects of reverberation on natural speech in many classrooms. They did not consider cases where the possible benefits of reflected sounds were present and they did not include younger listeners.

Nábělek and Pickett also demonstrated the binaural advantage of listening with two ears compared to monaural listening. Their results clearly demonstrate that the results of monaural listening tests (e.g., Finitzo-Hieber and Tillman[4] and Johnson[5]) are not representative of normal listening conditions in rooms.

Neuman and Hochberg[6] assessed the effects of reverberation on the intelligibility of speech for children aged 5, 7, 9, 11, and 13 years old as well as adults. They used a speech test consisting of nonsense syllables and reverberation times of 0, 0.4, and 0.6 s. All speech samples were pre-

a)Electronic mail: john.bradley@nrc-cnrc.gc.ca

sented at the same level and in "quiet" conditions. They obtained increasing intelligibility scores with increasing age of the listeners and with decreasing reverberation time. They also demonstrated the advantage of binaural listening for the 0.6 s reverberation time case. This was similar to a constant S/N experiment except that the noise level was very low. It is not possible to estimate the combined effect of reverberation time and S/N from these results.

Although studies in actual classrooms would be expected to more realistically determine the combined effects of S/N and reverberation time, in previous efforts it was not possible to find test classrooms with a wide range of reverberation times, S/N, and ages of listeners. An earlier study by Bradley[7] determined the combined effects of *A*-weighted speech-noise level differences [S/N(A)] and reverberation times (T_{60}) for 12 to 13 year olds in their classrooms using regression analyses of combinations of predictors. Although S/N(A) values were the major determinant of intelligibility scores, reverberation time had a significant effect such that decreased reverberation time related to increased intelligibility scores. In a more recent classroom study,[2] there were effects that indicated small increases in intelligibility scores with decreased reverberation times but not for the youngest subjects, i.e., the grade 1 students. Both results indicated that for a given S/N, increased reverberation time led to decreased speech intelligibility scores.

Most previous studies of the effect of reverberation on speech intelligibility have been for constant S/N or quiet conditions with a presumably high S/N. None have specifically considered the possible benefits of added early-arriving reflected sounds that could increase effective S/N values. It has been shown for adult listeners that added early reflections arriving within about 50 ms after the direct sound have the same effect as increasing the level of the direct sound and hence the added early-arriving reflections can usefully increase the S/N by 7 dB or more.[8] However, it has at other times been argued that increased reflected sound would increase both speech and noise levels and would result in no change to S/N values. Hodgson and Nosal[9] explained that what is critical is the relative distances of the speech and noise sources from the listener. Their calculations, based on simple diffuse field theory, showed that when the noise source is closer to the listener than the talker, then added early reflections would usefully increase S/N values and hence would be expected to improve speech intelligibility. Yang and Hodgson[10] carried out speech intelligibility tests by auralizing virtual sound fields to support the earlier work.[9] Although they were largely successful, they did not give the actual signal-to-noise ratios of their conditions and they made no attempt to confirm that their conditions would represent the balance between early-and late-arriving sounds that would commonly occur in real rooms.

As the predominant source of interfering sound in classrooms is usually the children, it seems that the most common situation in elementary school classrooms is the case where the noise source is closer than the talker to the listener. For this case we would expect increased levels of early-arriving reflections to increase intelligibility scores because they would be relatively more important for the more distant speech source. Of course there are also many particular situations where early-arriving reflections are critical to understanding speech, such as when the talker is not facing the listener, or is at a more distant position in the classroom from the talker where the level of early-arriving speech energy can be as much as 7 dB or more greater than the direct sound.[8] These issues are rarely considered in more general discussions of classroom acoustics requirements, but commonly occur in classrooms.

This new research was planned to address several questions related to better understanding the effects of room acoustics on the intelligibility of speech for children in classrooms. It was thought important to understand the combined effects of reverberation time and S/N, which might occur in school classrooms, on the intelligibility of speech. In the new tests children should be listening naturally with two ears so that they could benefit from any binaural advantage that the realistic sound fields provided.

The new tests described in this paper were carried out using binaural playback of speech test material recorded in simulated conditions representative of real classrooms. Although tests in actual classrooms with varied T_{60} might be better, it was not possible to find the necessary combinations of room acoustic conditions and children's ages. Considerable effort was made to ensure that the simulated conditions were realistic representations of conditions in typical classrooms. Two types of combinations of T_{60} and S/N were created. In one series of conditions, reverberation time was varied and S/N was held constant. In a second series of tests, S/N values increased with the energy of the added reflected sound as longer T_{60} values were created. In a third experiment, some further conditions were created to determine how listeners benefited from added early reflections of the speech sounds. Tests were carried out on grade 1, 3, and 6 students (6, 8, and 11 year olds) as well as adults.

II. EXPERIMENTAL PROCEDURE

The experimental procedure was to carry out speech intelligibility tests on elementary school students using speech test material binaurally recorded in simulated conditions representative of those in real classrooms.

A. Requirements for simulated classroom acoustics conditions

The intelligibility of speech is related to the level of the speech relative to the level of concurrent interfering noises. However, not all speech sound increases the intelligibility of the speech. Increased levels of the direct speech and early reflections of the speech arriving within about 50 ms after the direct sound lead to increased intelligibility, but later-arriving reflections reduce the intelligibility of the speech.[8] In simulating room acoustics conditions it is not good enough to simply vary reverberation times. It is possible to create unrealistic conditions with too much or too little early reflection energy that will lead to results that are not representative of conditions in actual classrooms.

The relative level of early reflection energy can be measured by C_{50} values, where C_{50} is an early-to-late-arriving

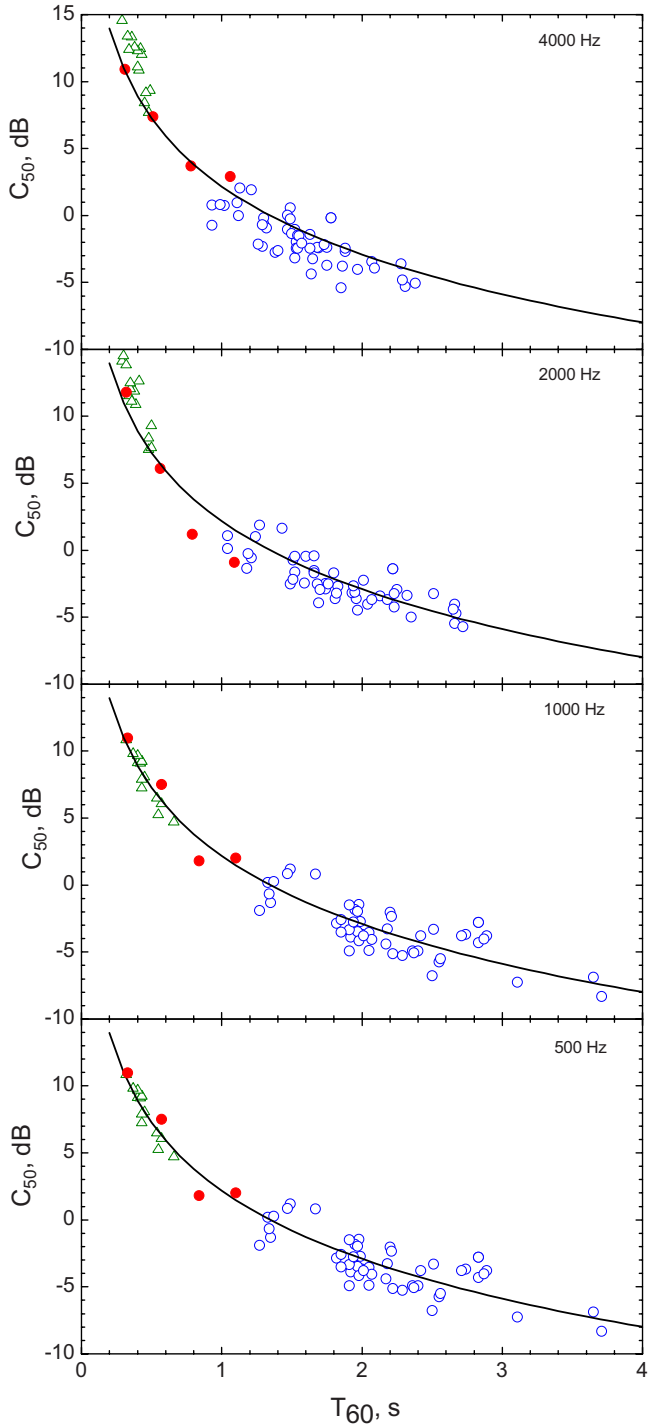


FIG. 1. (Color online) Measured octave band values of C_{50} plotted vs the corresponding T_{60} values. Open triangles: classroom data, open circles: measured auditorium data, closed circles: simulated sound fields, and solid line: best-fit regression line.

sound energy ratio with a 50 ms early time interval.[11] When simulating conditions with varied reverberation time (T_{60}), it is important that C_{50} values are also appropriate for the corresponding T_{60}. Figure 1 illustrates combinations of C_{50} and T_{60} obtained from measurements in both classrooms[2] and auditoria.

It was desired to create test conditions with T_{60} values of 0.3, 0.6, 0.9, and 1.2 s, which were thought to correspond to the full range of likely conditions in typical elementary school classrooms. A T_{60} of 0.6 s is often thought to be near optimum[7] and is referred to in the ANSI S12.60 classroom acoustics standard.[12] A T_{60} of 0.3 s is representative of the lowest T_{60} values likely to be found in a normal classroom. T_{60} values of 0.9 and 1.2 s could occur in real classrooms but were expected to lead to increasingly less suitable conditions with lower speech intelligibility scores. Figure 1 also shows the combinations of measured C_{50} and T_{60} for the four simulated conditions. They are seen to be close to the mean trend of the results from the real rooms and hence corresponded to realistic ratios of early- and late-arriving reflections.

One set of test conditions included these four T_{60} values and with a constant S/N. These would represent conditions in which the added reflected sounds equally influenced speech and noise levels. A second set of conditions was created in which speech levels increased as more reflected sound was added, while noise levels were held constant, leading to varied S/N. It was important to ensure that the increased speech levels with increasing T_{60} values realistically represented what would occur in real rooms.

The desired increase in speech levels with increasing T_{60} was determined from Beranek's compilation of measurement data. Figure 9.4 of Ref. 13 plotted values of (EDT/V) versus G_{mid} levels. (EDT is the early decay time (s), V is the room volume (m^3), G_{mid} is the relative level or strength of the sound in the rooms and both EDT and G_{mid} are for combined 500 and 1000 Hz octave band results.) (See Ref. 11 for definitions of EDT and G.) Beranek's plot relates the average variation in decay times to the average variation in levels for a large number of auditoria. These data were combined with data from several classroom-sized rooms and a new regression line fitted to the combined data, which was only very slightly different than Beranek's original line for large auditoria. The new line was used to predict the desired increases in level with varied decay time in the simulated conditions.

Beranek did not give the equation of his best-fit line but it was determined from his text and a manual fit of points from the line which indicates it is

$$10\log\{(\mathrm{EDT}/V)10^6\} = G_{\mathrm{mid}} + 16. \quad (1)$$

Because we would like to predict G_{mid} values it is necessary to reverse x and y values as follows:

$$G_{\mathrm{mid}} = 10\log\{(\mathrm{EDT}/V)10^6\} - 16. \quad (2)$$

Fitting this form of equation to Beranek's large hall data combined with data for classroom-sized rooms resulted in the following relationship:

$$G_{\mathrm{mid}} = 10.75\log\{(\mathrm{EDT}/V)10^6\} - 17.6. \quad (3)$$

Equation (3) was used to estimate increases in sound levels with increasing decay time for a 198 m^3 room volume which was the average room volume of the 41 elementary school classrooms recently studied.[1] The expected increases in level associated with the increased decay times using Eq. (3) are listed in Table I and plotted in Fig. 2.

For comparison, the expected changes of level with decay time were also calculated using Barron's revised theory and diffuse field theory[14,15] using a source-receiver distance of 5 m as representative of an average seat in a classroom.

TABLE I. Expected increases in sound levels with increasing decay time relative to the case of a T_{60} value of 0.3 s as well as the level increases measured in the simulated sound fields.

	Increases in $G(500,1000)$ (dB)			
T_{60} (s)	Beranek	Barron	Diffuse	Measured
0.3	0.0	0.0	0.0	0.0
0.6	3.2	3.7	2.8	2.8
0.9	5.1	5.8	4.5	5.5
1.2	6.5	7.2	5.8	7.3

These calculated level changes were based on the measured T_{60} values of the simulated sound fields and the results are also included in Table I and Fig. 2. The three approaches led to similar predicted increases in levels. Although the three sets of calculated level increases are all based on measured decay times, Beranek's relationship was based on EDT values while the others were based on measured T_{60} values. However, the changes in levels were expected to be similarly related to EDT and T_{60} values. Measurements of the simulated sound fields demonstrated that the increases in speech levels varied in a similar manner, as shown in Table I and Fig. 2.

B. Subjects and speech test procedures

The word intelligibility by picture identification (WIPI) speech test was used because it is a very simple test that 6 year olds and older students can quickly learn and respond to individually without significant training. It includes four lists of 25 phonetically balanced simple nouns.[16,17] The test words were each presented in the carrier phrase, "Please mark the — now" spoken by a clear speaking female voice. These tests used exactly the same speech test recordings as in the previous classroom studies.[1]

In the previous classroom study[1,2] students carried out the tests as groups while seated in their regular seats in their own classrooms and marked responses in a book of pictures illustrating the possible responses. In the current investigations, individual students were tested one at a time using headphone (HP) presentation of the speech material. The subject and experimenter were located in a quiet room without acoustical distractions and with no other people present. The processed speech test material was stored as wav files on a portable computer. These were presented to each listener using specially developed software that played the speech files and displayed the pictures corresponding to the possible six responses on a touch screen. The listener touched one of the six pictures to indicate the correct response. The program put an X through the touched picture to confirm which had been selected. The students found the test easy to perform and seemed to regard it as simple computer game.

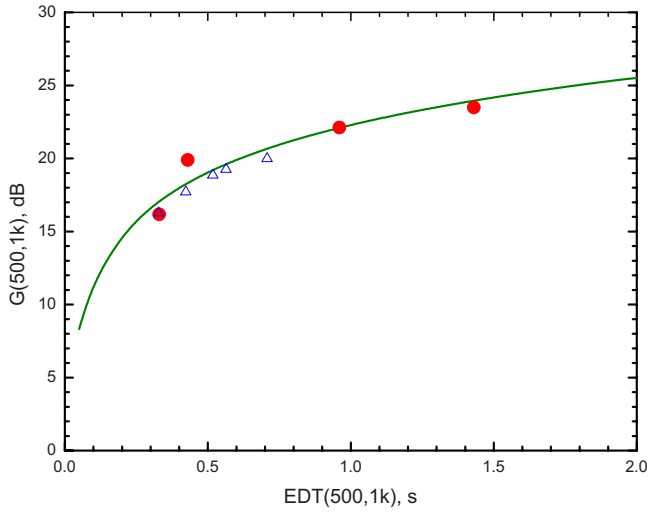


FIG. 2. (Color online) Variation in sound levels with decay time plotted as midfrequency G values vs measured midfrequency EDT values corresponding to the conditions with T_{60} values of 0.3, 0.6, 0.9, and 1.2 s for 198 m^3 room. Solid circles: simulated conditions, open triangles: classroom-sized room data, and solid line: Eq. (3).

All subjects first carried out a short practice test to be sure that they were familiar with the process of the test. If they had no problems they would then carry out the actual tests. The grade 1 students each carried out tests of three different conditions. The grade 3 and 6 students as well as the adults each carried out tests for four different conditions. The use of each of the four word lists was rotated so that all four word lists were used an approximately equal number of times to assess the nine different acoustical conditions by each age group of subjects.

The students were from several schools in the Ottawa Carleton District School Board (OCDSB). Permission to invite schools and students to participate in our tests was obtained from the OCDSB Research Advisory Committee. Ethics approval was obtained from both the University of Ottawa Research Ethics Board (protocol H 03-07-06) and the National Research Council Research Ethics Board (protocol 2007-10). All students volunteered to participate with the written consent of their parents. Adult participants volunteered and each signed consent forms. A total of 77 grade 1 students, 75 grade 3 students, and 65 grade 6 students participated. In addition 17 adults were tested.

C. Sound field simulation and headphone playback procedure

Conditions simulating those in classrooms were created using an eight channel electroacoustic sound field simulation system located in an anechoic room and quite similar to a previously described system.[8] The system consisted of eight Tannoy model 800A loudspeakers that surrounded the listening position. Five of the loudspeakers were in the horizontal plane of the listener's ears and the other three were raised up above this plane in front of the listener.

The signals to each loudspeaker were processed by four Yamaha DME32 digital signal-processing units connected together to form one large unit. A direct speech sound arrived first from the loudspeaker directly in front of the listening position. A total of 31 early reflections were created that arrived from the eight loudspeakers within 50 ms after the direct sound and realistically decreased in level with increasing time. Reverberant decays followed the discrete early reflections. Reverberation times were varied by varying the decay times of the digital reverberator components in the DME32 units. Adjusting the balance between the combina-

tion of direct sound and early reflections versus late-arriving sound made it possible to adjust C_{50} values independent of T_{60} values to create the desired combinations of C_{50} and T_{60} values in each octave band from 125 to 8000 Hz. This setup made it possible to systematically vary the most relevant aspects of the sound fields and to ensure that realistic combinations were obtained.

To record speech test material for each test condition, an acoustical mannequin (Brüel and Kjaer type 4128) was placed at the listener position. For the younger (and smaller) listeners, a smaller head would have been desirable but such heads are not commercially available. The speech test material was played through the eight loudspeakers of the simulation system and recorded using the microphones in the acoustical mannequin. It was subsequently played back to listeners over Sennheiser type HD280 HPs. In recording the speech in this way, the frequency response of the speech was modified by the frequency response of the acoustical mannequin. When playing the recordings back to subjects, the frequency response of the speech was further modified by the characteristics of the HPs. The frequency response of the test speech material was corrected by measuring the combined response of the HPs and acoustical mannequin. The transfer function of the combined HPs and acoustical mannequin was obtained by measuring the impulse response of the HPs while placed on the acoustical mannequin.

One of the major difficulties of using HP playback is that repositioning the HPs leads to different HPs transfer functions and in some cases these differences can be quite large.[18] Initial tests confirmed that large variations in the measured transfer functions are possible and to minimize these effects, the head and HP transfer function was determined from the average of ten different placements of the HPs on the acoustical mannequin. For each repeat, the HPs were carefully positioned on the mannequin so that the HP cushions completely covered the pinna of the acoustical mannequin. The average transfer function was determined from the average measured impulse response after carefully aligning the start of each measured impulse response.

The recorded speech test material was equalized to correct for the head-HP transfer function. This was done by deconvolution of the recorded speech with the average measured impulse response of the head-HP combination to extract the effects of the head and HPs from the recorded speech. The process was evaluated by comparing the levels of speech initially recorded at the mannequin with the levels of the same speech played back from HPs after processing and again recorded at the microphones of the mannequin. The differences are plotted as 1/3-octave band levels in Fig. 3 for conditions 1–4 having T_{60} values of 0.3, 0.6, 0.9, and 1.2 s (see Sec. II D and Table II for description of conditions). For frequencies from 250 to 6300 Hz inclusive the differences were 1 dB or less. However, the HP playback always had slightly lower levels with an average difference over the 250–6300 Hz range of 0.6 dB. This was thought to be due to using an average correction. There were larger differences at frequencies below 250 Hz and these differences increased with decreasing reverberation time of the test condition. These effects are not important for speech

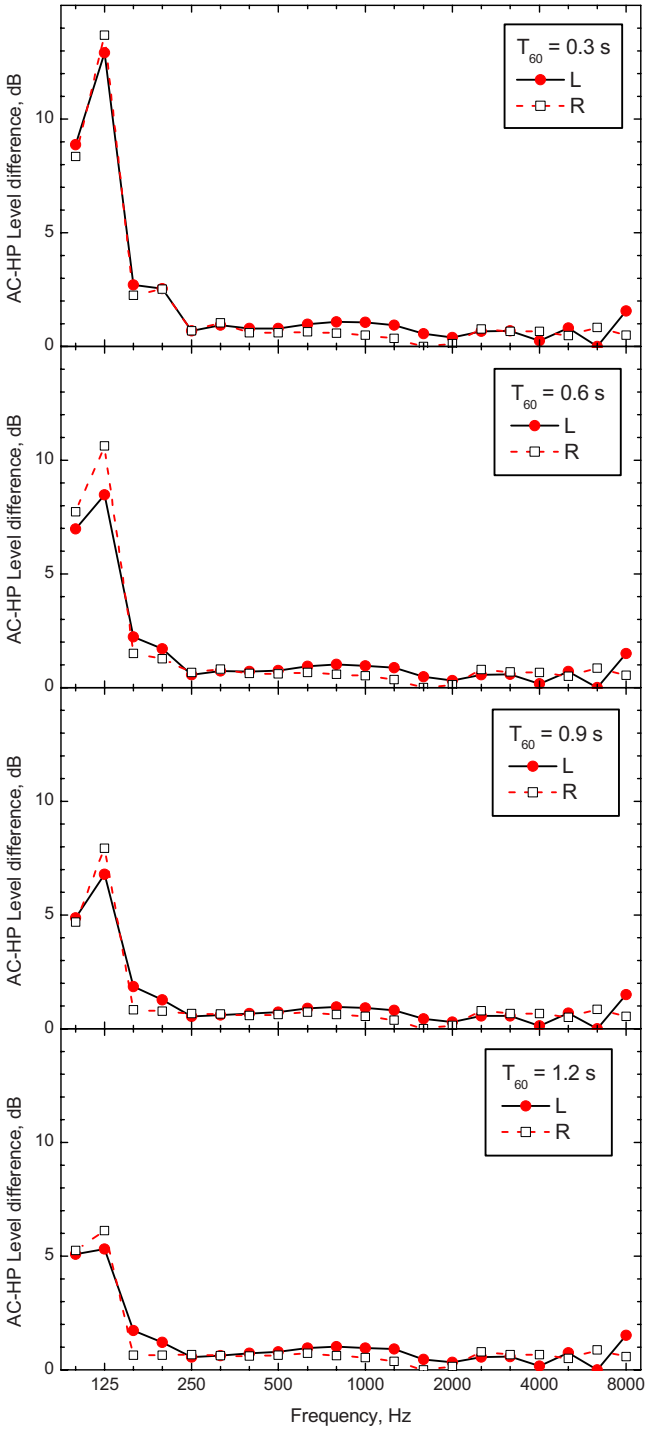


FIG. 3. (Color online) Level differences between 1/3 octave band speech levels of the initial acoustical mannequin recordings (AC) and recordings of the processed initial recordings played back over HPs. The differences for the left (L) and right (R) ear recording for conditions 1–4 having T_{60} values of 0.3–1.2 s are shown.

intelligibility[19] but are similar to previous observations that auralization of more absorptive conditions can be more difficult.[20]

Simulated ambient noise was separately recorded binaurally in a similar manner. Noise with a −5 dB per octave spectrum shape was produced and radiated incoherently from all eight loudspeakers in the sound field simulation system. This spectrum shape has been shown to approximate typical

TABLE II. Description of the nine acoustical conditions used in the speech tests.

Condition	T_{60} (s)	Speech level (dBA)	Noise level (dBA)	S/N(A)
1	0.3	62	67	−5
2	0.6	65	67	−2
3	0.9	67	67	0
4	1.2	69	67	2
5	Direct only	60	67	−7
6	Direct+early	66	67	−1
7	0.3	62	64	−2
8	0.9	67	69	−2
9	1.2	69	71	−2

indoor ambient noise such as that from ventilation systems and is often referred to as a "neutral" spectrum.[21,22] The binaural noise recordings were corrected for the response of the HPs and mannequin as described for the speech sounds. The noise recordings were mixed with the speech recordings at levels to provide the desired signal-to-noise ratios.

D. Test conditions

Speech tests were carried out for nine different acoustical conditions making it possible to carry out three different experiments. Table II summarizes the nine test conditions.

Conditions 1–4 were used in experiment No. 1 in which reverberation time was varied (0.3, 0.6, 0.9, and 1.2 s) and the ambient noise level was held constant. As a result, S/N(A) increased with increasing reverberation time representing the expected increase in speech level due to the addition of reflected speech sounds with increasing T_{60}.

Conditions 7, 2, 8, and 9 were used in experiment No. 2. Again reverberation time was varied (0.3, 0.6, 0.9, and 1.2 s) but the S/N(A) was kept constant in this experiment. This experiment would represent the condition where added reflected sound leads to equal increases in both speech and noise levels.

Experiment No. 3 included conditions 5, 6, and 3. Condition 5 included direct speech sound only. In condition 6 early reflections were added which increased the total sound level. Finally, condition 3 had the same level of direct sound and early reflections, but with added late-arriving sound. The ambient noise level was held constant and hence the overall speech levels increased as reflected sounds were added. This experiment was intended to determine whether young children benefit from added early-arriving reflections in a manner similar to adults.

The number of subjects tested for each of the nine conditions varied a little with the age of the subjects and slightly among the different conditions for each age group as summarized in Table III.

E. Validation of headphone playback procedure

Acoustical conditions A–D were used in initial tests to validate that the HP playback process led to the same intelligibility scores as direct playback of speech sounds for the same acoustical conditions. Conditions A–D were the same as conditions 1–4 except that some S/N(A) values were a little different. The comparison test used 16 adult subjects who each carried out the tests both by direct listening in the anechoic chamber (AC) simulation system and also by listening over HPs. Figure 4 shows that the mean scores for each condition were very similar for the two types of playback of the speech and noise sounds.

TABLE III. Number of subjects (N) that participated in each test condition for each age group.

Age	N
Grade 1	24–26
Grade 3	29–36
Grade 6	26–31
Adults	14–16

The differences were tested using a paired-sample t-test. When all conditions were included as a group, there was not a significant difference between the two playback methods. When the pairs of results for each of the four acoustical conditions were separately tested, in all cases there were no significant differences between the two playback methods. That is, we can be reasonably confident that our processed recordings played back over HPs were equally intelligible to the speech in the original sound fields. This confirmed earlier exploratory studies to consider the viability of the HP playback method.[23]

Marshall[16] found that the four word lists of the WIPI test did not yield identical scores for evaluations of the same acoustical conditions. As all of the four word lists were used approximately equally for each acoustical condition, it was possible to compare the mean scores from each word list averaged over all acoustical conditions. This was done first for the adult listeners so that they could be used for the results of the initial validation tests of the playback procedure. Table IV lists the mean adult scores for each of the four word lists of the WIPI test averaged over all acoustical conditions. This is followed by the corresponding corrections for

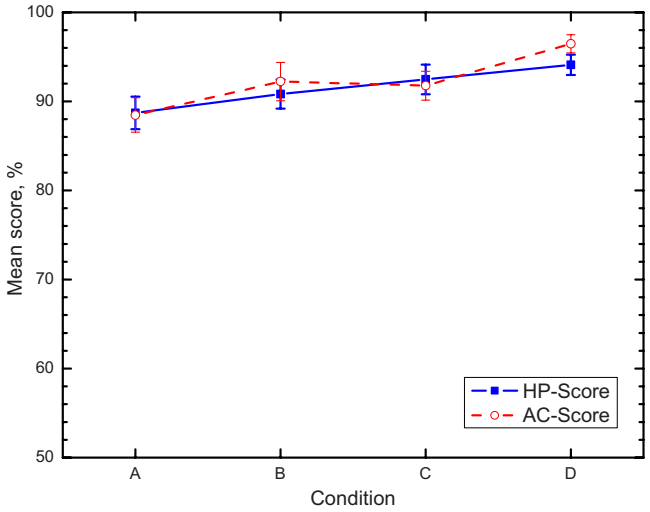


FIG. 4. (Color online) Comparison of mean speech intelligibility scores for HP playback and direct playback in the AC simulation system. Error bars indicate the standard errors of each of the mean values.

TABLE IV. Mean scores for each word list and correction factors of the WIPI test for adult listeners, followed by correction for adults, children, both (adults and children), and Marshall's corrections from Table 22 of Ref. 16.

Word list	Mean score	Correction adults	Correction children	Correction both	Correction Marshall
1	93.78	1.0509	1.0621	1.0565	1.0115
2	90.59	1.0153	1.0249	1.0201	0.9606
3	82.63	0.9259	0.9112	0.9186	0.9022
4	89.94	1.0079	1.0018	1.0048	1.1257
Average	89.233	1.0000	1.0621	1.0565	1.0115

the adult data for the children's responses and for the combined adult's and children's responses ("both"). (Of course the corrections for children and both adults and children were determined later but are included here for easy comparison.) Marshall's corrections for children aged 5–11 years old are included in the final column and are seen to be reasonably similar to those for children from the current study.

The corrections indicate how the average response for each word list differed from the average of all word lists. The adult corrections shown in Table IV for adult subjects were used to correct the scores from the validation test results by dividing each score by the appropriate correction value depending on the word list that was used. The resulting corrected scores are shown in Fig. 5.

The corrections result in a little closer agreement between the two sets of data. However, the pairs of results were not significantly different before correcting for word list differences and were again not significantly different after correcting for the word list differences (paired-sample *t*-tests). The results do suggest that there is a small benefit in correcting scores for word list differences, and the mean squared difference between HP and direct acoustic playback was reduced from 2.02 to 1.57 when the scores were adjusted to correct for word list differences.

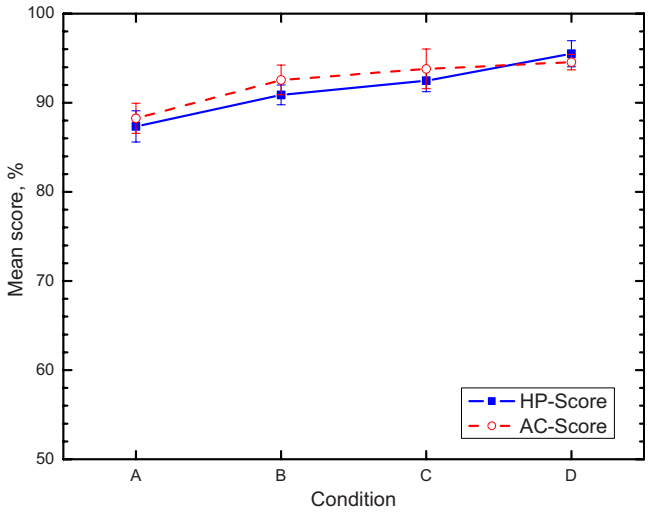


FIG. 5. (Color online) Comparison of corrected mean speech intelligibility scores for HP playback and direct playback in the anechoic room simulation system (AC). Error bars indicate the standard errors of each of the mean values.

The differences among the word lists may be due to a number of factors. Most obvious would be the different test words that make up each list. Some lists may contain a few more difficult words than other lists. However, there may also be differences related to how well the talker spoke each test word and how well each test word was recorded. In addition the age of the listener may influence the corrections because younger listeners would be more affected by more difficult words. The corrections included in Table IV were probably influenced by all of these factors and so we would not expect our new corrections to be the same as Marshall's.

III. RESULTS OF THE THREE MAIN EXPERIMENTS

The results of all three experiments described in Secs. III A–III C were first analyzed in terms of the uncorrected speech intelligibility scores and subsequently using the corrected scores as described in Sec. II using the both correction values from Table IV. In all cases using the corrected scores did not change the pattern of results but led to small improvements in the significance of the results. Therefore, to avoid unnecessary confusion, the following results of the three main experiments are described only in terms of the corrected scores.

A. Experiment No. 1 (varied S/N)

In experiment No. 1 subjects listened to speech for conditions 1–4 (described in Table II). These were conditions of varied T_{60} for constant noise level resulting in varied S/N as might occur when added room reflections of speech sounds increase the effective S/N. An analysis of variance of the corrected speech intelligibility scores indicated significant main effects of age ($p<0.001$) and condition ($p<0.003$). There was not a significant interaction effect. A Tukey honestly significant difference (HSD) *post hoc* test of the data indicated that the differences between each of the four age groups were all significant ($p<0.014$ or better).

The mean corrected scores are plotted versus condition for each age group in Fig. 6. The error bars show the standard errors of each mean score. A fifth line in Fig. 6 plots the average results over all age groups versus acoustic condition. Although there are not large variations in the scores with varied T_{60}, the average of all ages tends to peak for condition 3 with a T_{60} of 0.9 s. For these cases, where added reflected

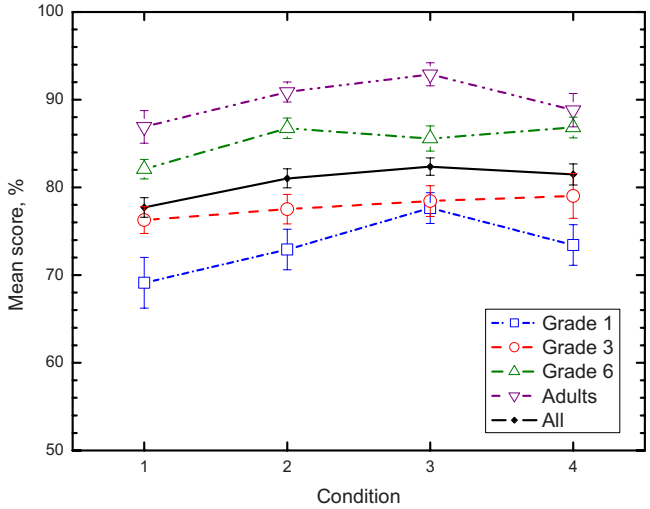


FIG. 6. (Color online) Mean corrected scores for conditions 1–4 having T_{60} values of 0.3, 0.6, 0.9, and 1.2 s, respectively. Each line refers to the data from a different age group and the error bars are the standard errors of each mean value. A fifth line indicates the averages of all four age groups.

sound increased both S/N and T_{60}, there is a range of conditions that lead to approximately the same speech intelligibility scores within each age group.

B. Experiment No. 2 (fixed S/N)

Experiment No. 2 included conditions 7, 2, 8, and 9 that had a constant S/N(A) of −2 dB for cases with T_{60} varying from 0.3 to 1.2 s, as described in Table II. An analysis of variance of the corrected scores for all age groups and these four conditions indicated highly significant main effects of both condition and age ($p<0.001$). There was not a significant interaction effect. A Tukey HSD *post hoc* test of the data indicated that the differences between pairs of the four age groups were all significantly different ($p<0.001$ or better). The mean values and their standard errors for the corrected scores are plotted in Fig. 7.

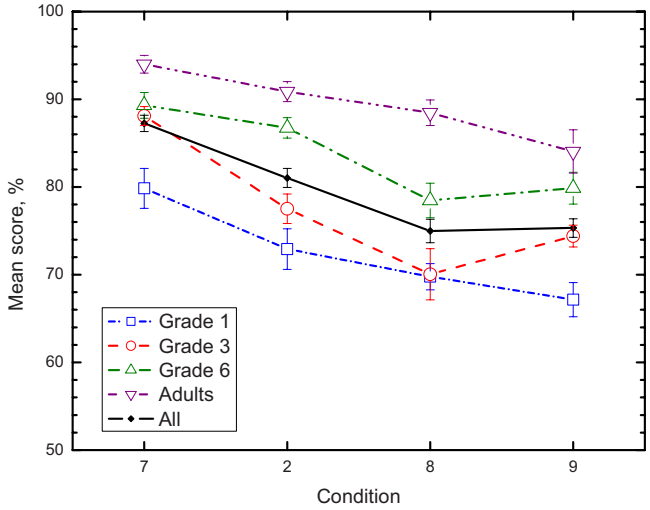


FIG. 7. (Color online) Mean corrected scores for conditions 7, 2, 8, and 9 having T_{60} values of 0.3, 0.6, 0.9, and 1.2 s, respectively, and a constant S/N(A)=−2 dB. Each line refers to the data from a different age group and the error bars are the standard errors of each mean value. A fifth line indicates the averages of all four age groups.

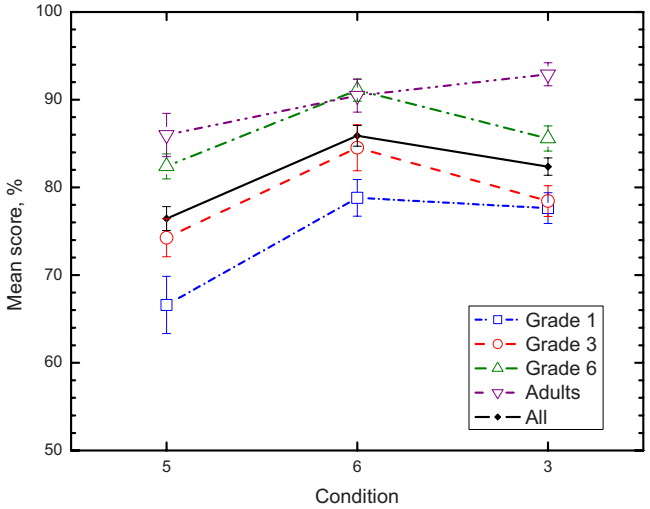


FIG. 8. (Color online) Mean corrected scores for condition 5 (direct sound only), condition 6 (direct sound and early-arriving reflections), and condition 3 (direct sound with early and late-arriving reflections). Each line refers to the data from a different age group and the error bars are the standard errors of the mean values. A fifth line indicates the averages of all four age groups.

When the S/N is held constant, as in these results, there is no beneficial effect of increased reflected speech sound and there is a trend for speech intelligibility to decrease with increasing reverberation time.

C. Experiment No. 3 (added reflections)

Conditions 5, 6, and 3 were used in experiment No. 3 to examine the basic effects of first adding early-arriving reflections to the direct sound, and second adding late-arriving reflections. By comparing the results from condition 6 with those of condition 5 we can determine the effects of adding early-arriving reflections to the direct speech sounds. An analysis of variance of the corrected results from conditions 5 and 6 showed that there were significant changes in the intelligibility scores with condition ($p<0.001$) and age ($p<0.001$) but no interaction effect. The lack of a significant interaction effect indicates that all ages of listener benefited equally when early-arriving reflections were added. A Tukey HSD *post hoc* test of these data indicated that the grade 6 and adult results were not significantly different but the results of all other age groups were different from each other ($p<0.011$ or better).

The mean corrected scores are plotted in Fig. 8. Adding early reflections increased speech intelligibility for all age groups but the scores of the adults were not significantly different than those of the grade 6 students.

Comparing the scores from conditions 3 and 6 makes it possible to determine the effect of adding late-arriving speech sounds with a 0.9 s reverberation time. An analysis of variance of the corrected data from these two cases indicated a significant effect of age ($p<0.001$) but no significant effect of condition. A Tukey HSD *post hoc* test showed that the age differences were not significant for all age groups. The results of the grade 1 and 3 students were not significantly different and the results of the grade 6 and adult listeners were not significantly different, but other differences among

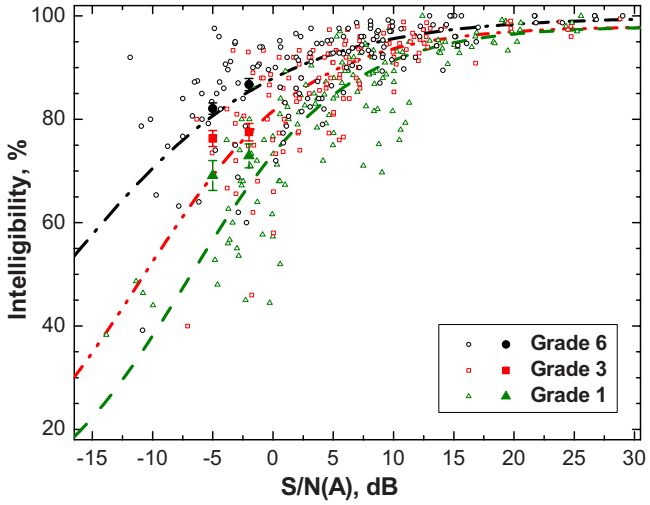


FIG. 9. (Color online) Comparison of mean speech intelligibility scores from conditions 1 and 2 (T_{60} 0.3 and 0.6 s) with previous classroom study results. Large filled symbols are the new results; small open symbols and regression lines are from the previous classroom study (Fig. 2 of Ref. 1).

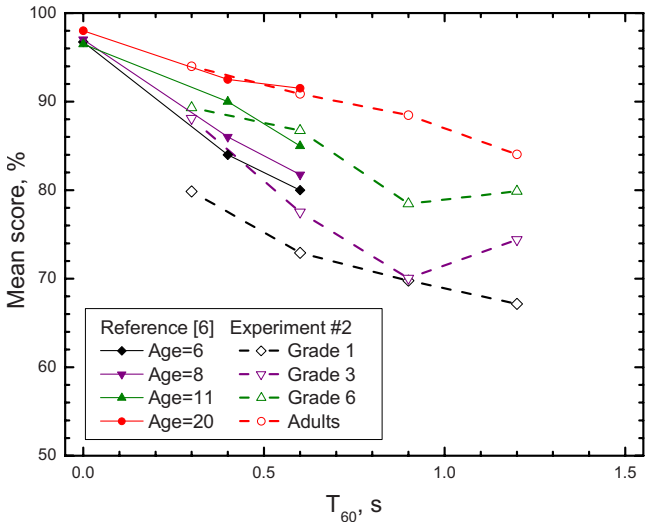


FIG. 10. (Color online) Comparison of experiment No. 2 results with those of Neuman and Hochberg (Ref. 6). The 6 and 8 year old data were from interpolations of Neuman and Hochberg's data for 5, 7, and 9 year olds.

age groups were significant. Adding late-arriving reflections did not significantly change speech intelligibility scores even though the overall speech level increased when the reverberant speech was added. The averages of all age groups shown in Fig. 8 suggest a small decrease in intelligibility but this was not statistically significant (i.e., $p=0.07$).

IV. DISCUSSION

A. Comparisons with previous results

It is of interest to compare the new results from the current study with previous results to confirm that they are representative of children's experience in real classrooms and that the effects of reverberation are similar to those in previous studies.

Previous speech intelligibility tests in classrooms[1,2] related speech intelligibility scores using the WIPI test to S/N(A) values. In the previous classroom study, the predominant source of interfering sound was concluded to be the children because occupied noise levels were higher than unoccupied noise levels even when the children were inactive and quiet.[2] Therefore we can assume that the results of experiment No. 1 are most representative of the conditions in the classrooms. Figure 9 compares mean speech intelligibility scores from the current study with the speech intelligibility scores versus S/N(A) values for grade 1, 3, and 6 students from the previous classroom study.

For each age group in the current study, the results of conditions 1 and 2, corresponding to T_{60} values of 0.3 and 0.6 s, are plotted in Fig. 9 at the appropriate S/N(A) values. The mean occupied classroom reverberation time was 0.41 s (Ref. 2) and was intermediate to the two conditions plotted from the current data. For the grade 6 results there is near perfect agreement between the current results and the classroom study results. The grade 3 results from the current study indicate slightly higher mean intelligibility scores than the classroom study and the grade 1 results indicate a little larger difference. The two studies used exactly the same speech test material, the same age groups, and the acoustical conditions of the new study were intended to closely model those in classrooms. However, there are other differences that might have affected the youngest listeners. While the acoustical conditions may have been quite similar, in the actual classrooms there were many other forms of distraction that might have reduced the scores of the youngest listeners. These other distractions would include visual distractions such as those of the other children's actions. In addition, the interfering sounds in the classroom were not always meaningless broadband noise, but at times were recognizable sounds from both within their classroom and from adjacent spaces. These may have had a larger negative effect on speech intelligibility scores. Considering the differences in the two experimental procedures, the agreement is very good and confirms that classroom conditions were accurately simulated.

There are little previous data available that can be compared with the current results indicating the effects of varied reverberation time for young children in conditions representative of classrooms. Most previous studies have included major procedural differences such as monaural presentation of the speech, different speech test material, or quite different and often unrealistic acoustical conditions. In spite of some differences in experimental methods, the current results of experiment No. 2 were compared with the results of Neuman and Hochberg[6] in Fig. 10.

Neuman and Hochberg tested children aged 5, 7, 9, 11, and 13 years old as well as adults. They included three acoustical conditions corresponding to no reverberation and reverberation times of 0.4 and 0.6 s. However, they did not specify the ambient noise level during the tests and only indicated it to be quiet. In addition, their speech test material was different than in the current study and consisted of nonsense syllables.

To obtain more comparable results, their scores for 5, 7, and 9 year olds were interpolated to get values representative of 6 and 8 year olds. Figure 10 indicates reasonable agree-

ment in the general trends of the data with intelligibility scores increasing with decreasing reverberation time. The adult and 11 year old (grade 6) data for the two studies agree very well for comparable T_{60} values. The results of the 8 year olds (grade 3) indicate some differences and for the data of the 6 year olds (grade 1), the current study produced much lower speech intelligibility scores. This is probably largely due to different signal-to-noise ratios between the two tests, which would more adversely affect the youngest listeners.[1] In experiment No. 2 the S/N(A) was −2 dB and was presumably much lower than for the Neuman and Hochberg results in quiet conditions. In view of the significant differences in the procedures of the two studies, the agreement seems reasonably good and generally indicates the same effects of reverberation for cases of constant S/N.

The results of experiment No. 3 cannot be directly compared with previous results because no previous study could be found that considered whether young children benefit from added early-arriving reflections of speech sounds. Although studies with adults have clearly demonstrated that the added energy of early-arriving reflections within about 50 ms of the direct sound increases speech intelligibility equivalent to a similar increase in the direct sound level,[8] this has not been demonstrated for children. The results of experiment No. 3 confirm that children do benefit as much as adults do when early reflections are added. The nonsignificant effect of adding later arriving sound is also similar to previous results for adult listeners.[8]

B. Determining ideal conditions for speech communication in classrooms

In experiment No. 1, speech intelligibility scores tended to peak at some intermediate T_{60} value as expected for the conditions with varied S/N, but there were not large variations in intelligibility scores over the included range of T_{60} values. Because the experiment No. 1 results shown in Fig. 6 were based on data from only four conditions, and a small range of T_{60} values, it is difficult to accurately determine the mean trends.

To obtain a better estimate of the mean trends, the speech intelligibility scores for all of the nine conditions were plotted versus the corresponding useful-to-detrimental sound ratios (U_{50}) for each of the conditions. This made it possible to use nine data points rather than four to determine the mean trends of the data. It is well known that U_{50} values can explain the combined effects of varied S/N and varied T_{60} on speech intelligibility scores.[7,9,24,25] Useful-to-detrimental sound ratios were calculated from measured C_{50} values along with measured speech and noise levels in the six octave bands from 125 to 4000 Hz as described in Ref. 25. The octave band U_{50} values were arithmetically added with a uniform frequency weighting. The mean scores from all nine conditions are plotted versus U_{50} values for each age group in Fig. 11. Because the range of conditions is not large, the variation in speech intelligibility scores with U_{50} values is approximated by linear regression lines in Fig. 11. Smoothed speech intelligibility scores that represent the average trend of the data can be determined from these linear

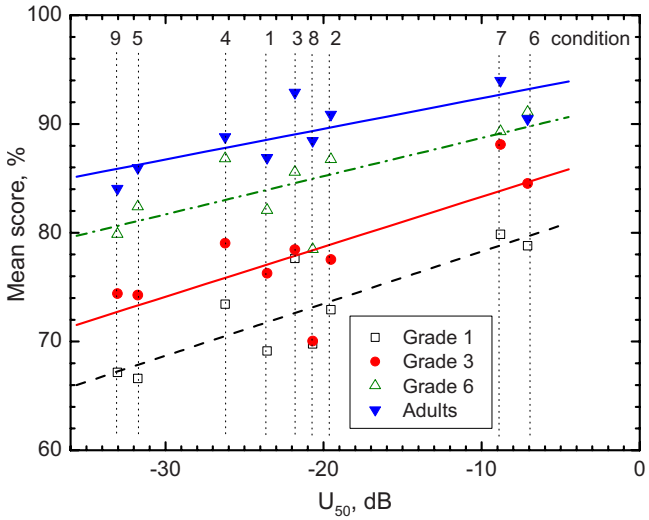


FIG. 11. (Color online) Plot of mean speech intelligibility scores vs U_{50} values for each of the nine conditions and for each age group with associated linear regression lines for each age group. Each vertical dotted line indicates the data for one condition as labeled at the top of the graph.

regression lines. These smoothed scores should provide a more accurate indication of the mean trend of the data for each of the experiments.

The smoothed speech intelligibility scores for the experiment No. 1 conditions from the regression lines in Fig. 11 are plotted versus T_{60} values in Fig. 12. These show what is believed to be better estimates of the mean trends of the experiment No. 1 results. Figure 12 shows approximately parallel curves peaking at a T_{60} of 0.68 s. That is, for these conditions this T_{60} value provides the best speech intelligibility. However, speech intelligibility scores are not substantially lower for a wide range of reverberation times. From Fig. 12 one could conclude that of the test conditions only the 1.2 s reverberation time condition showed a significant reduction in mean speech intelligibility score for the smoothed results of experiment No. 1. When the curves in

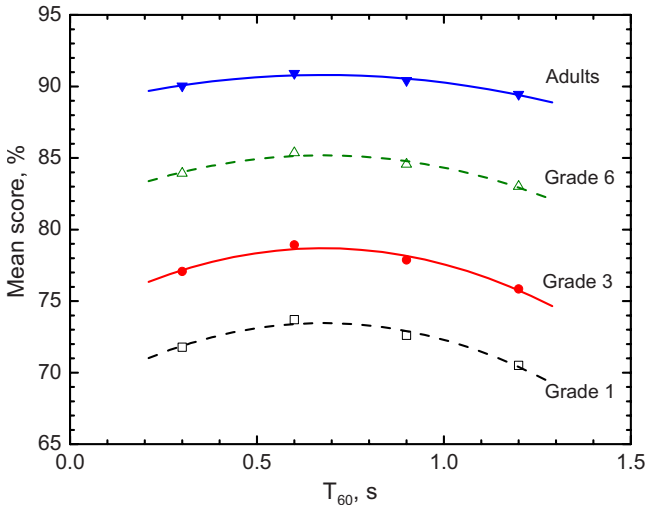


FIG. 12. (Color online) Smoothed speech intelligibility scores plotted vs T_{60} values for the results of experiment No. 1 with conditions having varied S/N values. The curved lines are second order polynomial regression lines to the data.

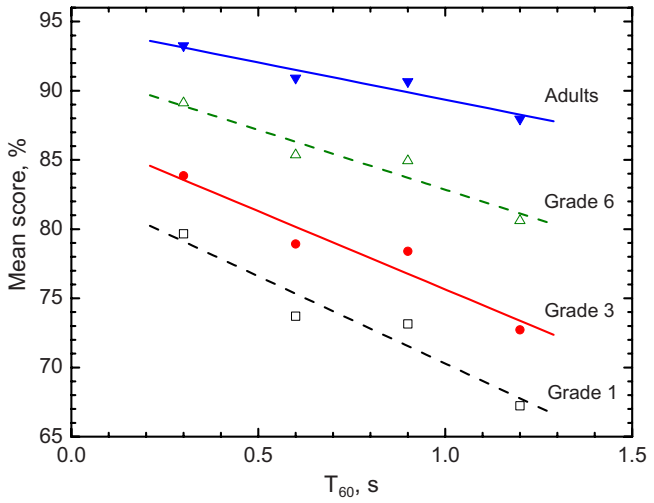


FIG. 13. (Color online) Smoothed speech intelligibility scores plotted vs T_{60} values for the results of experiment No. 2 with conditions having constant S/N value. The lines are linear regression lines to the data for each age group.

Fig. 12 are examined more carefully, they are seen to vary in curvature and are not quite parallel. The curvature increases with decreasing age of the listeners, possibly suggesting that younger listeners are more sensitive to the negative effects of reverberation. However, these effects are too small to be practically important and were not statistically substantiated.

Smoothed values for the experiment No. 2 results were also obtained from Fig. 11 and are plotted versus T_{60} values in Fig. 13. As expected this figure shows speech intelligibility scores increasing with decreasing T_{60}. However, it can now be seen that the rate of variation in speech intelligibility scores with T_{60} is greatest for the youngest listeners. That is, the negative effects of increasing reverberation time more rapidly degrade conditions for the youngest listeners.

V. CONCLUSIONS

The new results in this study provide statistically significant evidence of the effects of reverberation time and the age of the listeners on the intelligibility of speech in elementary school classrooms.

For the conditions of constant noise level and varied S/N in experiment No. 1, speech intelligibility scores were near maximum (within 1%) for a wide range of reverberation times. The new results indicate that for these varied S/N conditions, acceptable reverberation times can be described as the range from about 0.3 to 0.9 s reverberation time. The varied S/N conditions of experiment No. 1 are thought to be most representative of conditions in elementary school classrooms where the dominant sources of interfering sounds are the nearby children.

These results suggest that the natural increase in speech levels with the increased early reflection energy associated with increased reverberation time compensates for the negative effects of the concurrent increase in late-arriving speech sound with increasing reverberation time. However, if the constant noise level used in experiment No. 1 were increased or decreased the range of acceptable reverberation times would change. Previous studies have demonstrated that preferred reverberation times for speech increase with increased noise levels.[26]

For conditions of constant S/N (experiment No. 2), speech intelligibility scores increased with decreasing reverberation times and the effect was most rapid for the youngest listeners. However, even for high S/N conditions, having some reflected sound can be critical in understanding speech and hence very low reverberation times should not be recommended. For example, when the talker's head is turned or when listeners are more distant from the talker, adequate speech intelligibility depends on reflected sound and in such cases early-arriving reflections can increase S/N by 7 dB or more.[8]

The addition of early-arriving reflections of speech sounds was confirmed to be beneficial for young children and for adults.

While the younger children always had lower speech intelligibility scores, this was mostly due to younger children being more adversely affected by interfering noise.[1] However, there were small indications that younger children were more adversely affected by reverberation. For the varied S/N conditions (experiment No. 1), the range of acceptable reverberation times decreased very slightly with decreasing age of the listener. For the constant S/N conditions (experiment No. 2), the decrease in intelligibility scores with increasing reverberation time was a little more rapid for younger listeners. However, the magnitude of the negative effects of reverberation on speech intelligibility was much smaller than previously found for varied S/N and the effects of reverberation varied much less with the age of the listener.

An ideal approach to the acoustical design of classrooms would be to first reduce all noise levels (at the source if possible) and then design the reverberation time of the room to optimize the provision of added reflected sound to enhance speech levels. The current results suggest that design criteria should not specify maximum reverberation times. They should specify a range of acceptable values. Too little reflected sound is a potentially expensive and serious problem.

This study has considered how the physical characteristics of the classrooms affect the intelligibility of speech. The situation in real classrooms is more complicated than in the current tests because often the major factor influencing intelligibility is the interfering sounds made by the children. The levels of sound from the children and their behavior may also be affected by the acoustical treatment of the classroom. Further studies are needed to compare conditions in treated and untreated classrooms to help understand the interactions of the behavior of students and teachers with the acoustical treatment of classrooms.

ACKNOWLEDGMENTS

This work was supported by a grant from the Canadian Language and Literacy Research Network. The authors are grateful to the Ottawa Carleton District School Board and the many school principals and teachers who made it possible

for us to carry out this research. The authors would like to thank Dr. Brad Gover for his help with the processing of the speech recordings.

[1]J. S. Bradley and H. Sato, "The intelligibility of speech in elementary school classrooms," J. Acoust. Soc. Am. **123**, 2078–2086 (2008).
[2]H. Sato and J. S. Bradley, "Evaluation of acoustical conditions for speech communication in working elementary school classrooms," J. Acoust. Soc. Am. **123**, 2064–2077 (2008).
[3]A. K. Nábělek and J. M. Pickett, "Reception of consonants in a classroom as affected by monaural and binaural listening, noise, reverberation and hearing aids," J. Acoust. Soc. Am. **56**, 628–639 (1974).
[4]T. Finitzo-Hieber and T. W. Tillman, "Room acoustics effects on monosyllabic word discrimination ability for normal and hearing-impaired children," J. Speech Hear. Res. **21**, 440–458 (1978).
[5]C. E. Johnson, "Children's phoneme identification in reverberation and noise," J. Speech Lang. Hear. Res. **43**, 144–157 (2000).
[6]A. Neuman and I. Hochberg, "Children's perception of speech in reverberation," J. Acoust. Soc. Am. **73**, 2145–2149 (1983).
[7]J. S. Bradley, "Speech intelligibility studies in classrooms," J. Acoust. Soc. Am. **80**, 846–854 (1986).
[8]J. S. Bradley, H. Sato, and M. Picard, "On the importance of early reflections for speech in rooms," J. Acoust. Soc. Am. **113**, 3233–3244 (2003).
[9]M. Hodgson and E.-M. Nosal, "Effect of noise and occupancy on optimal reverberation times for speech intelligibility in classrooms," J. Acoust. Soc. Am. **111**, 931–938 (2002).
[10]W. Yang and M. Hodgson, "Auralization study of optimum reverberation times for speech intelligibility for normal and hearing-impaired listeners in classrooms with diffuse sound fields," J. Acoust. Soc. Am. **120**, 801–807 (2008).
[11]ISO3382, "Acoustics—Measurement of the reverberation time of rooms with reference to other acoustical parameters," International Organisation for Standardisation, Geneva, Switzerland (1998).
[12]American National Standards Institute (ANSI) Standard S12.60, "Acoustical performance criteria, design requirements, and guidelines for schools" (American National Standards Institute, New York).
[13]L. L. Beranek, *Concert and Opera Halls: How They Sound* (Acoustical Society of America, New York, 1996).
[14]M. Barron and L. J. Lee, "Energy relations in concert auditoriums I," J. Acoust. Soc. Am. **84**, 618–628 (1988).
[15]S. Chiles and M. Barron, "Sound level distribution and scatter in proportionate spaces," J. Acoust. Soc. Am. **226**, 1585–1595 (2004).
[16]N. B. Marshall, "The effects of different signal-to-noise ratios on the speech recognition scores of children," Ph.D. thesis, University of Alabama, Tuscaloosa, AL (1987).
[17]M. Ross and J. Lerman, "A picture identification test for hearing impaired children," J. Speech Hear. Res. **13**, 44–53 (1970).
[18]A. Kulkarni and H. S. Colburn, "Variability in the characterization of the headphone transfer-function," J. Acoust. Soc. Am. **102**, 1071–1074 (2000).
[19]ANSI S3.5-1997, "Methods for calculation of the speech intelligibility Index," American National Standard, Standards Secretariat, Acoustical Society of America, New York.
[20]W. Yang and M. Hodgson, "Validation of the auralization technique: Comparative speech-intelligibility tests in real and virtual classrooms," Acta. Acust. Acust. **93**, 991–999 (2007).
[21]D. F. Hoth, "Room noise spectra at subscribers' telephone locations," J. Acoust. Soc. Am. **12**, 449–504 (1941).
[22]W. E. Blazier, "Revised noise criteria for application in the acoustical design and rating of HVAC systems," Noise Control Eng. **16**, 64–73 (1981).
[23]J. S. Bradley, H. Sato, B. N. Gover, and N. York, "Comparison of speech intelligibility scores for direct listening and headphone playback," J. Acoust. Soc. Am. **117**, 2465 (2005).
[24]J. S. Bradley, "Predictors of speech intelligibility in Rooms," J. Acoust. Soc. Am. **80**, 837–845 (1986).
[25]J. S. Bradley, R. D. Reich, and S. G. Norcross, "On the combined effects of signal-to-noise ratio and room acoustics on speech intelligibility," J. Acoust. Soc. Am. **106**, 1820–1828 (1999).
[26]R. Reich and J. S. Bradley, "Optimizing classroom acoustics using computer model studies," Can. Acoust. **26**, 15–21 (1998).

Optimal design of minimum mean-square error noise reduction algorithms using the simulated annealing technique

Mingsian R. Bai,[a] Ping-Ju Hsieh, and Kur-Nan Hur
Department of Mechanical Engineering, National Chiao-Tung University, 1001 Ta-Hsueh Road, Hsin-Chu 300, Taiwan

(Received 8 August 2008; revised 20 November 2008; accepted 21 November 2008)

The performance of the minimum mean-square error noise reduction (MMSE-NR) algorithm in conjunction with time-recursive averaging (TRA) for noise estimation is found to be very sensitive to the choice of two recursion parameters. To address this problem in a more systematic manner, this paper proposes an optimization method to efficiently search the optimal parameters of the MMSE-TRA-NR algorithms. The objective function is based on a regression model, whereas the optimization process is carried out with the simulated annealing algorithm that is well suited for problems with many local optima. Another NR algorithm proposed in the paper employs linear prediction coding as a preprocessor for extracting the correlated portion of human speech. Objective and subjective tests were undertaken to compare the optimized MMSE-TRA-NR algorithm with several conventional NR algorithms. The results of subjective tests were processed by using analysis of variance to justify the statistic significance. A *post hoc* test, Tukey's Honestly Significant Difference, was conducted to further assess the pairwise difference between the NR algorithms.
 [DOI: 10.1121/1.3050292]

PACS number(s): 43.60.Dh, 43.60.Np, 43.60.Uv, 43.72.Kb [EJS] Pages: 934–943

I. INTRODUCTION

In recent years, applications of mobile communication, video conferencing, and peer-to-peer internet telephony networks, such as SKYPE®, hands-free car kits, etc., are rapidly advancing in modern daily life. In these applications, effective communication in noisy environments has been one of the pressing problems. To enhance speech quality, noise reduction (NR) technology has been extensively studied in the communication community. The main problem with most NR algorithms is that sheer NR does not necessarily lead to the general preference of the users. Overly aggressive NR schemes often result in processing artifacts and degradation of speech quality. How to effectively reduce background noise without impairing speech quality has become an imminent issue for NR algorithm design.

NR algorithms fall into three categories: spectral-subtraction algorithms, statistical-model-based algorithms, and subspace algorithms. Spectral-subtraction algorithms[1–6] subtract directly the estimated noise spectrum from the spectrum of the noisy speech. Statistical-model-based algorithms estimate Fourier coefficients using statistically optimal linear or nonlinear estimators of clean signals. The Wiener algorithm[7–10] and the minimum mean-square error (MMSE)[1,11] algorithm belong to this class. Subspace algorithms are based on the principle that the vector space of the noisy signal can be decomposed into the "signal" and "noise" subspaces. Noise is suppressed by projecting the noisy signals onto the signal subspace and nullifying the components in the noise subspace. The decomposition of these two orthogonal subspaces can be done by using the singular value decomposition or the eigenvalue decomposition. The Karhunen–Loéve transform (KLT) algorithm[11,12] falls into this category. All NR algorithms require the information of noise spectra or noise covariance matrices, which must be estimated and updated from frame to frame. Noise estimation can be carried out either during speech pauses, which requires a voice activity detector (VAD), or continuously using time-recursive averaging (TRA) algorithms. A more comprehensive review of speech enhancement and NR methods can be found in the monograph by Loizou.[11]

In this paper, a MMSE-NR algorithm based on TRA[11,13] noise estimation (denoted as MMSE-TRA-NR) is investigated. This algorithm is found to be very sensitive to the choice of two recursion parameters. To address this problem in a more systematic manner, this paper proposes an optimization method to efficiently search the optimal parameters of the MMSE-TRA-NR algorithms. A global optimization technique, simulated annealing (SA)[14–16] algorithm, is exploited for locating the optimal parameters. The objective function is a combined objective measure for NR and the incurred distortion of processed signals. Sensitivity analysis of the TRA parameters obtained using the SA optimization was undertaken for nine types of background noise. In addition to the optimized MMSE-TRA-NR, the possibility of using linear prediction coding (LPC)[6,17–19] as a preprocessor to the NR algorithm is also explored.

In order to evaluate the proposed optimized algorithm and the other NR algorithms, objective and subjective tests were carried out. The objective tests were conducted according to ITU-T P.862.[20] The subjective listening tests were conducted according to ITU-T P.835.[21] The test data were processed by using analysis of variance (ANOVA) to justify the statistic significance of the difference among the NR algorithms. A *post hoc* test, Tukey's HSD, was also employed in

[a]Author to whom correspondence should be addressed. Electronic mail: msbai@mail.nctu.edu.tw

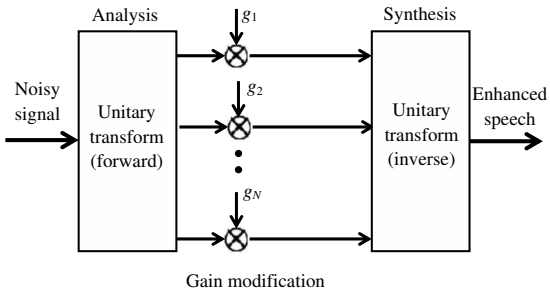


FIG. 1. General structure of NR algorithms (adapted from Ref. 11).

the paired comparison between the NR algorithms.

II. NOISE REDUCTION ALGORITHMS

Figure 1 illustrates the general three-step structure of NR algorithms.[11] The noisy signal is forward transformed using unitary transformations, e.g., Fourier transform, discrete cosine transform, and KLT transform. Next, gain modification, the major NR operation, takes place in the transformed domain. Finally, the time-domain signal of the enhanced speech is recovered by an overlap-and-add procedure. In this section, the MMSE-NR algorithm will be reviewed. The other traditional NR algorithms, such as the spectral subtraction, the Wiener filtering, and the KLT, to be compared in this paper are only mentioned in Sec. I with references.

A. Statistical-model-based noise reduction algorithm

The MMSE-NR algorithm is also based on a statistical model. Instead of the complex spectrum as in the Wiener filter method, a nonlinear estimator of the magnitude spectrum is optimized in the MMSE-NR algorithm. It is assumed that the discrete Fourier transform (DFT) coefficients are statistically independent and follow the Gaussian distribution. The mean-square error between the estimated ($\hat{S}_k$) and the true (S_k) magnitudes of the clean speech signal is

$$E_{\text{mse}} = E\{(\hat{S}_k - S_k)^2\}. \tag{1}$$

This expectation can be estimated using the following Bayesian mean-square error approach:

$$B_{\text{mse}}(\hat{S}_k) = \int\int (S_k - \hat{S}_k)^2 p(\mathbf{Y}, S_k) d\mathbf{Y} dS_k, \tag{2}$$

where $\mathbf{Y}=[Y(\omega_0)Y(\omega_1)\cdots Y(\omega_{N-1})]$ is the noisy speech spectrum and $p(\mathbf{Y},S_k)$ is the joint probability density function (pdf). The posterior pdf of S_k can be determined by using Bayes' rule. Minimization of the Bayesian MSE with respect to $\hat{S}_k$ leads to the optimal MMSE estimator,

$$\hat{S}_k = E[S_k|Y(\omega_k)] = \int_0^\infty s_k p(s_k|Y(\omega_k)) ds_k$$

$$= \frac{\int_0^\infty s_k p(Y(\omega_k)|s_k) p(s_k) ds_k}{\int_0^\infty p(Y(\omega_k)|s_k) p(s_k) ds_k}, \tag{3}$$

where s_k is a realization of the random variable S_k and $p(s_k|Y(\omega_k))$ is the conditional posterior pdf of s_k under the observation $Y(\omega_k)$. Assuming that the pdf of the noise Fourier coefficients is Gaussian, it was shown by Ephraim and Malah that the statistically optimal MMSE magnitude estimator takes the form[1]

$$\hat{S}_k = \frac{\sqrt{\pi}}{2}\frac{\sqrt{v_k}}{\gamma_k}\exp\left(-\frac{v_k}{2}\right)\left[(1+v_k)I_0\left(\frac{v_k}{2}\right) + v_k I_1\left(\frac{v_k}{2}\right)\right] Y_k, \tag{4}$$

where $I_0(\cdot)$ and $I_1(\cdot)$ are the modified Bessel functions of the zero and the first order, respectively, Y_k is the spectral magnitude of the noisy signal, and v_k is defined by

$$v_k = \frac{\xi_k}{1+\xi_k}\gamma_k, \tag{5}$$

where γ_k denotes the *a posteriori* signal-to-noise ratio (SNR) given by

$$\gamma_k \triangleq \frac{Y_k^2}{P_{vv}(\omega_k)} = \frac{Y_k^2}{E\{|V(\omega_k)^2|\}}. \tag{6}$$

In practice, the noise variance and hence the *a priori* SNR ξ_k are unknown, given the noisy signal $y(n)$. Thus, noise spectrum must be estimated prior to NR processing. First, the noise variance is estimated during speech pauses with the aid of a VAD (Ref. 22) provided the noise is stationary. For example, the following statistical-model-based VAD can be used:

$$\frac{1}{N}\sum_{k=1}^{N-1} \log\left(\frac{1}{1+\xi_k}\exp\left(\frac{\gamma_k\xi_k}{1+\xi_k}\right)\right) \underset{H_0}{\overset{H_1}{\gtrless}} \Delta, \tag{7}$$

where N is the Fast Fourier transform size, H_0 and H_1 denote the hypotheses of speech absence and speech presence, respectively, and the threshold Δ is usually set to 0.15. Here, the MMSE-NR algorithm used in conjunction with VAD for noise estimation is denoted as "MMSE-VAD-NR." Next, the *a priori* SNR ξ_k is estimated with a "decision-directed" approach using the recursive formula

$$\hat{\xi}_k(m) = a\frac{\hat{S}_k^2(m-1)}{P_{vv}(\omega_k, m-1)} + (1-a)\max(\gamma_k(m)-1, 0), \tag{8}$$

where m is the frame number and $0<a<1$ is a weighting factor commonly chosen to be $a=0.98$.

As mentioned above, Eq. (4) is only a spectral magnitude estimator. To recover the enhanced signal, one needs to estimate the phase of the clean speech signal. It was shown by Ephraim and Malah[1] that the optimal phase estimate is simply the noisy phase. Thus, the enhanced complex signal spectrum is calculated by combing the preceding estimated magnitude spectrum $\hat{S}_k$ and the noisy signal phase spectrum $j\theta_y(k)$, i.e., $\hat{S}(\omega_k)=\hat{S}_k\exp(j\theta_y(k))$.

III. ENHANCMENT OF MMSE-NR ALGORITHMS

In this section, three approaches of technical refinement are exploited to enhance the aforementioned MMSE-NR algorithm.

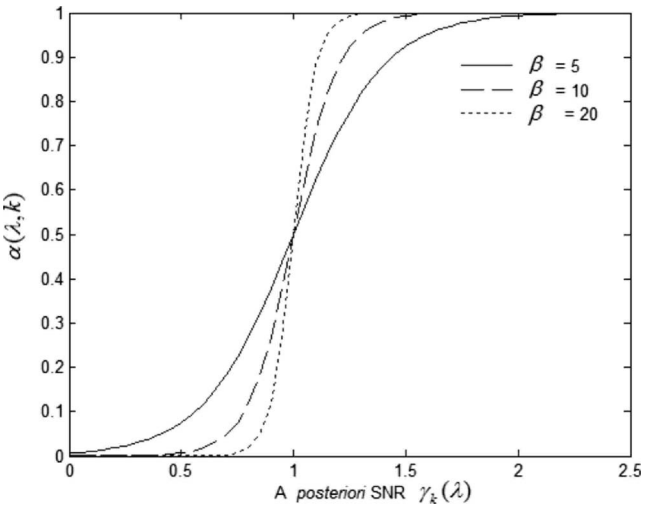


FIG. 2. The smoothing factor $\alpha(\lambda,k)$ calculated according to Eq. (10)for different values of the parameter β with $\delta=1$. (Solid line: $\beta=5$; dashed line: $\beta=10$; dotted line: $\beta=20$).

A. MMSE-time recursive averaging noise reduction

As mentioned earlier in the MMSE-VAD-NR algorithm, the noise variance can be estimated and updated during speech pauses via a VAD provided the noise is stationary. In practice, however, many background noises are often transient and nonstationary. For background noise of this kind, a more practical noise estimation algorithm called the TRA algorithm[13] can be used.

In the TRA algorithm, noise variance $\hat{\sigma}_v^2(\lambda,k)$ at the frame λ and the frequency k is estimated with the following recursive formula:

$$\hat{\sigma}_v^2(\lambda,k)=\alpha(\lambda,k)\hat{\sigma}_v^2(\lambda-1,k)+(1-\alpha(\lambda,k))|Y(\lambda,k)|^2, \tag{9}$$

where $|Y(\lambda,k)|$ is the noisy speech magnitude spectrum and $\alpha(\lambda,k)$ is a time and frequency dependent smoothing factor. The smoothing factor α in the one-pole recursive formula was used to avoid the excessive fluctuations during the process of noise estimation. Various algorithms were proposed to determine the smoothing factor $\alpha(\lambda,k)$ on the basis of the estimated SNR or the probability of speech presence. In this paper, a SNR-based smoothing factor $\alpha(\lambda,k)$ is selected to follow a sigmoid function,

$$\alpha(\lambda,k)=\frac{1}{1+e^{-\beta[\gamma_k(\lambda)-\delta]}}, \tag{10}$$

where β and δ are constants and the *a posteriori* SNR $\gamma_k(\lambda)$ is calculated by averaging the estimated noise variance in the past ten frames,

$$\gamma_k(\lambda)=\frac{|Y(\lambda,k)|^2}{\frac{1}{10}\sum_{m=1}^{10}\hat{\sigma}_v^2(\lambda-m,k)}. \tag{11}$$

Figure 2 plots the smoothing factor α for different values of the parameter β and $\delta=1$. Equations (10) and (11) can be interpreted as follows. If the speech is present, the *a posteriori* SNR $\gamma_k(\lambda)$ will be large, and therefore $\alpha(\lambda,k)\approx 1$. In this case, the noise update will cease and the noise estimate will remain the same as that of the previous frame [the first term of Eq. (9)]. Conversely, if the speech is absent, the *a posteriori* SNR $\gamma_k(\lambda)$ will be small, and therefore $\alpha(\lambda,k)\approx 0$. That is, the noise estimate will follow the power spectral density of the noisy spectrum [the second term of Eq. (10)]. In a long stationary noise period, α would stay at a very small value. As a consequence, $\hat{\sigma}_v^2(\lambda,k)\approx|Y(\lambda,k)|^2$. This ensures an accurate and robust estimation of noise level, which gives rise to good reduction performance. Thus, α is strongly dependent on the *a posteriori* SNR $\gamma_k(\lambda)$. The choice of parameters β and δ dictates the slope and the location of the transition of the sigmoid function. This transition can be considered as a "soft switch" between the bistates of speech presence and absence. How to select these two parameters to maximize the NR performance is crucial to the resulting NR performance, as will be explored in the subsequent sections.

When noise is strong and the SNR becomes rather low, the distinction of speech and noise segments could be difficult. Moreover, the noise is estimated intermittently and updated only during the speech silent periods. This may cause problems if the noise is nonstationary, which is the case in many applications. The recursive nature of the TRA algorithm enables estimating noise variance continuously, even during speech activities, which is advantageous in dealing with nonstationary noises. Figure 3 compares NR performance between the VAD and the TRA algorithms. The test signal is a speech signal corrupted by random noise (solid line) varied with three different levels (low-high-medium). The noise (dotted line) estimated using the VAD and the TRA algorithms are also superimposed in the left side of the top and the bottom panels in Fig. 3, respectively. Unlike the VAD algorithm that fails to respond to the noise level variation, the TRA is capable of estimating the noise with drastically transient fluctuation. In other words, VAD and TRA deal with different noise scenarios. VAD is suited for the estimation of stationary noise during speech absence, while TRA is preferred for estimating transient noise, where synchronization of noise estimation is crucial. As a result, a marked difference in NR performance is observed in the enhanced signals using these two noise estimation methods. The right side of the top and the bottom panels in Fig. 3 shows the signals (dotted lines) processed by the MMSE-NR using VAD and TRA, respectively, for noise estimation. The noisy signals (solid lines) are also superimposed to ease comparison. Obviously, the TRA is more superior to the VAD in estimating nonstationary background noise. Thus, the MMSE with TRA noise estimation (denoted as MMSE-TRA-NR) will be employed in the following presentation.

B. Intelligent tuning of the parameters for the MMSE-TRA-NR algorithm

As mentioned previously, the parameters β and δ are used in the sigmoid function of the TRA algorithm for noise estimation. Conventionally, choices such as $\delta=1.5$, $15\leqslant\beta\leqslant 30$ are recommended in the literature.[11] To our surprise, we found that these two parameters β and δ have profound impact on noise estimation and hence on the NR performance

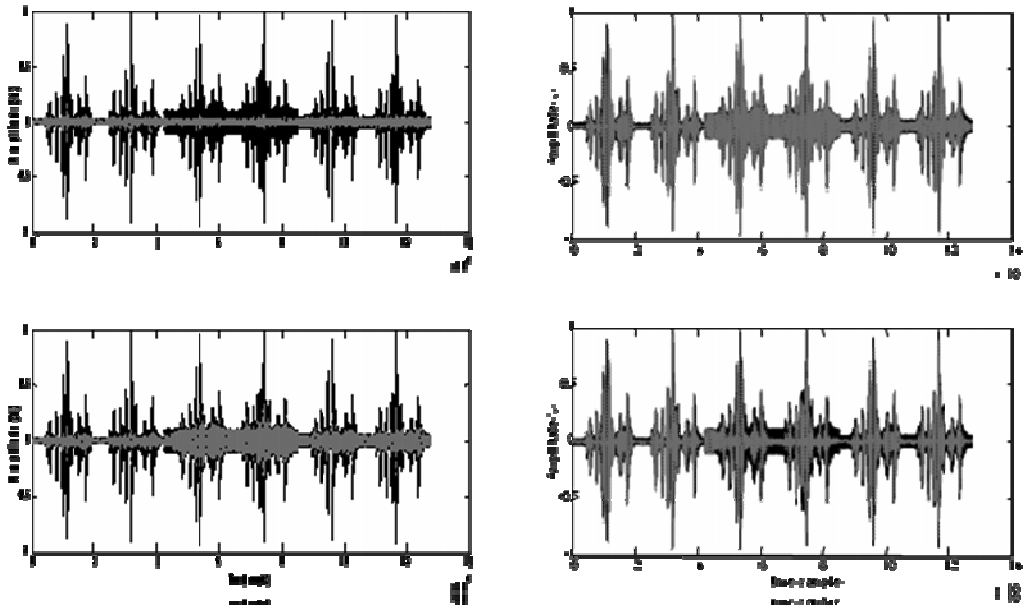


FIG. 3. Comparison of the VAD and TRA algorithms. The noise estimated using the VAD and the TRA algorithms are superimposed in the left side of the top and the bottom panels. The processed speech signals using the MMSE-VAD-NR and MMSE-TRA-NR algorithms are superimposed in the right side of the top and the bottom panels.

of the MMSE-TRA-NR algorithm. Therefore, it is worth exploring how to adjust these two parameters such that NR performance can be maximized without too much speech quality degradation. In the following, a procedure based on the SA optimization method is presented for automated tuning of the TRA parameters.

1. Simulated annealing algorithm

SA is a generic probabilistic meta-algorithm for the global optimization problem, namely, locating a good approximation to the global optimum of a given function in a large search space.[14–16] SA is a technique well suited for solving global optimization problems with many local optima. The flowchart of the SA is illustrated in Fig. 4. In the SA method, each point in the search space is analogous to the thermal state of the annealing process in metallurgy. The objective function Q to be maximized is analogous to the internal energy of the system in that state. The goal of search is to bring

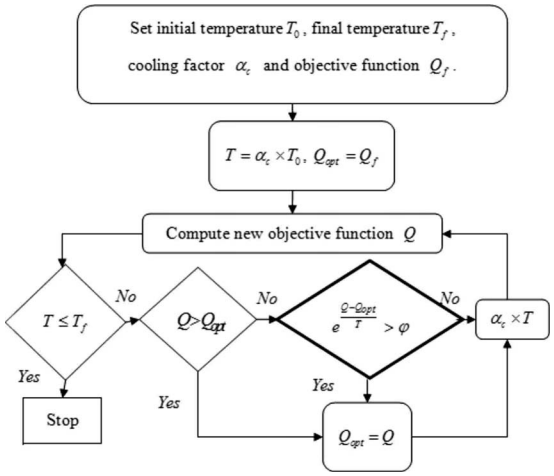


FIG. 4. The flowchart of the SA optimization algorithm.

the system from an initial state to a randomly generated state with the maximum possible objective function. Two conditions are used to determine whether or not to accept an improved solution. If the objective function is increased, the new state is always accepted. Conversely, if the objective function is decreased and the following condition holds, the new state is accepted:

$$p_{\mathrm{SA}} = \exp(\Delta Q/T) > \varphi, \tag{12}$$

where p_{SA} is the acceptance probability function, ΔQ denotes the increment of the objective function, T is the temperature that follows a certain annealing schedule, and φ is a random number generated subject to the uniform distribution on the interval [0, 1]. It follows that the system may possibly move to a new state that is "worse" than the present one. It is this mechanism that prevents the search from being trapped in a local maximum.

Initially, the high temperature T results in the high probability of accepting a move that decreases the objective function, which is analogous to a steel piece whose thermal state is highly active at high temperatures. As the annealing process goes on and T decreases, the probability of accepting a move becomes increasingly small until it finally converges to a stable solution.

A simple annealing schedule is the exponential cooling, which begins at some initial temperature T_0 and decreases temperature in steps according to

$$T_{k+1} = \alpha_c T_k, \tag{13}$$

where $0 < \alpha_c < 1$ is a cooling factor. It is likely that a number of moves are accepted at each temperature before proceeding to the new state. SA search is terminated at some final value T_f. An empirical choice for α_c is 0.95, and T_0 should be chosen such that the initial acceptance probability is higher than 0.8.

2. Objective function Q

Two objective measures, the segmental SNR (denoted as SNRseg) and the perceptual evaluation of sound quality (PESQ),[21] are considered in constructing the objective function for optimizing the performance in the MMSE-TRA-NR algorithm. The index SNRseg calculates SNR based on the noisy signals and the processed signals averaged over frames,

$$\mathrm{SNRseg} = \frac{10}{M_s}\sum_{m=0}^{M_s-1} \log_{10}\frac{\sum_{n=N_s m}^{N_s m+N_s-1} s^2(n)}{(s(n)-\hat{s}(n))^2}, \tag{14}$$

where N_s is the frame length and M_s is the number of frames. The SNRseg is a widely used objective measure for assessing NR performance in the telephony industry. The index PESQ is a more sophisticated objective measure for assessing speech quality, which takes into account psychoacoustic aspects of human hearing. The original and the processed signals are first level—equalized to a standard listening level and filtered by a filter, with a response similar to a standard telephone handset. The signals are aligned in time to correct for time delays and then processed through an auditory transform to obtain the loudness spectra. A more detailed information of the PESQ can be found in ITU-T P. 862.[21]

SNRseg and PESQ reflect the NR performance and the sound quality, respectively, of the processed signals. Hence, an objective function Q is constructed by combining the SNRseg and the PESQ using a weighting factor r, i.e.,

$$Q = r \times \mathrm{SNRseg} + \mathrm{PESQ}. \tag{15}$$

The weighting factor r will be found from a subjective listening test. Two kinds of background noise at the SNR level of 5 dB, white noise and car noise, were processed using five NR algorithms including spectral subtraction, Wiener filtering, MMSE-VAD-NR, MMSE-TRA-NR, and KLT-NR. The TRA parameters in MMSE-TRA-NR are chosen to be $\beta=0.6$ and $\delta=1.5$. Figure 6 shows the clean speech signal used in the simulation. The test signal is a male speech sentence sampled at 8 kHz and separated into 25 ms frames with 50% overlap. The test signals last for 2 s in duration. All test signals were adjusted to the same level of loudness. A headset was used as the means of audio rendering.

Owing to the space limitation, we show only the results processed using the MMSE-TRA-NR algorithm. Figures 5(a) and 5(b) show the spectrograms of the noise and the signal processed by the MMSE-TRA-NR algorithm for the white noise case. Figures 6(a) and 6(b) show the spectrograms of the noise and the signal processed by the MMSE-TRA-NR algorithm for the car noise case. Thirty-two experienced listeners participated in the listening test. Three subjective indices including NR, *sound quality*, and *total preference* were employed in the listening test. The grading scale is set to be −3 to 3. A multiple regression analysis based on five NR algorithms and two background noises was utilized to establish a linear model between the NR, sound quality, and total preference. The results of the multiple regression analysis determine the weighting factors between the SNRseg and the PESQ for the objective function. This gives the weighting

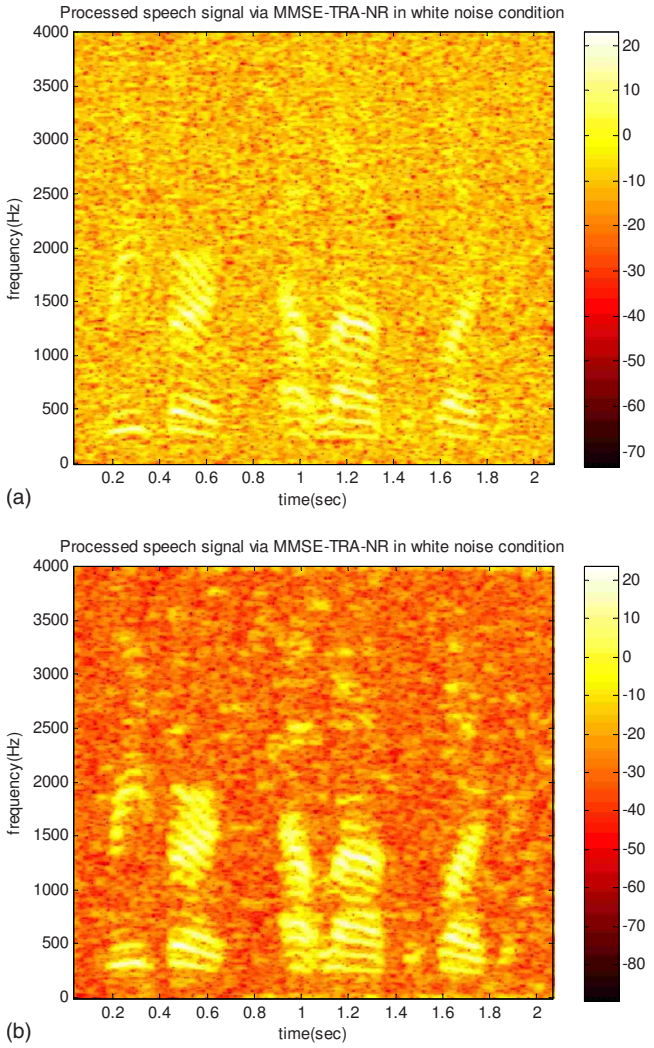


FIG. 5. (Color online) The spectrograms of a male speech sentence in the white noise scenario. (a) Speech corrupted by the white noise. (b) Enhanced speech signal processed by the MMSE-TRA-NR algorithm.

factor in Eq. (27), $r=1.867$, which will be used in the objective function in optimizing the MMSE-TRA-NR algorithm using the SA method next.

3. SA optimization of the MMSE-TRA-NR algorithm

The objective function with $r=1.867$ is employed in the SA optimization of the MMSE-TRA-NR algorithm. Initially, the TRA parameters are arbitrarily chosen to be $\beta=1.6$ and $\delta=1$. The parameters of SA are chosen as $T_0=1$ K, $T_f=10^{-9}$ K, and $\alpha_c=0.95$. With the SA optimization, the optimal parameters are obtained for the white noise ($\beta=0.6117$ and $\delta=0.5214$) and the car noise ($\beta=0.7128$ and $\delta=0.5265$). Figure 7 shows the "learning curve" of the SA for the car noise scenario, where the objective function Q settles to a constant value after about 500 iterations. To see the effect of optimization, NR performances in terms of the SNRseg and PESQ attained using the initial and the optimal parameters β and δ are compared in Table I. In comparison with the initial nonoptimal setting, a marked improvement in performance is obtained using the optimal TRA parameters.

To further justify the optimized NR algorithm, a subjective listening test was conducted. The test speech signal and

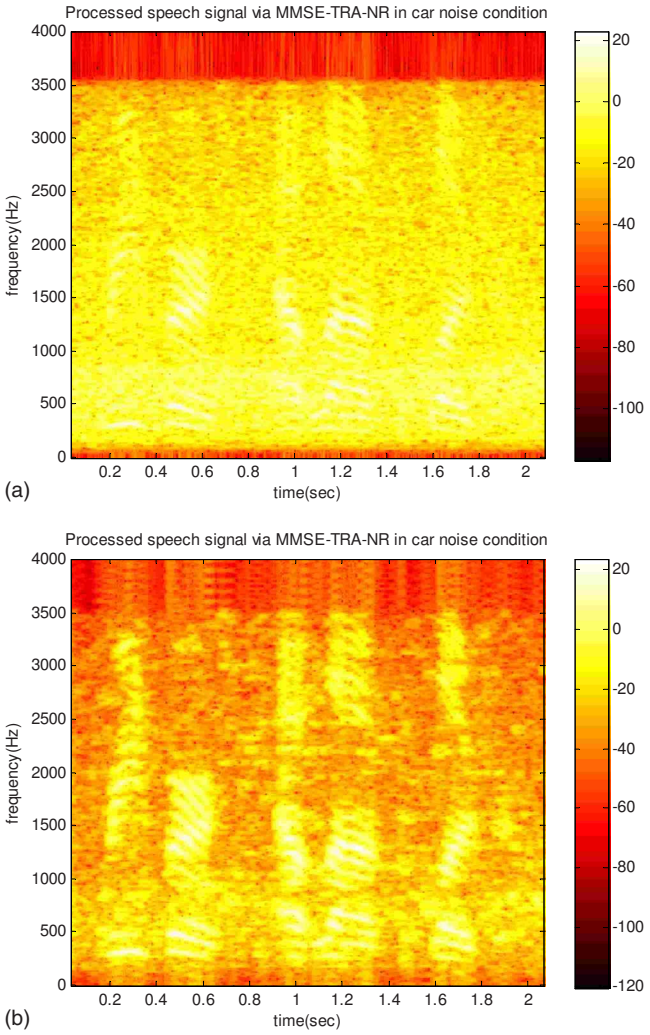


FIG. 6. (Color online) The spectrograms of a male speech sentence in the car noise scenario. (a) Speech corrupted by the car noise. (b) Enhanced speech signal processed by the MMSE-TRA-NR algorithm.

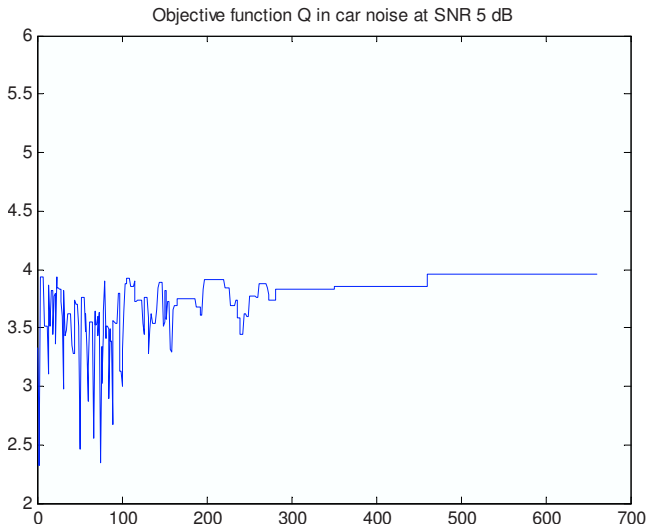


FIG. 7. (Color online) The learning curve of the SA optimization algorithm applied to the car noise.

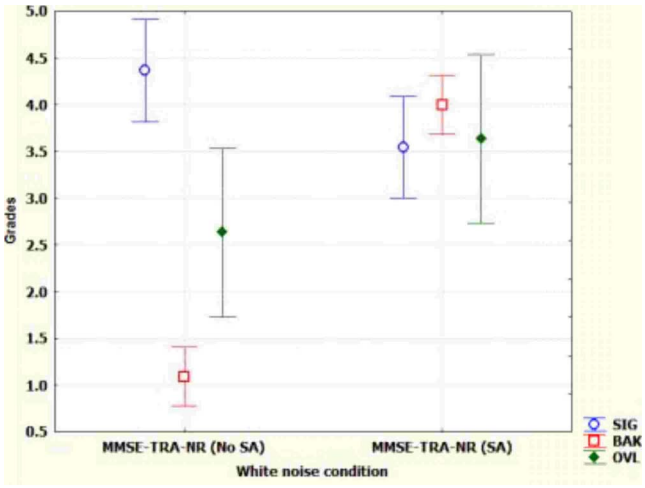
the test conditions are the same as those used in the listening test for the preceding regression analysis. The grading scale is set to be 1–5, as recommended by ITU-T P.835.[22] Three subjective indices, including *scale of signal distortion (SIG)*, *scale of background intrusiveness (BAK)*, and *scale of overall quality (OVL)*, were employed in the listening test. Every subject participating in the test was instructed with the definitions of the subjective indices prior to the listening test. Figures 8(a) and 8(b) show the results of the listening test for the white noise and car noise, respectively. The grades were also processed by using the Multivariate Analysis of Variance (MANOVA) (Ref. 23) to justify the statistical significance of the test results. The average—a 5%–95% bracket is shown in the figure—and the significance level of the grades were summarized in Table II. Cases with significance levels below 0.05 indicate that a statistically significant difference exists among methods. Although there is no significant difference in OVL, the difference in SIG and BAK between the initial and optimal results is significant. The trade-off between NR (BAK) and signal distortion (SIG) is clearly visible—the optimized algorithm has attained remarkable NR performance at some expense of speech quality. Thus, we choose the optimized MMSE-TRA-NR algorithm for the following objective and subjective comparison with several other NR algorithms.

C. Linear prediction coding preprocessor

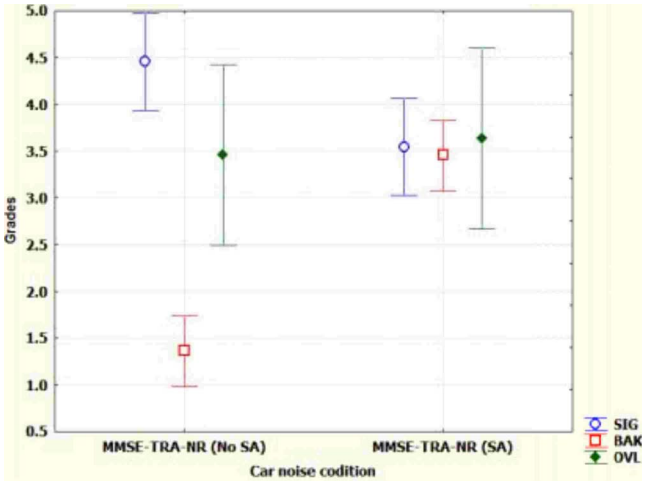
Another possibility of enhancing NR algorithms is to use LPC as the preprocessor. The underlying idea is that the highly correlated portion of human speech can be extracted by using the LPC approach. The timbral quality of voice is preserved as the spectral envelope is captured using the LPC. Figure 10(a) illustrates the one-step forward linear prediction problem.[17–19] The current input $x(n)$ is predicted by a linear combination of past input samples,

TABLE I. The NR performance of the MMSE-TRA-NR algorithm in terms of the objective measures SNRseg and PESQ for the initial and the optimized parameters β and δ (the optimal parameters are marked with *).

	MMSE-TRA-NR parameters				
Noise type	β	δ	SNRseg	PESQ	Q
White noise	1.6	1	−1.0942	1.9639	−0.1984
	0.6117*	0.5214*	1.5155	2.1619	4.8106
Car noise	1.6	1	−1.5609	2.2168	−0.2998
	0.7128*	0.5265*	0.7061	2.3145	3.9544



(a)



(b)

FIG. 8. (Color online) Comparison of the MMSE-TRA-NR algorithm with and without SA optimization. The results of the listening test are processed by using the MANOVA. (a) White noise. (b) Car noise.

$$\hat{x}(n) = \sum_{k=1}^{p} A_k x(n-k), \tag{16}$$

where p is the prediction order and A_k are the prediction coefficients. The associated prediction finite impulse response (FIR) filter is

$$P(z) = \sum_{k=1}^{p} A_k z^{-k}. \tag{17}$$

By minimizing the mean squares of the one-step forward prediction error, $E_p = E\{e^2(n)\} = E\{[x(n) - \hat{x}(n)]^2\}$, the following equation for the linear prediction problem can be derived:

TABLE II. The MANOVA output of the subjective listening test to compare the MMSE-TRA-NR algorithm with and without optimization. The background noises are the white noise and the car noise. Cases with significance value p below 0.05 indicate that statistically significant difference exists among all methods.

	Significance value		
Noise type	SIG	BAK	OVL
White noise	0.040	0.000	0.117
Car noise	0.017	0.000	0.784

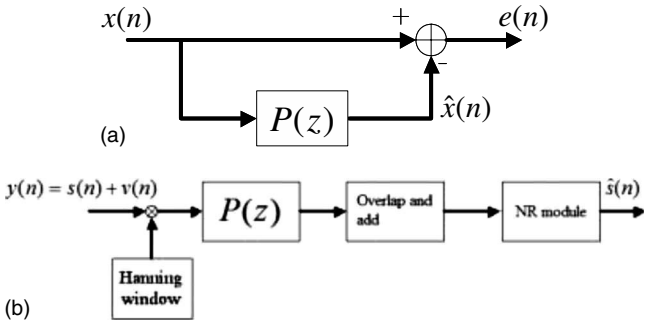


FIG. 9. The NR algorithm cascaded with a LPC preprocessor. (a) Feedforward linear prediction structure. (b) The cascaded LPC-NR system.

$$\sum_{k=0}^{p} A_k \gamma_{xx}(l-k) = \begin{cases} E_p^f, & l = 0 \\ 0, & l = 1,2,\ldots,p, \end{cases} \tag{18}$$

where E_p^f is the mean of the forward prediction error of order p and

$$\gamma_{xx}(m) = E\{x^*(n)x(n+m)\} = \lim_{N\to\infty} \frac{1}{2N+1} \sum_{n=-N}^{N} x^*(n)x(n+m) \tag{19}$$

is the autocorrelation sequence. The optimal LPC coefficients of the prediction filter can be efficiently calculated by using the Levinson–Durbin algorithm. According to the LPC coefficients, The noisy input can be preprocessed by using the prediction filter $P(z)$ in Eq. (17) to extract the correlated input with minimal timbral distortion for the MMSE-TRA-NR module. Figure 9(b) illustrates a MMSE algorithm concatenated with the LPC as its preprocessor (denoted as LPC-MMSE-TRA-NR).

IV. OBJECTIVE AND SUBJECTIVE EVALUATIONS OF NR ALGORITMS

Objective and subjective experiments were undertaken to compare the proposed optimized LPC-MMSE-TRA-NR algorithm with a number of other widely used NR algorithms.

A. Performance evaluation of NR algorithms by objective measures

The preceding objective measures SNRseg and the PESQ are employed to assess the performance of six NR algorithms (spectral subtraction, Wiener filtering, MMSE-VAD-NR, MMSE-TRA-NR, LPC-MMSE-TRA-NR, and KLT-NR algorithms) for the speech signal corrupted by two kinds of background noise (white noise and car noise). All test signals and conditions are similar to those used in the previous test.

According to Table III, the Wiener filtering algorithm tends to underestimate noise level and yield high residual noise (or low SNRseg). The KLT-NR algorithm attains the highest SNRseg. In addition, LPC seems to slightly improve the speech quality over the MMSE-TRA-NR algorithm for

TABLE III. Comparison of processing time and objective NR measures for six NR algorithms.

Noise type	SNRseg		PESQ	
	Noise type			
NR algorithm	White	Car	White	Car
Spectral subtraction	2.115	1.450	2.224	2.118
Wiener filtering	0.878	0.073	2.162	2.322
MMSE-VAD-NR	2.215	1.224	2.250	2.394
MMSE-TRA-NR	1.515	0.7061	2.161	2.314
LPC-MMSE-TRA-NR	1.439	0.3110	2.234	2.162
KLT-NR	3.177	1.856	2.400	2.367

the white noise case. As for the PESQ objective evaluation, there seems to be no significant difference in speech quality resulting from these NR algorithms.

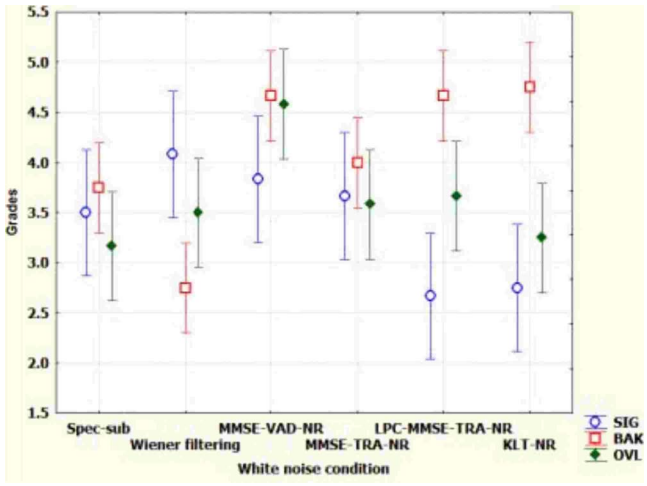
B. Performance evaluation of NR algorithms by subjective measures

In order to further compare the preceding NR algorithms, subjective listening tests were conducted according to the ITU-T P.835.[22] Thirty-two experienced listeners participated in the subjective tests. The six NR algorithms used in the objective test are compared again in this subjective test. The test signals and conditions remain the same as the preceding listening tests (Table IV). The mean and spread of the listening test results are shown in Figs. 10(a) and 10(b). The test results were processed using MANOVA (Ref. 23) with significance levels summarized in Table V. Cases with significance levels below 0.05 indicate that a statistically significant difference exists among methods. From Table V, the difference of the indices SIG, BAK, and OVL among the NR methods was found to be statistically significant (except for OVL in the car noise scenario).

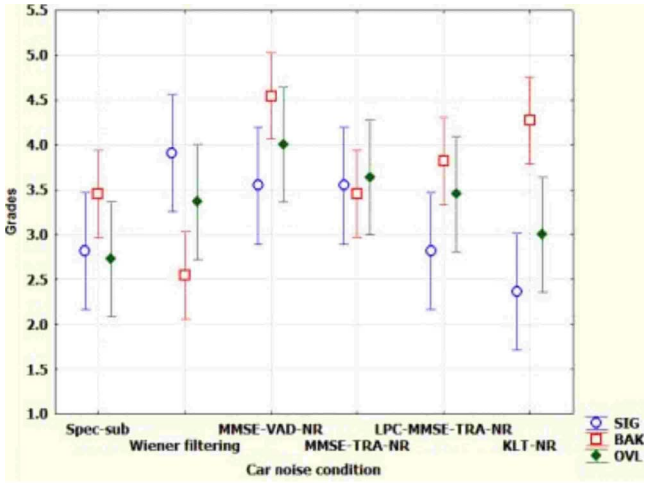
Next, a *post hoc* Tukey HSD test[23] was employed to perform multiple paired comparisons of the NR algorithms. *Post hoc* tests are generally performed after ANOVA, which is able to determine whether or not significant difference is present in the data of a number of cases. Tukey's HSD test is one of the commonly used *post hoc* tests for the assessment of differences in the means between pairs of populations following the ANOVA test. Table VI summarizes the results of

TABLE IV. The optimal parameters β and δ obtained using the SA search for nine types of background noise (babble, station, car, airport, street, train, exhibition, restaurant, and white noise).

Background noise	Optimal β	Optimal δ
White noise	0.6117	0.5214
Babble	0.7178	0.8710
Station	0.6889	0.5350
Car	0.7128	0.5265
Airport	0.6259	0.5016
Street	0.5266	0.5016
Train	0.4609	0.5043
Exhibition	0.5440	0.5026
Restaurant	0.5103	0.5310



(a)



(b)

FIG. 10. (Color online) Comparison of six NR algorithms in time-domain waveforms. (a) The noisy and processed signals in the white noise condition. (b) The noisy and processed signals in the car noise condition (dotted line: noisy speech signals; solid line: processed speech signals).

the test in terms of the subjective indices SIG, BAK, and OVL. To facilitate the comparison, the NR algorithms that have attained good subjective performance (with no statistical difference) are marked with asterisks in the table. In Figs. 10(a) and 10(b), surprisingly, in contrast to the results of objective evaluation, the KLT-NR algorithm performed quite poorly in SIG for all noise conditions. The price paid for high NR using the KLT-NR algorithm is obviously the signal distortion, which was noticed by many subjects. Despite the excellent performance in SIG, the Wiener filtering algorithm received the lowest scores in BAK for all noise conditions,

TABLE V. The MANOVA output of the listening test of the six NR algorithms. Cases with significance value p below 0.05 indicate that statistically significant difference exists among all methods.

	Significance value p		
Noise type	SIG	BAK	OVL
White noise	0.008	0.000	0.008
Car noise	0.011	0.000	0.093

TABLE VI. The *post hoc* Tukey HSD test of the subjective measures SIG, BAK, and OVL obtained using six NR algorithms. The NR algorithms that have attained good subjective performance (with no statistical difference) are marked with asterisks.

	SIG		BAK		OVL	
	Noise condition					
NR algorithms	White	Car	White	Car	White	Car
Spectral subtraction	*	*				*
Wiener filtering	*	*			*	*
MMSE-VAD-NR	*	*	*	*	*	*
MMSE-TRA-NR	*	*	*		*	*
LPC-MMSE-TRA-NR		*	*	*	*	*
KLT-NR			*	*		*

which is consistent with the observation in the objective evaluation. The spectral-subtraction algorithm received the lowest grade in BAK for all noise conditions because of the "musical noise"[11] problem, which is quite disturbing to the listeners. There is no significant difference in OVL among all NR algorithms for the car noise scenario. The spectral-subtraction and KLT-NR algorithms received lower scores in OVL than the other algorithms in the white noise case. It can be concluded that the MMSE-VAD-NR, MMSE-TRA-NR, and LPC-MMSE-TRA-NR algorithms are superior to the other algorithms.

Overall, these three algorithms performed equally well in terms of all subjective indices in the two noise scenarios. For background noise with rapidly varying levels, however, the MMSE-TRA-NR algorithm should be more practical than the MMSE-VAD-NR. The LPC preprocessor may contribute to enhancing the NR algorithms, albeit this observation is not statistically significant.

C. Sensitivity analysis in the MMSE-TRA-NR algorithm

In this section, a sensitivity analysis is presented to demonstrate the effect of the choice of TRA parameters. The SA method is employed to search for the optimal parameters β and δ of the aforementioned MMSE-TRA-NR algorithm in dealing with nine types of background noise at the SNR level of 5 dB. These nine types of noise include babble, station, car, airport, street, train, exhibition, restaurant, and white noise, which were taken from the database of Ref. 11. The results of the optimal parameters β and δ summarized in Table IV are plotted in a scatter diagram in Fig. 11 for each noise condition. It is worth noting that the NR performance of the MMSE-TRA-NR algorithm is very sensitive to the choice of the parameter β. The optimal parameter β falls in the range of $0 \leqslant \beta \leqslant 1$ for all noise conditions, which is quite different from the values of $15 \leqslant \beta \leqslant 30$ recommended in Ref. 11. By contrast, the optimal parameter δ is relatively constant (≈ 0.5) for all types of background noise except for "babble" ($\delta = 0.871$), which is also different from the value $\delta = 1.5$ recommended in Ref. 11. The recommended parameter δ should be in the range of $0.5 \leqslant \delta \leqslant 1.5$ because δ decides the transition point of the sigmoid function in the previous TRA algorithm. The transition point can be considered as a threshold to discriminate speech presence from speech absence according to the *a posteriori* SNR. In the present study, a judicious but more reasonable range of $0.5 \leqslant \delta \leqslant 1.5$ is recommended.

V. CONCLUSIONS

An optimized MMSE-TRA-NR algorithm has been presented. The SA optimization technique is exploited to search for optimal TRA parameters, especially the parameter β, which has a profound impact on the estimation of the noise spectrum and hence the resulting NR performance of the algorithm. The optimal parameter β generally falls in the range of $0 \leqslant \beta \leqslant 1$, whereas the optimal parameter δ stays at a relatively constant value of 0.5 for many types of background noise. In addition, a LPC preprocessor has been presented to enhance the MMSE-TRA-NR algorithm.

The proposed NR algorithms have been compared with several other widely used algorithms via extensive objective and subjective tests. These methods exhibit different degrees in trading off reduction performance and speech quality. It can be concluded that the MMSE-VAD-NR, MMSE-TRA-NR, and LPC-MMSE-TRA-NR algorithms are more superior

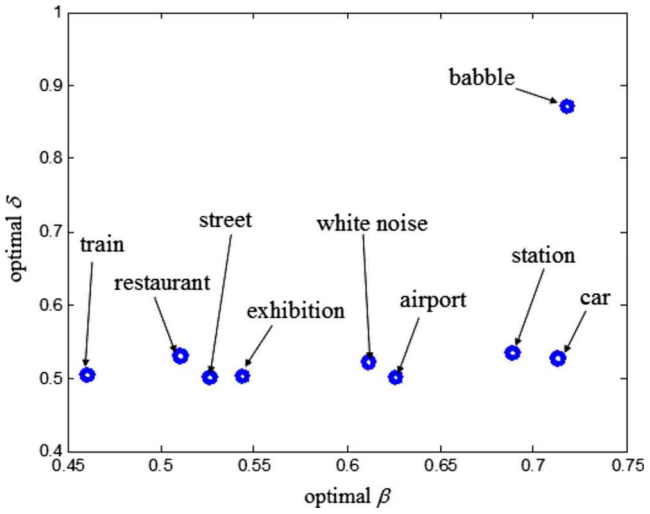


FIG. 11. (Color online) Sensitivity analysis of the optimal parameters β and δ of the MMSE-TRA-NR algorithm for nine kinds of background noise.

to the other algorithms. Overall, these three algorithms performed equally well in terms of all subjective indices in the white and car noise scenarios. For background noise with rapidly varying levels, however, the MMSE-TRA-NR algorithm is more practical than the MMSE-VAD-NR.

ACKNOWLEDGMENTS

This work was supported by the National Science Council of Republic of China, under Project No. NSC 95-2221-E-009-179.

[1]Y. Ephraim and D. Malah, "Speech enhancement using a minimum mean-square error short time spectral amplitude estimator," IEEE Trans. Acoust., Speech, Signal Process. **32**, 1109–1121 (1984).
[2]R. J. McAulay and M. L. Malpass, "Speech enhancement using a soft-decision noise suppression filter," IEEE Trans. Acoust., Speech, Signal Process. **28**, 137–145 (1980).
[3]E. Hänsler and G. Schmidt, *Acoustic Echo and Noise Control: A Practical Approach* (Wiley, New York, 2004).
[4]R. E. Crochiere, "A weighted overlap-add method of short-time Fourier analysis/synthesis," IEEE Trans. Acoust., Speech, Signal Process. **281**, 99–102 (1980).
[5]M. R. Portnoff, "Implementation of the digital phase vocoder using the fast Fourier transform," IEEE Trans. Acoust., Speech, Signal Process. **24**, 243–248 (1976).
[6]U. Zölzer, *DAFX—Digital Audio Effects* (Wiley, New York, 2002).
[7]S. L. Gay and J. Benesty, *Acoustic Signal Processing for Telecommunication* (Kluwer Academic, Norwell, MA, 2000).
[8]N. Wiener, *Extrapolation, Interpolation, and Smoothing of Stationary Time Series with Engineering Applications* (Wiley, New York, 1949).
[9]B. Farhang-Boroujeny, *Adaptive Filters Theory and Application* (Wiley, New York, 2000).
[10]S. V. Vaseghi, *Advanced Signal Processing and Digital Noise Reduction* (Wiley, New York, 1996).
[11]P. C. Loizou, *Speech Enhancement Theory and Practice* (CRC, New York, 2007).
[12]Y. Hu and P. C. Loizou, "A generalized subspace approach for enhancing speech corrupted by colored noise," IEEE Trans. Acoust., Speech, Signal Process. **11**, 334–341 (2003).
[13]L. Lin, W. Holmes, and E. Ambikairajah, "Adaptive noise estimation algorithm for speech enhancement," Electron. Lett. **39**, 754–755 (2003).
[14]N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, "Equations of state calculations by fast computing machines," J. Chem. Phys. **21**, 1087–1092 (1953).
[15]*Quantum Annealing and Related Optimization Methods*, edited by A. Das and B. K. Chakrabarti (Springer, Heidelberg, 2005).
[16]J. De Vicente, J. Lanchares, and R. Hermida, "Placement by thermodynamic simulated annealing," Phys. Lett. A **317**, 415–423 (2003).
[17]J. Makhoul, "Linear prediction: A tutorial review," Proc. IEEE **63**, 561–580 (1975).
[18]J. D. Markel and A. H. Gray, *Linear Prediction of Speech* (Springer-Verlag, Berlin, 1976).
[19]S. J. Orfanidis, *Optimum Signal Processing: An Introduction* (McGraw-Hill, New York, 1996).
[20]ITU-T Rec. P.862, "Perceptual evaluation of speech quality (PESQ), and objective method for end-to-end speech quality assessment of narrowband telephone networks and speech codecs," International Telecommunications Union, Geneva, Switzerland, 2000.
[21]ITU-T Rec. P.835, "Subjective test methodology for evaluating speech communication systems that include noise suppression algorithm," International Telecommunications Union, Geneva, Switzerland, 2003.
[22]R. Martin, "Noise power spectral density estimation based on optimal smoothing and minimum statistics," IEEE Trans. Acoust., Speech, Signal Process. **9**, 504–512 (2001).
[23]G. Keppel and S. Zedeck, *Data Analysis for Research Designs* (Freeman, New York, 1989).

Adaptive near-field beamforming techniques for sound source imaging

Yong Thung Cho[a] and Michael J. Roan
Department of Mechanical Engineering, Virginia Polytech Institute and State University, Blacksburg, Virginia 24061

(Received 5 May 2008; revised 4 November 2008; accepted 7 November 2008)

Phased array signal processing techniques such as beamforming have a long history in applications such as sonar for detection and localization of far-field sound sources. Two sometimes competing challenges arise in any type of spatial processing; these are to minimize contributions from directions other than the look direction and minimize the width of the main lobe. To tackle this problem a large body of work has been devoted to the development of adaptive procedures that attempt to minimize side lobe contributions to the spatial processor output. In this paper, two adaptive beamforming procedures—minimum variance distorsionless response and weight optimization to minimize maximum side lobes—are modified for use in source visualization applications to estimate beamforming pressure and intensity using near-field pressure measurements. These adaptive techniques are compared to a fixed near-field focusing technique (both techniques use near-field beamforming weightings focusing at source locations estimated based on spherical wave array manifold vectors with spatial windows). Sound source resolution accuracies of near-field imaging procedures with different weighting strategies are compared using numerical simulations both in anechoic and reverberant environments with random measurement noise. Also, experimental results are given for near-field sound pressure measurements of an enclosed loudspeaker. © *2009 Acoustical Society of America.* [DOI: 10.1121/1.3050248]

PACS number(s): 43.60.Fg, 43.60.Jn, 43.60.Lq, 43.60.Mn [EJS] Pages: 944–957

I. INTRODUCTION

Beamforming has a long history of use in applications such as sonar for detection and localization of far-field sound sources.[1,2] For sources that lie in the far-field, beamforming has two primary uses: first, determining the direction to the source and, second, enhancing the signal-to-noise ratio. With modification, standard beamforming procedures can be used for near-field sound source imaging. Traditional far-field delay-and-sum beamforming can be modified to give good imaging performance in the near-field by using beamforming weights that are inversely proportional to the distance from the source to the measurement locations. Previous work has shown the utility in applying standard beamforming techniques to sound source identification. It has been used to reduce the effect of wind noise on sound measurements in a wind tunnel using an array of microphones.[3] Also, the characteristics of jet noise sources were identified using beamformed far-field measurements.[4] Passby noise from a vehicle has been measured in far-field and visualized using beamforming.[5] The beamforming weighting was estimated using a maximum likelihood estimation of the amplitude of a single spherical source with additive white noise, which resulted in a weighting inversely proportional to the distance from the hypothesized source location to the measurement point.[5]

A large body of work has been devoted to the development of far-field adaptive beamforming procedures with the goal of reducing side lobe contributions to the beamformer output.[1,6–10] Reducing side lobes improves source resolution accuracy for measurements made with and without reverberation. For an equally spaced linear array, it is possible to analytically find weights that minimize side lobe level or main lobe beam width using far-field pressure measurements.[6] The minimum variance distorsionless response (MVDR) beamformer is one of the most widely used adaptive beamforming procedures and improves source resolution accuracy by adaptively finding the weights that minimize the output noise variance due to signals that arrive from directions other than the hypothesized source direction.[1,7,8] The side lobe level of beamformed pressure can also be reduced by finding weights to minimize the maximum side lobes using an optimization procedure.[9,10]

The conventional delay-and-sum beamforming procedure can be modified to visualize sound sources based on near-field measurements.[11] This is accomplished by modifying the form of the conventional beamforming (CBF) weights such that the beamformer focuses at specific points between the face of the array and the source rather than by steering the beamformer to coherently sum source contributions from a given direction. The weighting used to accomplish distance specific focusing is inversely proportional to the distance from the source to the measurement points. This reduces the weighting of measurements farther away from the source. This weighting method significantly improves the source resolution accuracy of the beamforming procedure based on near-field measurements.

In this work, two adaptive beamforming procedures: MVDR and optimized weights to minimize the maximum

[a]Author to whom correspondence should be addressed. Electronic mail: cho.yong@gmail.com

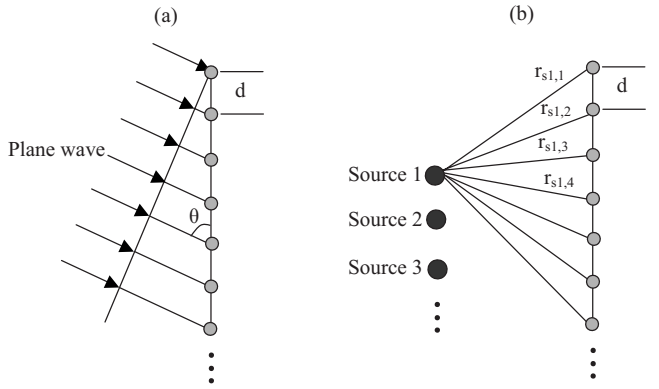


FIG. 1. Beamforming of measurement pressure in line array. (a) Plane wave source; (b) point sources.

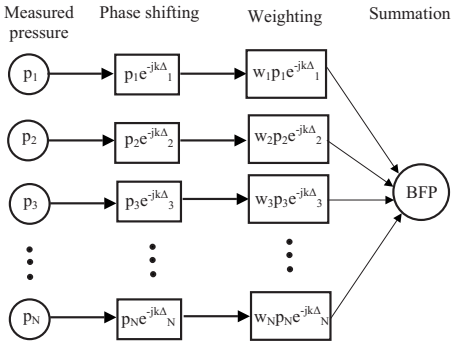


FIG. 2. Beamforming algorithm of measurement pressure.

side lobes (from here on referred to as optimized weights) are modified to perform acoustic source imaging based on near-field measurements. In addition, a weighting scheme is implemented where the weights are inversely proportional to higher orders of the distance from the hypothesized source to the measurement points. Near-field beamforming weights focusing at source locations are estimated based on spherical wave array manifold vectors with spatial windows. Numerical simulations compare the sound source resolution accuracies of the various weighting strategies. Multipole simulations compare the relative performance of the various weighting strategies when estimating beamformed intensity from near-field pressure measurements. The effects of random measurement noise and reverberation are quantified via appropriate simulations. In addition, near-field sound pressure of an enclosed loudspeaker was anechoically measured, and beamforming intensity estimates using various near-field beamforming procedures were compared.

II. SOUND SOURCE IMAGING USING HIGH-RESOLUTION NEAR-FIELD BEAMFORMING

This section introduces three high-resolution beamforming procedures for sound source imaging using near-field microphone measurements. In order to have good source resolution accuracy when using beamforming techniques, two criteria should be met: (1) the beamformer main lobe should be as narrow as possible and (2) side lobes should be as low as possible. The three high-resolution beamforming procedures introduced in this section each attempt to satisfy these criteria. The first technique uses weights that are inversely proportional to higher orders of the distance from the source to the measurement points. The net effect of this weighting scheme is that the beamformer focuses at points between the face of the array and the source. This reduces the contributions from all other points when reconstructing the source image. Second, an adaptive beamforming procedure, near-field MVDR, is introduced to minimize side lobe contributions by adaptively placing nulls in the direction of sources other than the steering directions. Lastly a procedure is introduced that optimizes weights to minimize the maximum side lobes. The accuracies of the latter methods are compared in Secs. III and IV using simulations and measurements of an enclosed loudspeaker in an anechoic chamber.

A. Near-field beamforming for sound source imaging

Beamforming is an effective technique to image sound sources using near-field measurements of the sound pressure field. This subsection provides the necessary background on the near-field beamforming technique.

The wave propagation from the sound source and the pressure measurement geometry are shown in Fig. 1. The beamforming algorithm for measurement pressure in the frequency domain is shown in Fig. 2. First, pressure is measured with an array of microphones, and the phase of the measurement pressure is shifted to account for the delay of wave propagation due to the difference in path length between microphones. Then, the weights are multiplied with the phase shifted measurement pressure and summed to estimate beamforming pressure for a given steering angle θ and frequency ω.

By focusing the beamformer at a point instead of a certain direction θ, as shown in Fig. 1(b), the locations of sources can be identified. Assuming $e^{-j\omega t}$ time convention, beamforming pressure focused on source j, BFP_j, can be represented as

$$\mathrm{BFP}_j = w_{j,1}p_1e^{-jk\Delta_{j,1}} + w_{j,2}p_2e^{-jk\Delta_{j,2}} + w_{j,3}p_3e^{-jk\Delta_{j,3}} + \cdots + w_{j,N}p_Ne^{-jk\Delta_{j,N}}. \quad (1)$$

The delay of the path length of pressure measured in each microphone focusing at source j, $\Delta_{j,i}$, is

$$\Delta_{j,i} = r_{sj,i}, \quad (2)$$

where $r_{sj,i}$ is the distance from the hypothesized source j to measurement location i. Weightings $w_{j,i}$ for uniform weighting is

$$w_{j,i} = \frac{1}{N}, \quad (3)$$

which is simply averaging the phase shifted measurement pressure. However, if weightings inversely proportional to distance from source to measurement location are assumed, then

$$w_{j,i} = \frac{1}{\Delta_{j,i}N} = \frac{1}{r_{sj,i}N}. \quad (4)$$

Normalizing the weights with respect to distance,

$$w_{j,i} = \frac{1}{r_{sj,i}\sum_{k=1}^{N}\frac{1}{r_{sj,k}}}, \tag{5}$$

ensures that the sum of the weights is 1 as for uniform weighting [Eq. (3)].

Beamforming pressure in Eq. (1) can also be applied to a two-dimensional array, and beamforming pressure can be estimated on the source surface. This process is similar to the back-projection procedure of acoustical holography.[12–17] Since beamforming pressure can be estimated on an infinite number of surfaces close to the source, the beamforming particle velocity on the source surface can be found using Euler's equation. Also beamforming intensity on the source surface can be calculated from beamforming pressure and particle velocity estimated on the source surface.

More generally, anechoic, noise-free sound pressure $\mathbf{p}$ due to source signal $\mathbf{s}$, can be represented as

$$\mathbf{p} = \mathbf{sv}, \tag{6}$$

where $\mathbf{v}$ is an array manifold vector, which is a transfer function between source signals and measurement pressure,

$$\mathbf{v} = [A_1 e^{j\phi_1}, \quad A_2 e^{j\phi_2}, \quad A_3 e^{j\phi_3}, \quad \ldots, \quad A_N e^{j\phi_N}], \tag{7}$$

where A_i's and ϕ_i's are the amplitude and phase of transfer function between the source and the measurement pressure. Source signal $\mathbf{s}$ can be estimated from measurement pressure $\mathbf{p}$ by multiplication of weight vector with phase compensation, $\mathbf{w}$, as

$$\mathbf{s} = \mathbf{pw}^T. \tag{8}$$

By postmultiplying $\mathbf{v}$ at both sides of Eq. (8) and is compared with Eq. (6),

$$\mathbf{p} = \mathbf{sv} = \mathbf{pw}^T\mathbf{v} \tag{9}$$

under the constraint[7]

$$\mathbf{w}^T\mathbf{v} = I, \tag{10}$$

where I is an identity matrix. The relationship in Eq. (10) is true regardless of the manifold vector or the source. By postmultiplying $\mathbf{v}^H$ at both sides of Eq. (10) and dividing by $\mathbf{vv}^H$, which is a constant, the weight vector is estimated as

$$\mathbf{w}^T = \mathbf{v}^H/\mathbf{vv}^H. \tag{11}$$

The weight vector estimate for anechoic and noise-free pressure measurement in Eq. (11) is a normalized complex conjugate of the array manifold vector and is independent of source signals.

By assuming spherical wave from the source, the array manifold vector is represented as

$$\mathbf{y} = \left[\frac{1}{r_1}e^{jkr_1}, \quad \frac{1}{r_2}e^{jkr_2}, \quad \frac{1}{r_3}e^{jkr_3}, \quad \ldots, \quad \frac{1}{r_N}e^{jkr_N}\right], \tag{12}$$

where r_i's represent the distance from the source to the measurement location. By substituting the array manifold vector in Eq. (12) into the weight vector in Eq. (11), amplitude weights are

$$w_i = \frac{1}{r_i\sum_{k=1}^{N}\frac{1}{r_k^2}}, \tag{13}$$

which is identical to the amplitude of weights presented in Eq. (5) except for the normalization of the amplitude of the weights.[5] Therefore, beamforming weights whose magnitudes are inversely proportional to the distance between the source and the measurement location are obtained by assuming spherical waves propagating from the source in an anechoic, noise-free environment.

B. High order inversely proportional beamforming

Inversely proportional beamforming (IWBF) weights are derived for spherical waves propagating from the source in an anechoic, noise-free environment, as shown in Sec. II A. For more general cases, sound radiation in the radial direction from motion of a sphere can be represented by a combination of spherical Hankel function of order m. For relatively large kr, e.g., $kr > 10$, the spherical Hankel function of order m converges to spherical waves (spherical Hankel function of order $m=0$). The latter implies that IWBF is more accurate for larger kr or at higher frequencies. However, for the smaller kr or at lower frequencies, spherical Hankel functions of higher order m decay more rapidly than $m=0$ as kr increases. The latter property of the spherical Hankel function causes the measurement pressure at smaller kr or at lower frequencies containing high order components to drop rapidly below the noise floor. The higher order measurement pressure is dominated by measurement noise as the measurement is taken farther away from the source. However, measurement pressure taken farther away from the source, especially at low frequencies, possibly corrupted by measurement noise can be filtered during the beamforming process by implementing high order IWBF.

A significantly higher resolution of the source can be obtained using weightings that are inversely proportional to higher orders of the distance from source to measurement location rather than uniform weightings. If higher order weightings inversely proportional to distance from source to measurement location are assumed, the weightings are

$$w_{j,i} = \frac{1}{\Delta_{j,i}^n N} = \frac{1}{r_{sj,i}^n N}, \tag{14}$$

where n is the order of inversely proportional weighting. Normalizing the higher order weights with respect to distance gives

$$w_{j,i} = \frac{1}{r_{sj,i}^n\sum_{k=1}^{N}\frac{1}{r_{sj,k}^n}}. \tag{15}$$

This ensures that the sum of the weights is 1 as for uniform weighting [Eq. (3)].

Beamforming weights shown in Eqs. (5), (14), and (15) can also be represented as spatial windows applied to mea-

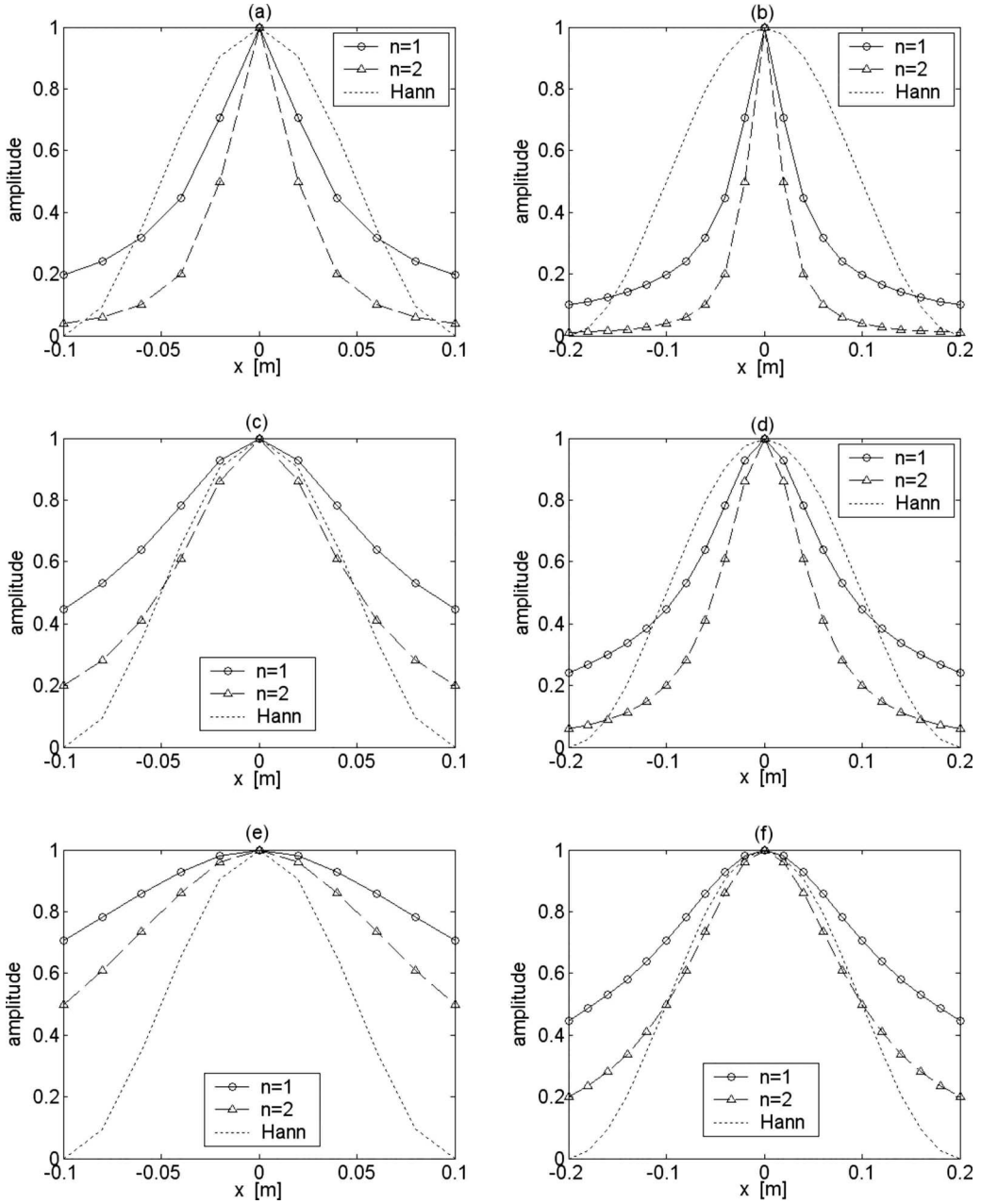


FIG. 3. Normalized amplitude of IWBF window and Hanning window applied to measurement pressure with various measurement locations and measurement aperture sizes. (a) L_x=0.2 m, z=0.02 m; (b) L_x=0.4 m, z=0.02 m; (c) L_x=0.2 m, z=0.05 m; (d) L_x=0.4 m, z=0.05 m; (e) L_x=0.2 m, z=0.1 m; (f) L_x =0.4 m, z=0.1 m.

surement pressure. Normalized amplitudes of the IWBF and Hanning windows applied to measurement pressure with various measurement locations and measurement sizes are shown in Fig. 3. The lengths of the linear measurement array, L_x, are 0.2 and 0.4 m, and the spacing between measurement points is 2 cm. A point source is supposed to be located at the coordinate origin, and the measurement array is located at z=2 cm, z=5 cm, and z=10 cm.

C. Minimum variance distorsionless response beamformer

In this section, the adaptive near-field MVDR beamforming procedure is derived to minimize contributions from sources that lie in directions other than the focusing point. Near-field MVDR beamformer weights are derived to minimize the output noise power or maximize the array gain with a distortionless response constraint.[1,7,8]

First, measurement pressure can be represented as a superposition of the pressure directly radiated from source and measurement noise,

$$\mathbf{p} = \mathbf{p_s} + \mathbf{p_n}, \tag{16}$$

where $\mathbf{p_s}$ is the pressure directly radiated from source and $\mathbf{p_n}$ is the measurement noise. The source property estimate, $\mathbf{s_n}$, based on measurement pressure, $\mathbf{p}$, can be represented using a beamforming weight vector, which is,[1]

$$\mathbf{s_n} = \mathbf{p}\mathbf{w}^T. \tag{17}$$

Cross-power spectral matrix or variation of the source output estimate based on measurement pressure, $\mathbf{S_n}$, is

$$\mathbf{S_n} = \mathbf{s_n}^H\mathbf{s_n} = \mathbf{w}^*\mathbf{P}\mathbf{w}^T, \tag{18}$$

where $\mathbf{P}$ is the cross-power spectral matrix of the measurement pressure with noise. To find the maximum value of the variance of the source output and the corresponding weights, a Lagrange multiplier was used, satisfying the distorsionless response relationship between the array manifold vector and weights, which is[1]

$$\mathbf{v}\mathbf{w}^T = 1. \tag{19}$$

The maximization function, F, is

$$F = \mathbf{w}^*\mathbf{P}\mathbf{w}^T + \lambda(1 - \mathbf{v}\mathbf{w}^T), \tag{20}$$

where λ is the Lagrange multiplier. By taking the derivative of the maximization function, F, with respect to the weight vector, $\mathbf{w}^T$, and if the derivative is zero when maximum,

$$\frac{\partial F}{\partial \mathbf{w}^T} = \mathbf{w}^*\mathbf{P} - \lambda\mathbf{v} = 0, \tag{21}$$

and $\mathbf{w}^*$ reduces to

$$\mathbf{w}^* = \lambda\mathbf{v}\mathbf{P}^{-1}. \tag{22}$$

By substituting Eq. (22) into the Hermitian of Eq. (19), the Lagrange multiplier, λ, is estimated as

$$\lambda = \frac{1}{\mathbf{v}\mathbf{P}^{-1}\mathbf{v}^H}. \tag{23}$$

By substituting Eq. (23) into Eq. (22), $\mathbf{w}^*$ is estimated as

$$\mathbf{w}^* = \frac{\mathbf{v}\mathbf{P}^{-1}}{\mathbf{v}\mathbf{P}^{-1}\mathbf{v}^H}. \tag{24}$$

The weight vector derived in Eq. (24) maximizes the beamformed source output and is a function of the array manifold vector and cross-power spectrum of the measurement pressure. No specific wave type was assumed in the array manifold vector while deriving the weight vector except for the constraint between the weight and the array manifold vectors. There is no restriction about the source type when estimating the array manifold vector in Eq. (24). Either planar wave or spherical wave sources can be implemented when estimating the array manifold vector. Also spatial windows can be incorporated in the array manifold vector combined with spherical wave sources. As a result, measurement pressure taken farther away from the source especially at low frequencies possibly corrupted by measurement noise can be filtered during the beamforming process.

For a beamformer *focusing* at source j, $\mathbf{w}_j$ is the vector of weights with compensated phase, defined as

$$\mathbf{w}_j = [w_{j,1}e^{-jk\Delta_{j,1}}, w_{j,2}e^{-jk\Delta_{j,2}}, w_{j,3}e^{-jk\Delta_{j,3}}, \ldots, w_{j,N}e^{-jk\Delta_{j,N}}]. \tag{25}$$

The MVDR weight vector minimizing the output noise power or maximizing the array gain focusing at source j is derived to be[1,7,8]

$$\mathbf{w}_j^* = \frac{\mathbf{v}_j\mathbf{P}^{-1}}{\mathbf{v}_j\mathbf{P}^{-1}\mathbf{v}_j^H}, \tag{26}$$

where $\mathbf{P}$ is a cross-power spectral matrix of measurement pressure. The cross-power spectral matrix of measurement pressure is estimated by the multiplication of Hermitian of the measurement pressure vector and the measurement pressure vector itself.

However, cross-power spectral matrix may be ill conditioned and requires regularization for estimating the inverse matrix. For a more accurate near-field implementation of MVDR, spatial filtering and spherical wave are incorporated in the array manifold vector. In order to have accurate near-field focusing based on the MVDR algorithm, a new array manifold vector $\mathbf{X}_j$ with spatial filtering and spherical wave is introduced as

$$\mathbf{X}_j = \left[\frac{e^{jk\Delta_{j,1}}}{\mu_{j,1}^n}, \frac{e^{jk\Delta_{j,2}}}{\mu_{j,2}^n}, \frac{e^{jk\Delta_{j,3}}}{\mu_{j,3}^n}, \ldots, \frac{e^{jk\Delta_{j,N}}}{\mu_{j,N}^n}\right], \tag{27}$$

where $n-1$ is the order of the spatial window and $n=1$ represents a spherical wave with uniform window. The delay of the path length of pressure measured by each microphone is then focused on a surface between source j and measurement surface, and $\mu_{j,i}$ is

$$\mu_{j,i} = r_{fsj,i}, \tag{28}$$

where $r_{fsj,i}$ is the distance from point j on surface between source and measurement surface to measurement location i. The amplitude of the near-field MVDR weight vector is

$$\mathbf{A}_j = |\mathbf{X}_j\mathbf{P}^{-1}/(\mathbf{X}_j\mathbf{P}^{-1}\mathbf{X}_j^H)|. \tag{29}$$

The normalized and phase corrected near-field MVDR weight vector is now given by

$$\mathbf{w}_{j,i} = \mathbf{A}_{j,i}\mathbf{v}_{\mathbf{p}j,i}^* \Big/ \sum_{k=1}^{N}\mathbf{A}_{j,k}, \tag{30}$$

where $\mathbf{v_p}$ is defined as

$$\mathbf{v}_{\mathbf{p}j} = [e^{jk\Delta_{j,i}}, e^{jk\Delta_{j,2}}, e^{jk\Delta_{j,3}}, \ldots, e^{jk\Delta_{j,N}}], \tag{31}$$

which is the plane wave array manifold vector focusing at source locations.

The weighting strategy given in Eq. (30) combines the amplitude weighting of MVDR with spherical wave and spatial filtering of measurement pressure and the phase information as used in standard frequency domain delay-and-sum beamforming to give robust adaptive near-field performance. The optimal location of the focusing surface for the calculation of MVDR weight amplitudes in Eq. (29) may or may not coincide with the hypothesized source location in Eq. (31), depending on the measurement geometry, projection distance, etc. This is because the shape of the spatial filter depends on both measurement geometry and projection distance, as shown in Fig. 3, and optimal spatial filtering depends on the source, measurement noise, etc. As the distance between the measurement and the source surfaces increases, the shape of the spatial filtering becomes relatively more

uniform, as shown in Fig. 3. However, for both lower frequencies and signal-to-noise ratio of measurement pressure, a relatively sharper spatial window should be applied to the measurement pressure even though measurement surface is located farther away from the source surface. This can be accomplished by introducing hypothesized source location that is located closer to the measurement surface than on the actual source. The shape of the spatial window is influenced by the choice of hypothesized source location and the order of beamforming spatial window. This type of weighting significantly improves source resolution performance when using near-field measurements and is referred to as near-field MVDR in the present work.

D. Optimization of weights to minimize the maximum side lobe level

Beamforming with optimized array element position and weights to minimize both the number of elements in array and the maximum side lobe level of linear and sparse arrays was investigated extensively for various applications.[18–26] Optimal weights to minimize main lobe width or maximum side lobe level for an equally spaced linear array can be analytically calculated using Dolph–Chebyshev array weighting.[1,6] Optimal weights to minimize maximum side lobe level for both equally spaced and sparse linear and two-dimensional arrays can be estimated using linear programming.[9,10]

Similar to Holm's method,[9,10] the maximum side lobe level for near-field beamforming pressure, that is, maximum beamforming pressure level except the main lobe region, can be minimized by finding the appropriate weighting using optimization. The normalized maximum side lobe level of near-field beamforming pressure δ_s can be estimated from Eq. (1), excluding the main lobe region, as

$$\delta_s = \frac{\max(|\mathrm{BFP}_j|)}{|\mathrm{BFP}_{\max}|} \quad (\text{for } j = 1,2,\cdots,N \text{ except the main lobe region}), \tag{32}$$

where $\mathrm{BFP}_{\max}$ represents the maximum value of BFP_j in the main lobe region. The optimal weights to minimize the maximum side lobe level δ_s are found using the constraint,

$$\sum_{i=1}^{N} w_{j,i} = 1 \quad (\text{for } j = 1,2,\cdots,N). \tag{33}$$

However, in the present work, the weighting is supposed to be a function of only the distance between the focusing point and measurement location as

$$w_{j,i} = f(r_{sj,i}), \tag{34}$$

which remarkably reduces the number of optimal weightings to be estimated and reduces the required computation time by orders of magnitude especially when the number of measurements is large. Higher resolution near-field beamforming with optimized weights to minimize maximum side lobe level can be applied for equally spaced and sparse linear and two-dimensional arrays. Optimized weights to minimize the maximum side lobe level is referred to as optimized weights in the present work.

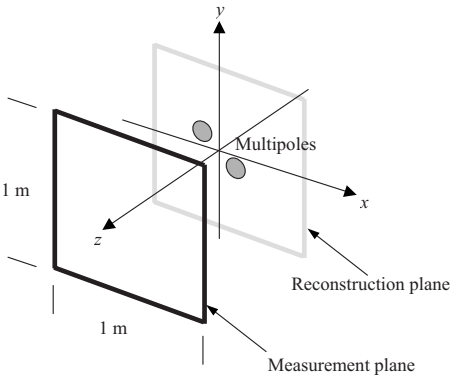


FIG. 4. Measurement geometry definition of multipole simulation with planar array.

III. NUMERICAL SIMULATIONS

Several sets of numerical simulations of sound source imaging based on near-field acoustic measurements with two-dimensional arrays were performed to validate and compare the accuracy of the high-resolution near-field beamforming approaches. Details and results of the numerical simulations are described in this section.

First, anechoic sound fields radiated by out-of-plane multipoles separated by a distance smaller than wavelength were generated and measured numerically without measurement noise using a two-dimensional array at different frequencies and distances from the sources. In addition, reverberant sound fields radiated by the same multipoles were generated and measured numerically with random measurement noise using a two-dimensional array at different frequencies and distances from the sources. Normalized beamforming pressure and intensity were compared with those estimated using uniform weighting, higher order inversely proportional weighting, optimized weighting to minimize maximum side lobe level, and near-field MVDR weighting. Numerical simulations are described in more detail, and results are given in Secs. III A and III B.

A. Anechoic multipole simulation with a two-dimensional array

In this section, a multipole simulation is performed to compare the relative performance of near-field beamforming procedures with different weighting strategies for complicated sources.

The multipole simulations consist of the pressure field generated by ten monopoles. The measurement geometry of the multipole simulation with a 1 m square planar array is shown in Fig. 4. Measurement spacing is 5 cm both in x-and y-directions. Sound pressure is measured at $z=0.1$ m and $z=0.05$ m for 1000 and 3000 Hz, respectively. The location and amplitudes of monopoles that make up the multipoles are shown in Fig. 5. The centroids of the multipoles are

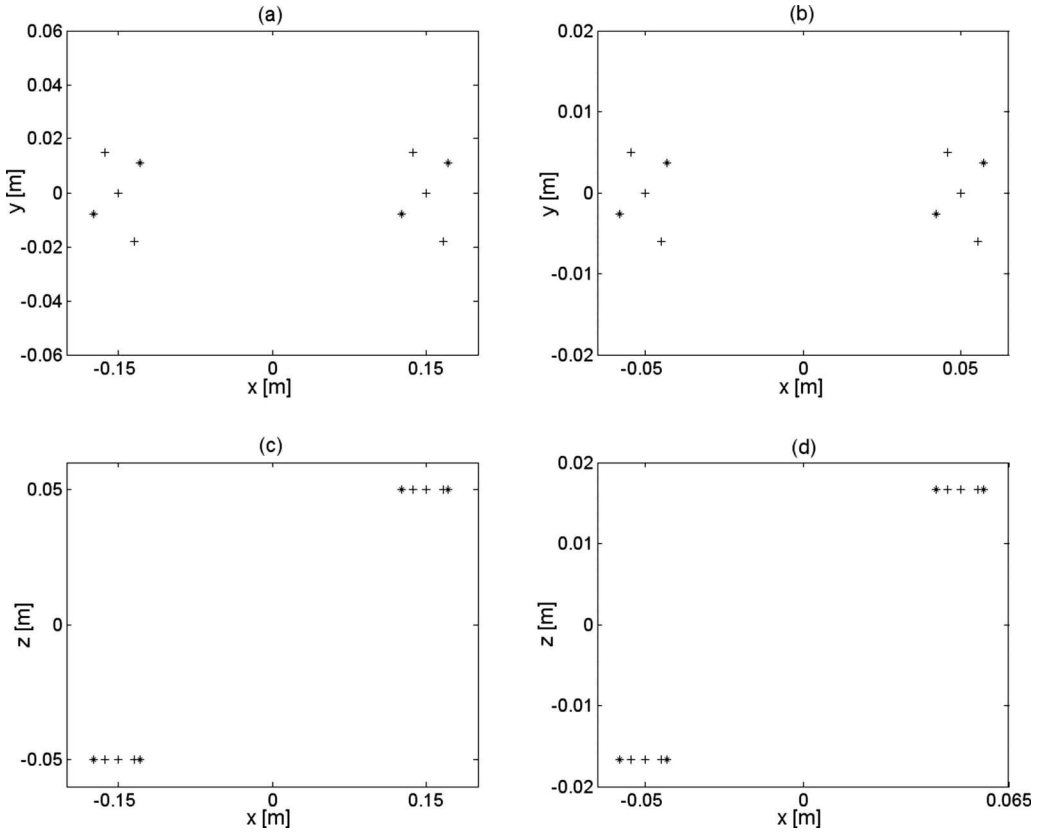


FIG. 5. Location and rms amplitude of monopoles consist of multipoles used for simulation at 1000 and 3000 Hz, where "+" indicates amplitude of positive 1 and " *" indicates amplitude of negative 1. (a) 1000 Hz, *xy*-coordinate; (b) 3000 Hz, *xy*-coordinate; (c) 1000 Hz, *xz*-coordinate; (d) 3000 Hz, *xz*-coordinate.

located off the x-axis, these are $(-0.15, 0, -0.05)$, $(0.15, 0, 0.05)$ m for 1000 Hz and $(-0.05, 0, -0.0167)$, $(0.05, 0, 0.0167)$ m for 3000 Hz.

The normalized amplitude of multipole source beamforming intensity estimated at the $z=0$ plane using sound pressure measurements at $z=0.1$ m and $z=0.05$ m and frequencies of 1000 and 3000 Hz are shown in Figs. 6 and 7. The order of near-field MVDR BF for both frequencies was $n=1$, which corresponds to the array manifold vector for spherical wave propagation. The hypothesized source surfaces to estimate near-field MVDR weights are located at $z=0$ m and $z=0.02$ m for frequencies of 1000 and 3000 Hz, respectively. The normalized beamforming intensity level away from the actual source region for first order inversely weighted beamforming is lower than that of uniformly weighted CBF. So the first order inversely weighted beamforming procedure is more accurate than uniformly weighted CBF for resolving closely located sound sources both at 1000 and 3000 Hz. As the order of inversely weighted beamforming is increased, the intensity level away from the actual source region is decreased. However the intensity level of one of the sources located further away from the measurement surface is also decreased as the order of inversely weighted beamforming is increased. The intensity level of third order inversely weighted beamforming in the vicinity of the source located further away from the measurement surface is significantly lower than that of uniformly weighted CBF or first order inversely weighted beamforming. The intensity level away from the sources is not significantly lower than that of first order inversely weighted beamforming. So for both frequencies, first or second order inversely weighted beamforming performs best for source localization and visualization of nonreverberant multipole sources among IWBF procedures considered in this section.

The next algorithm considered was the optimized weight algorithm. For this algorithm, the optimized weights to minimize the maximum side lobe level were calculated under the assumption that the monopole source was located at the coordinate origin. Since weights are optimized for the monopole source located at the coordinate origin, beamforming intensity estimated at 3000 Hz represents the location of sources better than that estimated at 1000 Hz, probably due to the fact that the source location at 3000 Hz is closer to the coordinate origin than that at 1000 Hz.

The results in Figs. 6 and 7 show that the near-field MVDR beamforming intensity level estimated from near-field measurements provides the clearest image in the source region. The amplitude of beamforming intensity represented as $|\mathrm{In}|$ in the figures and captions is estimated using CBF, IWBF with different orders, optimized weight beamforming, and MVDR. However, second order inversely weighted beamforming provides the lower beamforming intensity level outside of the source region compared to that of near-field MVDR beamforming. Among the beamforming procedures considered in the present work, near-field MVDR beamforming appears to give the best sound source visualization for the slightly out-of-plane multipole simulation.

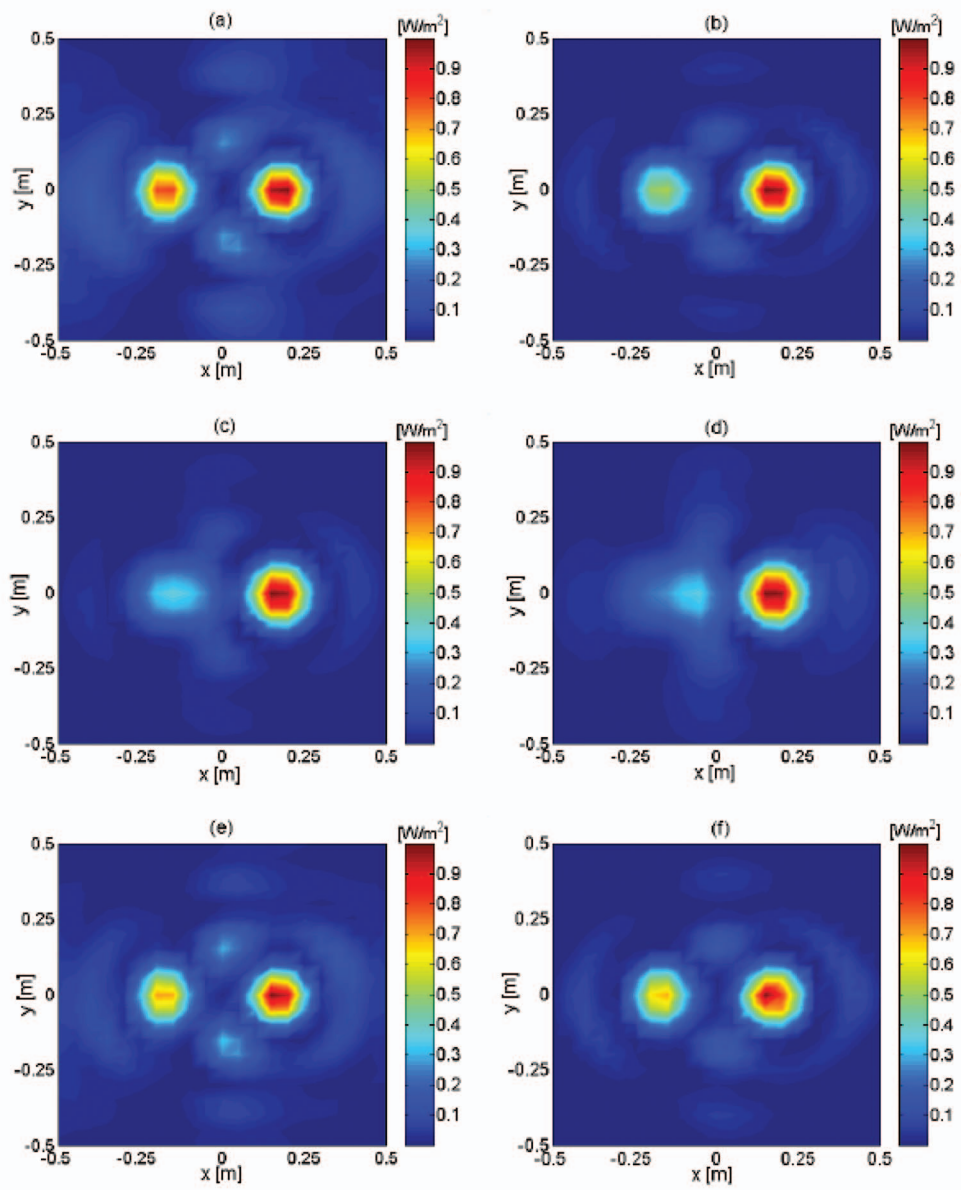


FIG. 6. Normalized amplitude of estimated beamforming multipole intensity based on anechoic measurement at 1000 Hz, z=0.1 m using planar array. (a) CBF |In|, z=0; (b) IWBF 1/R, BF |In|, z=0; (c) IWBF $1/R^2$, BF |In|, z=0; (d) IWBF $1/R^3$, BF |In|, z=0; (e) optimized weights, BF |In|, z=0; (f) MVDR, BF |In|, z=0.

B. Reverberant multipole simulation with two-dimensional array

In Sec. III A, the relative performances of the various near-field beamforming procedures were compared for the out-of-plane multipoles based on noiseless anechoic measurements. In this section, the multipole simulation is performed in a reverberant environment with added random measurement noise. The method of images of sources was used to generate the sound pressure with reverberation.

An identical multipole source and measurement geometry, as described in Sec. III A, is used for the multipole simulation in this section except that now two rigid surfaces are located normal to each other to simulate reverberation. The location of the rigid surfaces relative to the multipole sources is shown in Fig. 8. The reverberant-to-direct energy ratio of the measurement pressure for the multipole simulation is estimated as 6.5% and 5.3% for 1000 and 3000 Hz, respectively. Also the root-mean-square (rms) error between the directly measured pressure with and without reverberation is estimated as 25.5% and 23.1% for 1000 and 3000 Hz, respectively. Since the measurement surface is located close to the source, the reverberant-to-direct energy ratio is small although the rms error between the directly measured pressure with and without reverberation is not small. However, the dimension of measurement geometry shown in Fig. 8 represents pressure measurement in a practical reverberant measurement environment well.

The normalized amplitude of multipole beamforming intensity estimated using MVDR and optimized weights from measurements with reverberation and 20 dB additive random noise at 1000 and 3000 Hz is shown in Fig. 9. By comparing the amplitude of beamforming intensity estimated using both anechoic and reverberant pressure measurements with and without 20 dB additive random noise, it is observed that beamforming intensity estimates are very similar. It is also true for the results of IWBF, which is not shown in the present work. So near-field beamforming procedures with

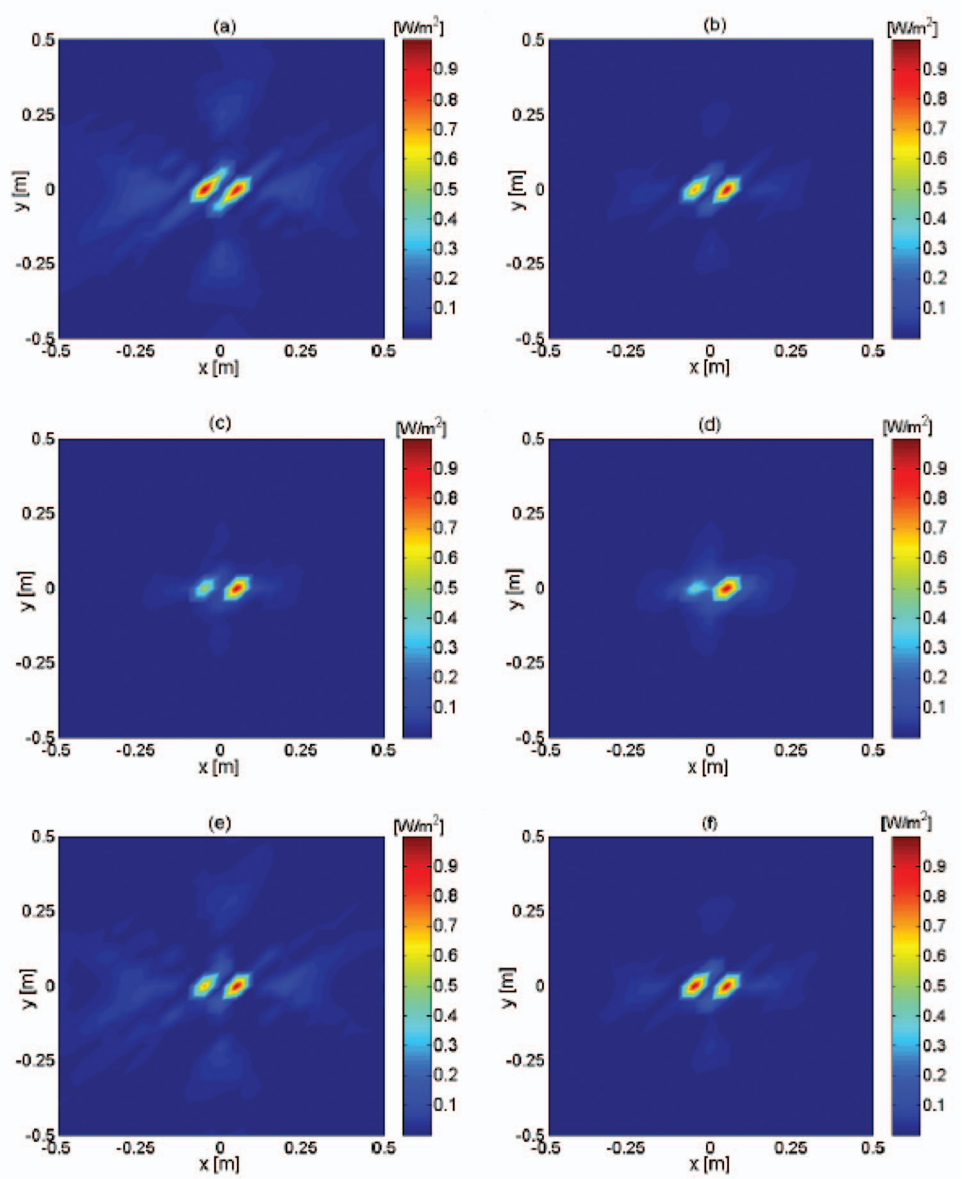


FIG. 7. Normalized amplitude of estimated beamforming multipole intensity based on anechoic measurement at 3000 Hz, z=0.05 m using planar array. (a) CBF |In|, z=0; (b) IWBF 1/R, BF |In|, z=0; (c) IWBF $1/R^2$, BF |In|, z=0; (d) IWBF $1/R^3$, BF |In|, z=0; (e) optimized weights, BF |In|, z=0; (f) MVDR, BF |In|, z=0.

different weighting strategies are effective both for anechoic and reverberant pressure measurements with random measurement noise.

IV. ENCLOSED LOUDSPEAKER MEASUREMENT

The results presented in Sec. III were based on numerical simulations. In this section, experimental results are presented using an enclosed loudspeaker measurement setup. First, the loudspeaker measurement geometry and experimental apparatus are described in Sec. IV A. Then, in Sec. IV B, IWBF, optimized weight beamforming, and near-field MVDR beamforming intensity estimates using a planar measurement surface are compared.

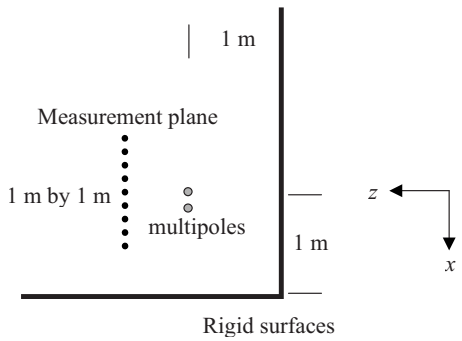


FIG. 8. Location of rigid surfaces relative to multipole sources to simulate reverberation. The centroid of multipoles is located at the coordinate origin.

A. Enclosed loudspeaker measurement description

The near-field measurement geometry for the enclosed loudspeaker experiment is shown in Fig. 10. A 12.7 cm diameter loudspeaker mounted in an enclosure was used as the source. All measurements were done in an anechoic chamber. The actual loudspeaker and enclosure are shown in Fig. 11. The outer surface of the loudspeaker is on the same plane with the surface of the enclosure. The measurement plane was 2 cm above the surface of the enclosure. An array of 11 microphones was used to take simultaneous measurements in the x-direction. The array was then moved in increments of 2 cm in the y-direction to take 16 y-direction measurements. This resulted in a 22×32 cm^2 rectangular measurement sur-

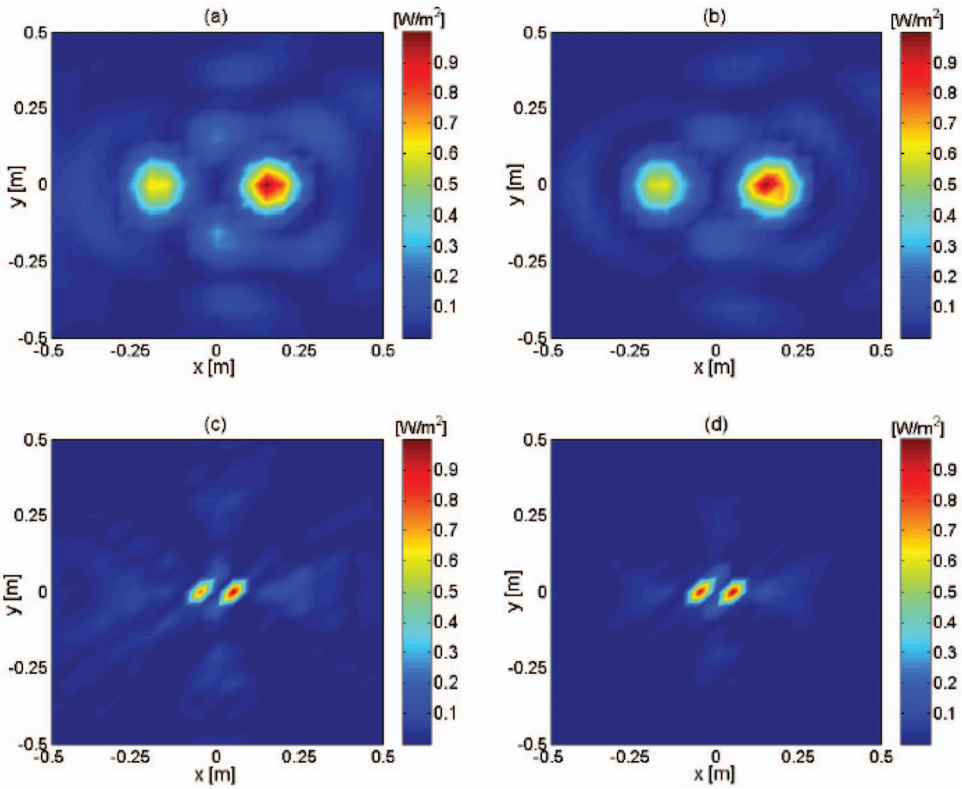


FIG. 9. Normalized amplitude of estimated beamforming multipole intensity at $z=0$ from measurement with reverberation and 20 dB additive random noise at 1000 Hz, $z=0.1$ m and 3000 Hz, $z=0.05$ m using planar array. (a) Optimized weights, BF |In|, 1000 Hz; (b) MVDR, BF |In|, 1000 Hz; (c) optimized weights, BF |In|, 3000 Hz; (d) MVDR, BF |In|, 3000 Hz.

face. The center of both the measurement surface and the source were on the z-axis. The loudspeaker enclosure surface was a 17.3×26.3 cm^2 rectangle and the center of the top surface of the enclosure coincided with the coordinate system origin.

The measurement system consisted of a National Instruments (NI) CompactDAQ chassis, NI cDAQ-9172, and NI 9233 signal conditioner, and a Dell Inspiron 640 m laptop computer was used to run the NI LABVIEW 8.2 data acquisition software. Eleven array microphones (G.R.A.S. Sound & Vibration Type 40 PR) were used to make the sound pressure measurements: they are also shown in Fig. 11. The array microphones were calibrated using a B&K Type 4230 microphone calibrator.

A random signal with a cutoff frequency of 6 kHz was computer generated and played through a JBL power amplifier model 6260. The output of the JBL power amplifier was directly provided as input to the loudspeaker. Also the computer generated random signal was fed directly to the NI 9233 signal conditioner as the reference signal.

Field microphone signals are sampled at 20 kHz with LABVIEW. A 0.25 s long Hanning window was applied to each temporal data record. The low pass filtered signals with cutoff frequency of 8 kHz were fast Fourier transformed and were averaged 119 times with 50% overlap and 4 Hz resolution to estimate the required transfer functions between reference and field microphone signals. The transfer functions between reference and field microphone signals were considered as measurement pressure in the present work.

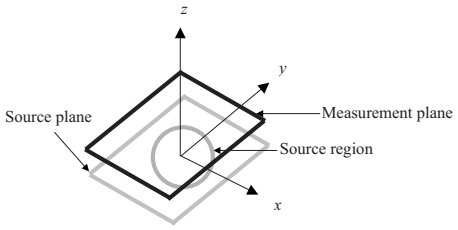


FIG. 10. Enclosed loudspeaker near-field measurement geometry definition.

B. Near-field measurement results

The spatial rms amplitude of the near-field sound pressure measurement of the enclosed loudspeaker is shown in

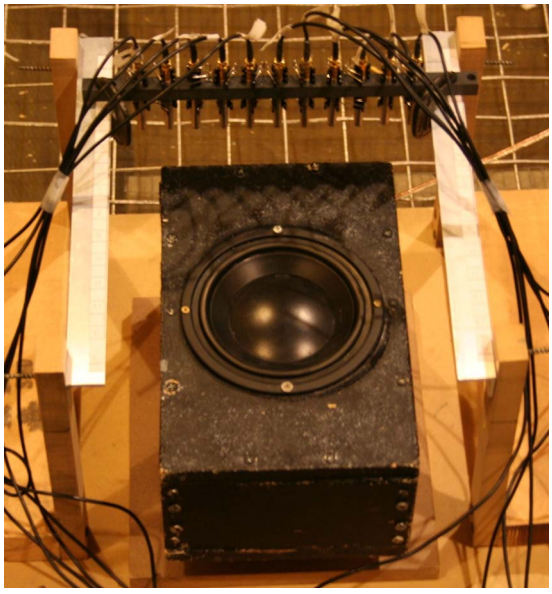


FIG. 11. (Color online) Enclosed loudspeaker source and microphone array for near-field measurement.

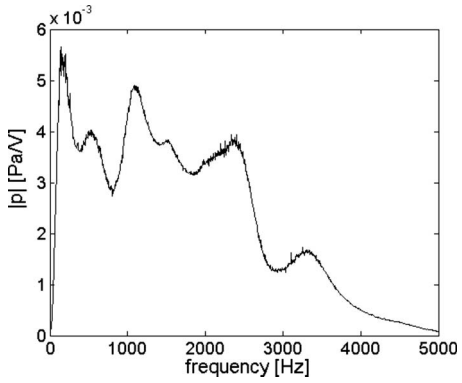


FIG. 12. Spatial rms of near-field pressure measurement.

Fig. 12. IWBF and near-field MVDR beamforming were implemented to operate at frequencies corresponding to the major peaks and dips of the spatial rms amplitude of the near-field sound pressure measurement. Specifically, four frequencies (804, 1088, 2928, and 3312 Hz) were selected from Fig. 12 and the corresponding measurement pressure is shown in Fig. 13. These frequencies correspond to interesting mode shapes of the loudspeaker. 804 and 1088 Hz are modes where the entire surface of the loudspeaker moves in phase. At 2928 Hz, a nodal line appears diagonally across the loudspeaker. Lastly, at 3312 Hz, a mode exists where the center of speaker moves out of phase with the surrounding cone.

The magnitudes of the frequency response functions (the H_1 transfer functions between the response of the microphones and the input of the JBL amplifier) representing the measurement pressure amplitudes for the enclosed loudspeaker are shown in Fig. 12. Also BF intensity estimates 2.5 cm below the measurement surface or 0.5 cm below the loudspeaker enclosure surface using IWBF, optimized weight BF, and near-field MVDR BF are shown in Figs. 14 and 15. The order of IWBF, $n=2$ is used for all frequencies. For near-field MVDR BF, $n=1$ is used for 2928 and 3312 Hz and $n=2$ is used for 804 and 1088 Hz. This was done because MVDR BF, $n=2$, removes measurement noise better than MVDR BF, $n=1$, especially at low frequencies. However, using MVDR BF, $n=1$, provides more detailed information about the source at higher frequencies. Optimized weights are estimated based on a monopole source located 0.5 cm below the coordinate origin and the actual size of the loudspeaker. Both BF intensity estimates at 804 and 1088 Hz are similar in terms of the shape of source even though 804 Hz is one of the lowest dips and 1088 Hz is one of the highest peaks in spatial rms of measurement pressure. Although not shown in the results, the shape of the source at frequencies below 2336 Hz is typically the same as that approximated from the BF intensity estimates at 804 and 1088 Hz. The size of the source approximated from BF intensity estimates at 804 and 1088 Hz represents the actual size of the loudspeaker very well.

The BF intensity estimate of the enclosed loudspeaker measurement using near-field MVDR BF and higher order IWBF is very similar over the range of frequencies, except that measurement noise is removed relatively well in the intensity estimate using either higher order IWBF or near-field MVDR BF. This is not shown in the present work. The

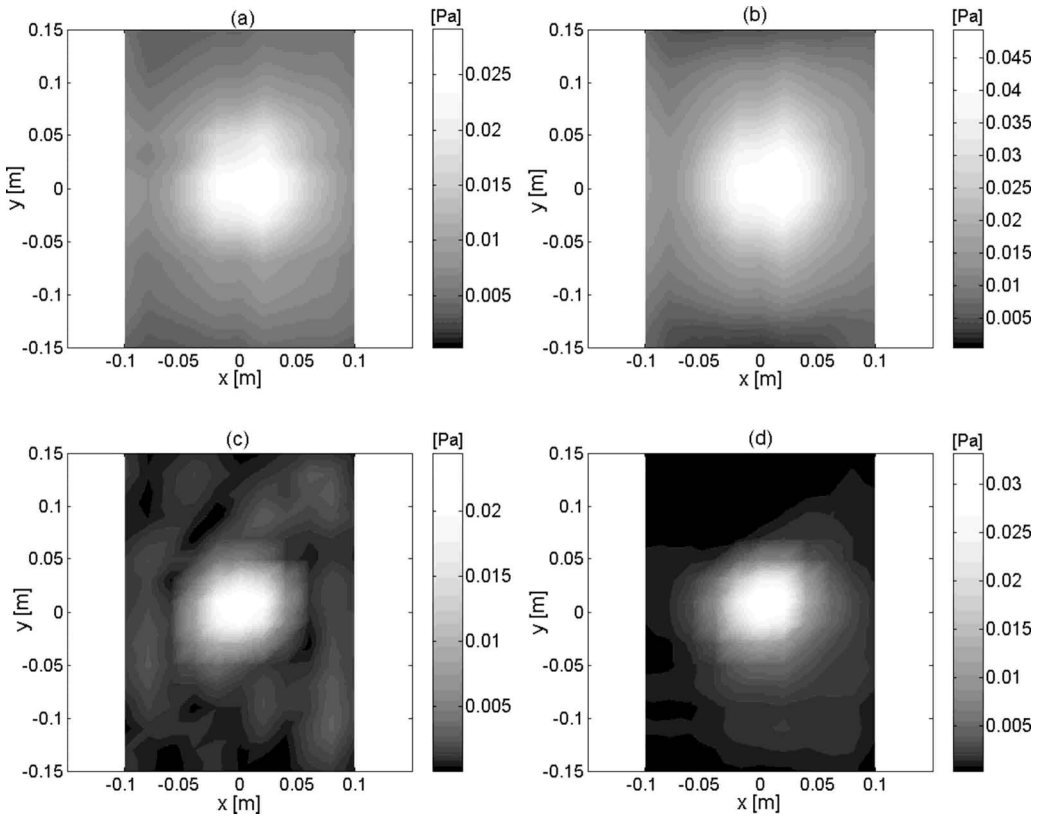


FIG. 13. Enclosed loudspeaker measurement pressure. (a) $|p|$, 804 Hz; (b) $|p|$, 1088 Hz; (c) $|p|$, 2928 Hz; (d) $|p|$, 3312 Hz.

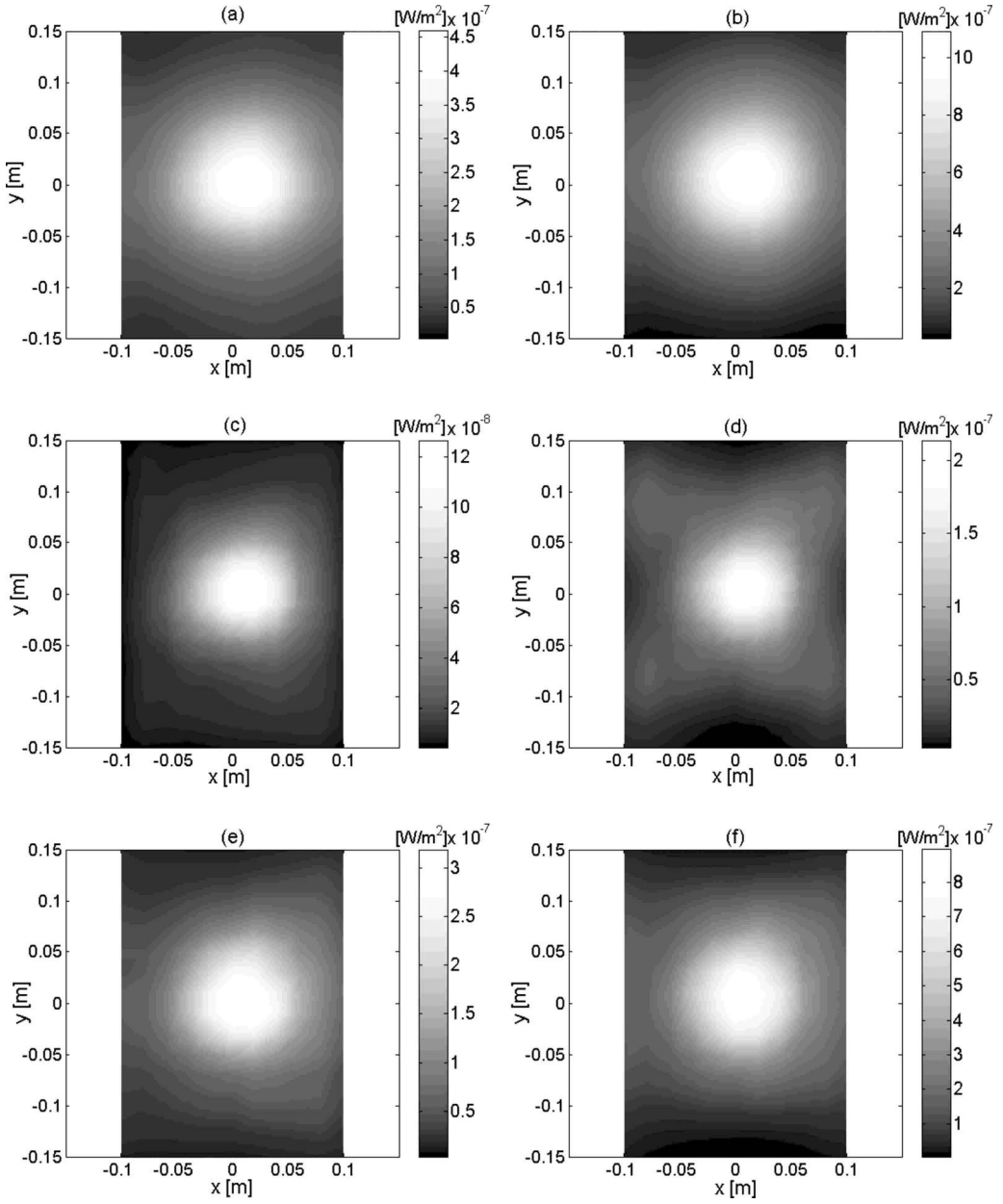


FIG. 14. Beamformed intensity of enclosed loudspeaker measurement. (a) IWBF |In|, 804 Hz; (b) IWBF |In|, 1088 Hz; (c) optimized weights, BF |In|, 804 Hz; (d) optimized weight, BF |In|, 1088 Hz; (e) MVDR BF |In|, 804 Hz; (f) MVDR BF |In|, 1088 Hz.

shape of the source approximated from the BF intensity estimate is very similar for frequencies of 2336 Hz or below. However, both BF intensity estimates at 2928 and 3312 Hz are quite different from those at other frequencies. The lowest dip in spatial rms of pressure measurement is 2928 Hz, and the highest peak in spatial rms of pressure measurements above 3 kHz is 3312 Hz. It appears that near-field MVDR BF intensity provides more detailed information about the source than IWBF or optimized weight BF. The nodal line in BF intensity at 2928 Hz using near-field MVDR BF shown in Fig. 15(e) indicates that the even mode dominates at this frequency. Since even modes are very inefficient sound radiators[27] and 2928 Hz corresponds to the lowest dip in the spatial rms of measurement pressure, the even mode shape at this frequency is reasonable. At 3312 Hz, the BF intensity estimates using IWBF and near-field MVDR BF are similar in that it can be seen that the center of the loudspeaker radiates sound at the highest amplitude. Overall, the size of the source approximated using IWBF, optimized weight BF, and near-field MVDR BF intensity is reasonable compared to the actual size of the loudspeaker.

V. CONCLUSIONS

In the present work, fixed and adaptive beamforming algorithms are modified to provide very effective acoustic source imaging capabilities using near-field measurements. Near-field beamforming weightings are estimated based on spherical wave array manifold vectors with spatial windows. To show this, both simulations and experiments were done for complex sound sources. The improved source resolution accuracy is accomplished by application of different weight-

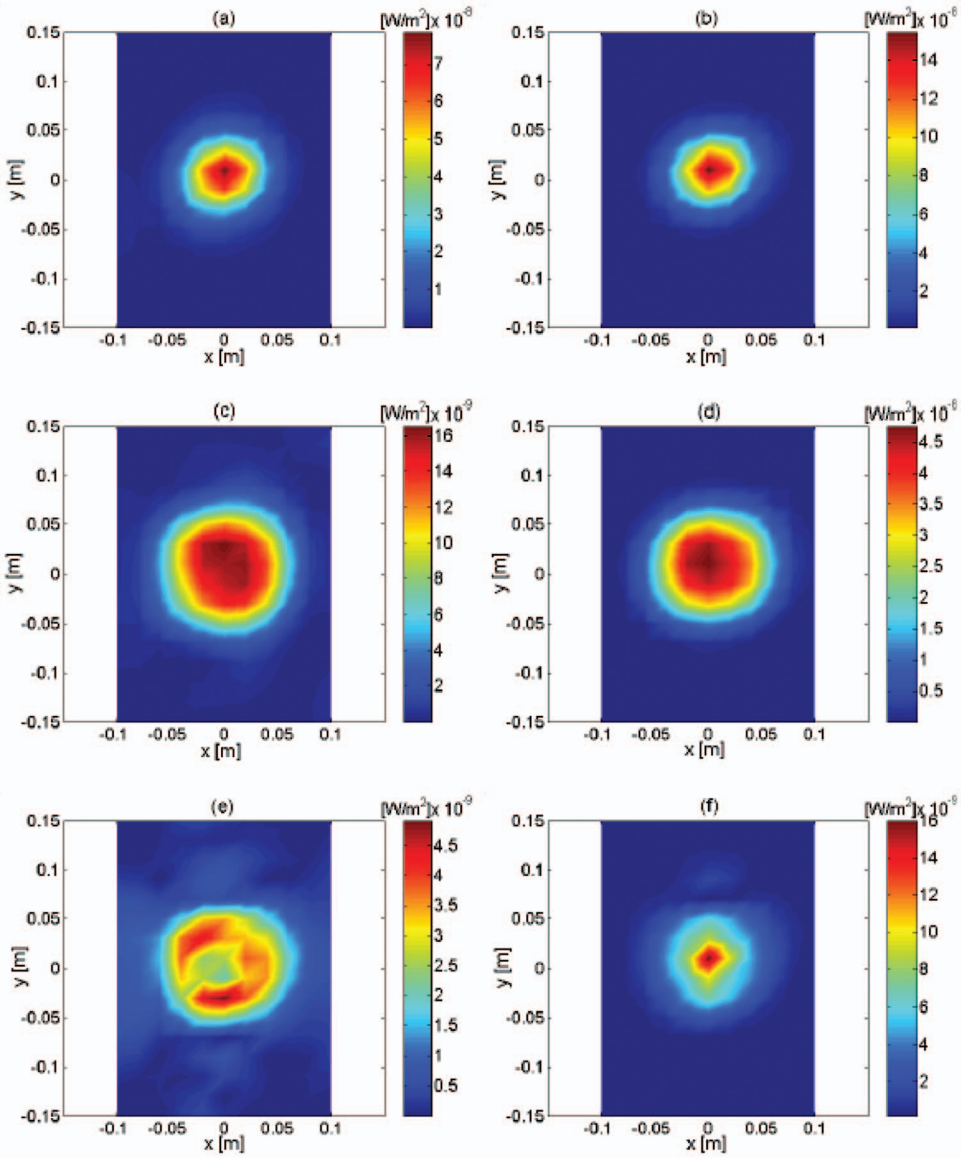


FIG. 15. Beamformed intensity of enclosed loudspeaker measurement. (a) IWBF |In|, 2928 Hz; (b)IWBF |In|, 3312 Hz; (c) optimized weights, BF |In|, 2928 Hz; (d) optimized weight, BF |In|, 3312 Hz; (e) MVDR BF |In|, 2928 Hz; (f) MVDR BF |In|, 3312 Hz.

ing strategies, including higher order inversely proportional weighting, optimized weights to reduce side lobe level, and near-field phase corrected MVDR beamforming with spherical wave array manifold vectors with spatial windows.

The numerical simulation results show that using higher order inversely proportional weighting ($n>2$) does not necessarily improve the source resolution accuracy. The optimized weight algorithm provides superior performance (in terms of source resolution accuracy) only when true source location is known. However, source resolution accuracy degrades when the weights are optimized for unknown or inaccurate source location.

Source resolution accuracy of the standard MVDR beamformer using near-field measurement is not as good as beamforming with first order inversely proportional weights. However, in this work a near-field phase corrected version of MVDR with spherical wave array manifold vectors with spatial windows is introduced and provides significantly improved source resolution accuracy. This new near-field MVDR beamformer is accurate for visualization of sources based on near-field measurements with and without reverberation and random measurement noise.

Multipole simulation were performed both with and without reverberation and additive random measurement noise. Using high-resolution near-field beamforming procedures significantly removes both 20 dB additive random noise in the pressure measurements and reverberation created by two infinite rigid surfaces. In addition, near-field sound pressure of enclosed loudspeaker is measured, and the source is visualized using higher order IWBF, optimized weight BF, and near-field MVDR BF procedures. Both IWBF and near-field MVDR BF procedures provided similar results except at 2928 Hz, which is the lowest dip of the spatial rms of measurement pressure. More detailed visualization of the source is provided by near-field MVDR BF intensity compared to that provided by IWBF or optimized weight BF at 2928 Hz. Overall, it can be concluded that near-field MVDR beamforming and higher order inversely proportional weights with spatial windows can be implemented to visualize sound sources more accurately than CBF with other weighting strategies for various near-field sound pressure measurement environments.

[1]H. L. Van Trees, *Optimum Array Processing: Part IV of Detection, Estimation, and Modulation* (Wiley, New York, 2002).
[2]J. Billingsley and R. Kinns, "The acoustic telescope," J. Sound Vib. **48**, 485–510 (1976).
[3]P. T. Soderman and S. C. Noble, "Directional microphone array for acoustic studies of wind tunnel models," J. Aircr. **12**, 168–173 (1975).
[4]T. Suzuki, "Identification of multipole noise sources in low Mach number jets near the peak frequency," J. Acoust. Soc. Am. **119**, 3649–3659 (2006).
[5]H. Kook, G. B. Moebs, P. Davies, and J. S. Bolton, "An efficient procedure for visualizing the sound field radiated by vehicle during standardized passby tests," J. Sound Vib. **233**, 137–156 (2000).
[6]C. L. Dolph, "A current distribution for broadside arrays which optimizes the relationship between beam width and side-lobe level," Proc. IRE **34**, 335–348 (1946).
[7]J. Capon, "High-resolution frequency-wavenumber spectrum analysis," Proc. IEEE **57**, 1408–1418 (1969).
[8]B. G. Ferguson, "Minimum variance distortionless response beam forming of acoustic array data," J. Acoust. Soc. Am. **104**, 947–954 (1998).
[9]S. Holm and B. Elgetun, "Optimization of the beam pattern of 2D sparse arrays by weighting," Proc.-IEEE Ultrason. Symp. **1995**, 1345–1348.
[10]S. Holm, B. Elgetun, and G. Dahl, "Properties of the beampattern of weight- and layout-optimized sparse arrays," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **44**, 983–991 (1997).
[11]Y. T. Cho, M. J. Roan, and J. S. Bolton, "A comparison of beam forming and acoustical holography for sound source visualization," Proc. Inst. Mech. Eng., Part C: J. Mech. Eng. Sci. (In press, 2009).
[12]J. D. Maynard, E. G. Williams, and Y. Lee, "Nearfield acoustic holography: I. Theory of generalized holography and development of NAH," J. Acoust. Soc. Am. **78**, 1395–1413 (1985).
[13]E. G. Williams, H. D. Dardy, and K. B. Washburn, "Generalized near-field acoustic holography for cylindrical geometry: Theory and experiment," J. Acoust. Soc. Am. **81**, 389–407 (1987).
[14]E. G. Williams, *Fourier Acoustics: Sound Radiation and Near-Field Acoustical Holography* (Academic, London, UK, 1999).
[15]R. Steiner and J. Hald, "Near-field acoustical holography without the errors and limitations caused by the use of spatial DFT," Proceedings of ICSV6 (1999), pp. 843–850.
[16]J. Hald, "Patch near-field acoustical holography using a new statistically optimal method," Proceedings of INTER-NOISE 2003, pp. 2203–2210 (2003).
[17]Y. T. Cho, J. S. Bolton, and J. Hald, "Source visualization by using statistically optimized near-field acoustical holography in cylindrical coordinates," J. Acoust. Soc. Am. **118**, 2355–2365 (2005).
[18]H. Schjær-Jacobsen and K. Madsen, "Synthesis of nonuniformly spaced arrays using a general nonlinear minimax optimization method," IEEE Trans. Antennas Propag. **24**, 501–506 (1976).
[19]P. Jarske, T. Saramäki, S. K. Mitra, and Y. Neuvo, "On properties and design of nonuniformly spaced linear arrays," IEEE Trans. Acoust., Speech, Signal Process. **36**, 372–380 (1988).
[20]R. M. Leahy and B. D. Jeff, "On the design of maximally sparse beam forming arrays," IEEE Trans. Antennas Propag. **39**, 1178–1187 (1991).
[21]D. H. Turnbull and F. S. Foster, "Beam steering with pulsed two-dimensional transducer arrays," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **38**, 320–333 (1991).
[22]P. K. Weber, R. M. Schmitt, B. D. Tylkowski, and J. Steck, "Optimization of random sparse 2-D transducer arrays for 3-D electronic beam sterring and focusing," Proc.-IEEE Ultrason. Symp. **1994**, 1503–1506.
[23]R. E. Davidsen, J. A. Jensen, and S. W. Smith, "Two dimensional random arrays for real time volumetric imaging," Ultrason. Imaging **16**, 143–163 (1994).
[24]G. R. Lockwood, P. Li, M. O'Donnell, and F. S. Foster, "Optimizing the radiation pattern of sparse periodic linear arrays," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **43**, 7–14 (1996).
[25]G. R. Lockwood and F. S. Foster, "Optimizing the radiation pattern of sparse periodic two-dimensional arrays," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **43**, 15–19 (1996).
[26]V. Murino, A. Trucco, and C. S. Regazzoni, "Synthesis of unequally spaced arrays by simulated annealing," IEEE Trans. Signal Process. **44**, 119–123 (1996).
[27]F. Fahy, *Sound and Structural Vibration: Radiation, Transmission and Response* (Academic, London, UK, 1985).

Nonlinear acoustics in cicada mating calls enhance sound propagation

Derke R. Hughes,[a] Albert H. Nuttall,[b] Richard A. Katz, and G. Clifford Carter
Naval Undersea Warfare Center Division, Newport, 1176 Howell Street, Newport, Rhode Island 02841-1708

(Received 3 May 2008; revised 31 October 2008; accepted 15 November 2008)

An analysis of cicada mating calls, measured in field experiments, indicates that the very high levels of acoustic energy radiated by this relatively small insect are mainly attributed to the nonlinear characteristics of the signal. The cicada emits one of the loudest sounds in all of the insect population with a sound production system occupying a physical space typically less than 3 cc. The sounds made by tymbals are amplified by the hollow abdomen, functioning as a tuned resonator, but models of the signal based solely on linear techniques do not fully account for a sound radiation capability that is so disproportionate to the insect's size. The nonlinear behavior of the cicada signal is demonstrated by combining the mutual information and surrogate data techniques; the results obtained indicate decorrelation when the phase-randomized and non-phase-randomized data separate. The Volterra expansion technique is used to fit the nonlinearity in the insect's call. The second-order Volterra estimate provides further evidence that the cicada mating calls are dominated by nonlinear characteristics and also suggests that the medium contributes to the cicada's efficient sound propagation. Application of the same principles has the potential to improve radiated sound levels for sonar applications. [DOI: 10.1121/1.3050258]

PACS number(s): 43.60.Wy, 43.25.Ts, 43.80.Jz [EJS] Pages: 958–967

I. INTRODUCTION

A. Background

The objective of this research is to begin to understand how the cicada, a small insect, emits one of the loudest sounds in all of the insect population despite its relatively small size. Detailed knowledge of the characteristics of the cicada's acoustic signature is a necessary step toward the ultimate goal of transferring this biotechnological feat of nature to a manmade transduction system of similar proportions. The cicada's highly effective sound production system occupies a physical space typically less than 3 cc. Cicadas are sexually dimorphic, and only males possess the structures necessary for making loud audible sounds. Male sounds are broadcast advertisements for attracting females, and they are typically loud, rhythmic, and easily distinguished from background noise. Males create sound by flexing a pair of ridged abdominal membranes called tymbals. The sounds made by these tymbals are amplified by the hollow abdomen functioning as a tuned resonator, as described by current research.[1–6] Nevertheless, the tuned resonator explanation in the current literature does not account for the sound radiation capabilities of the cicada.[7]

Studying the sound production system of the cicada in captivity has some inherent difficulties: cicadas vocalize only on sunny days and do not respond as well to indoor lighting. However, the recent discovery of a female response to the mating call of the male has made it possible to conduct experiments on cicadas in the field. In many species, the females answer loud male signals with quiet wing flick responses.[8–10] Males perceiving such responses will approach that signal and continue to call even if disturbed. These female signals are easily imitated, which provides an important tool for collecting and manipulating cicadas: in an acoustical duet with a female, a male will become sexually excited and continue to sing even if disturbed or manipulated. Similar manipulations without a duet may cause a male to stop calling.

B. Current experimental opportunities with periodic cicadas

Opportunities to collect and study cicadas of the midwestern and eastern United States are limited by the insects' periodical cycles. The insects have a 13 or 17 year life cycle and emerge in mass numbers, known as broods, at predictable times in predictable locations.[11,12] Since the sounds and behaviors of the cicadas are not as yet fully characterized,[8] an upcoming emergence of the insects will provide an opportunity to test and report on the tymbal and abdomen structural dynamics that generate the cicada's mating call. Measuring the cicada's two most important anatomic structures for producing sound would add to the scientific understanding of the extent of the nonlinear nature of the cicada signals and would also help to explain the high sound levels produced by this small insect.

C. Current state of cicada research

A comprehensive review of the prominent journal articles on cicada sound production and mechanisms[13–19] and a survey of a number of subject area expert textbooks in this field[20–28] revealed that the mechanisms underlying the cicada's sound levels and efficient sound propagation are not

[a]Author to whom correspondence should be addressed. Electronic mail: derke.hughes@navy.mil
[b]Retired.

fully understood. The cicada song has been classically modeled using linear mathematical methods. However, these linear methods are insufficient for a true model of the system because the buckling tymbals within the cicada sound production system are essential to the acoustic level and propagation of this mating call. Inelastic buckling is well recognized as a nonlinear phenomenon among researchers.[29]

II. QUANTIFYING NONLINEARITY IN CICADA SIGNALS

A. Technical approach

A signal processing repertoire includes methods to test for (a) Gaussianity, (b) non-Gaussianity, (c) linearity, and (d) nonlinearity. The basic analytical tools used to perform these tests are the temporal power spectrum, the average mutual information (Mi) (i.e., an information theory technique), and surrogate data hypothesis testing (i.e., phase randomization). A preliminary determination of the degree and influence of these effects in the sound production system of the cicada is explored using the Volterra expansion.

B. Nonlinear signal processing with mutual information

In order to substantiate the existence of nonlinearity, a quantitative method must be established. This study used Mi (Ref. 30) and a surrogate data method to confirm nonlinearity. Equation (1), which defines the general Mi, is a probabilistic equation used quantitatively to assess information between two random variables A and B.

$$I_{A,B} = \sum_{i,j} P_{AB}(a_i,b_j)\log_2\left[\frac{P_{AB}(a_i,b_j)}{P_A(a_i)P_B(b_j)}\right], \tag{1}$$

where $P_{AB}(a_i,b_j)$ is the joint probability of events from sets $A=\{a_i\}$ and $B=\{b_j\}$, and $P_A(a_i)$ and $P_B(b_j)$ are the marginal individual probabilities associated with sets A and B, respectively. For example, set B can be taken as a collection of events that are time-delayed versions of the events in set A, which is the case for this research.

The surrogate data method consists of randomizing the phase of a signal spectrum as shown in

$$\begin{aligned}\mathbf{s}(n) &= \sum_{k=0}^{N-1} x(k)e^{-i2\pi nk/N} \quad \text{for } n=0{:}N-1.\\ \mathbf{S}(k) &= \sum_{n=0}^{N/2-1} \mathbf{s}(n)e^{i[\boldsymbol{\varphi}(n)+2\pi nk/N]}\\ &\quad + \sum_{n=N/2}^{N-1} \mathbf{s}(n)e^{i[-\boldsymbol{\varphi}(n)+2\pi nk/N]} \quad \text{for } k=0{:}N-1.\end{aligned} \tag{2}$$

$\mathbf{S}(k)$ is the overall phase-randomized discrete Fourier transform (DFT) of signal $\mathbf{x}(k)$, where phases $\{\boldsymbol{\varphi}(n)\}$ are independent, uniform, and randomly distributed over a 2π range. $\mathbf{s}(n)$ is the complex amplitude spectrum of the DFT of the original time series, which is altered by a random phase $\boldsymbol{\varphi}(n)$. In the surrogate method, the inverse DFT, $\mathbf{S}(k)$, is calculated from $\mathbf{s}(n)$, and this transformed time series is called the surrogate data and used for a comparison with the original signal $x(k)$. Nonlinearity exists if the randomized signal's Mi diverges from the original signal's Mi, thereby signifying that nonlinearity must be present in the time series.[31]

Linear calculations are plotted with the nonlinear signal results in order to display the contrast between linearity and nonlinearity. The Gaussian rule in Eq. (3) is presented to indicate how the correlation coefficient and Mi are linked for the special case of a pair of Gaussian random variables with normalized correlation coefficient ρ:[32]

$$I_{A,B} = -\tfrac{1}{2}\log_2(1-\rho^2). \tag{3}$$

Note that the Mi $I_{A,B}$ is 0 when the correlation coefficient ρ is 0, while the Mi goes to infinity as the correlation coefficient goes to positive or negative 1. This holds for all joint Gaussian processes, which are considered linear processes. The frequency domain equivalent of the correlation coefficient for measuring linearity is the subject of many published papers on coherence in the 1993 reprint text by Carter.[33] Papers on coherence include how to estimate coherence, how to minimize bias and variance, and how to determine confidence bounds for estimates of magnitude-squared coherence. In general, for stationary random process, proper averaging of large time segment improves estimation.

C. Higher-order spectral techniques using magnitude-squared bicoherence

Furthermore, multispectral techniques exist, such as bicoherence and tricoherence, which could provide additional understanding of a non-Gaussian process. If the cicada signals are non-Gaussian, the bicoherence could determine if a process is a mixed-phased process or a nonlinear process. Cumulants are higher-order statistical information about any data series. For example, the first-order cumulant for a stationary process is its mean value. The second-order cumulant for a zero-lag process is the covariance, the third is skewness, and the fourth is kurtosis. As the first step beyond first-order spectral analysis, this study analyzes the bicoherence. Equation (4) is the magnitude-squared bicoherence (MSB), which is used to provide evidence on whether a data sequence is linear or nonlinear:

$$B(f_1,f_2) = \frac{|S_2(f_1,f_2)|^2}{S(f_1)S(f_2)S(f_1+f_2)}. \tag{4}$$

D. Nonlinear signal processing using a Volterra expansion

A nonlinear fit is performed on the cicada data by means of the Volterra expansion, which describes the first-order and second-order signal dynamics present within the cicada time series. The data acquired on the cicada consist of two laser measurements, denoted by sequences $\{\mathbf{a}(n)\}$ and $\{\mathbf{b}(n)\}$, as well as a simultaneously sampled microphone sequence $\{\mathbf{z}(n)\}$. This situation will be considered to be a two-input, one-output, nonlinear "cicada system" with mathematical memory, instead of the traditional system of one input and one output.[34] This formulation lends itself to a Volterra ex-

pansion, which can be used to determine, quantitatively, the extent of nonlinearity present between the two inputs and the one output.

1. Second-order Volterra formulation

The standard Volterra expansion for a one-input system takes the form

$$\mathbf{y}(n) = \mathbf{h}_0 + \sum_{k=0}^{K_1-1} \mathbf{h}_1(k)\mathbf{a}(n-k) + \sum_{k=0}^{K_2-1}\sum_{j=k}^{K_2-1} \mathbf{h}_2(k,j)\mathbf{a}(n-k)\mathbf{a}(n-j), \tag{5}$$

when carried to the second order, where $\mathbf{y}(n)$ is the Volterrra fit. The three functions $\mathbf{h}_0, \mathbf{h}_1(k)$, and $\mathbf{h}_2(k,j)$ are the zeroth-order, first-order, and second-order kernels, respectively. The first-order terms are carried out to length K_1, while the second-order terms are carried out to length K_2 by K_2. Because of the symmetry inherent to $\mathbf{h}_2$ in Eq. (5), the summation index j can be limited to value k and above. The unknown kernels appear *linearly* in model Eq. (5), whereas the *known* excitation $\{\mathbf{a}(n)\}$ appears nonlinearly through a product of delayed versions.

With a two-input model, a generalization is necessary, namely,

$$\mathbf{y}(n) = \mathbf{y}_0 + \mathbf{y}_1(n) + \mathbf{y}_2(n), \tag{6}$$

where components

$$\mathbf{y}_1(n) = \sum_{k=0}^{K_1-1} \mathbf{h}_\mathbf{a}(k)\mathbf{a}(n-k) + \sum_{k=0}^{K_1-1} \mathbf{h}_\mathbf{b}(k)\mathbf{b}(n-k), \tag{7}$$

$$\mathbf{y}_2(n) = \sum_{k=0}^{K_2-1}\sum_{j=k}^{K_2-1} \mathbf{h}_\mathbf{aa}(k,j)\mathbf{a}(n-k)\mathbf{a}(n-j) + \sum_{k=0}^{K_2-1}\sum_{j=k}^{K_2-1} \mathbf{h}_\mathbf{bb}(k,j)\mathbf{b}(n-k)\mathbf{b}(n-j) + \sum_{k=0}^{K_2-1}\sum_{j=0}^{K_2-1} \mathbf{h}_\mathbf{ab}(k,j)\mathbf{a}(n-k)\mathbf{b}(n-j).$$

There are two linear components in $\{\mathbf{y}_1(n)\}$ each of length K_1, and three nonlinear (second-order) components in model output $\{\mathbf{y}_2(n)\}$. The advantage of the inherent symmetry in the two auto components in $\mathbf{y}_2(n)$ reduces the number of kernel values that have to be determined. However, the cross component $\mathbf{h}_\mathbf{ab}(k,j)$ in $\mathbf{y}_2(n)$ has no such symmetry and therefore requires a full K_2 by K_2 expansion. The total number of unknown kernel coefficients in Eqs. (6) and (7) is

$$K = 1 + 2K_1 + K_2(2K_2 + 1). \tag{8}$$

It is desired to choose these coefficients $\mathbf{h}$ so that the *total* model output, Eq. (6), fits the measured microphone output $\mathbf{z}(n)$ as well as possible using least squares, so that $\mathbf{y}(n) \approx \mathbf{z}(n)$. The least-squares approach is adopted because the simultaneous equations for the optimum kernel coefficients will then all be linear.

Although the laser and microphone data have been sampled at frequency f_s=96 kHz, the crucial frequency content of the cicada system itself is not believed to extend above 10 kHz. Therefore, reductions in the memory lengths K_1 and K_2 for the first-order and second-order kernels in the model will not alias the cicada statistical information. Also, the memory lengths are decreased to minimize the computer random access memory required to calculate Eq. (6). Consequently, a decimation factor of M is applied to the kernels. Thus, the model to be fitted is a modification of Eq. (7), namely,

$$\mathbf{y}_1(n) = \sum_{k=0}^{K_1-1} \mathbf{h}_\mathbf{a}(k)\mathbf{a}(n-Mk) + \sum_{k=0}^{K_1-1} \mathbf{h}_\mathbf{b}(k)\mathbf{b}(n-Mk) \equiv \mathbf{y}_\mathbf{a}(n) + \mathbf{y}_\mathbf{b}(n), \tag{9}$$

$$\mathbf{y}_2(n) = \sum_{k=0}^{K_2-1}\sum_{j=k}^{K_2-1} \mathbf{h}_\mathbf{aa}(k,j)\mathbf{a}(n-Mk)\mathbf{a}(n-Mj) + \sum_{k=0}^{K_2-1}\sum_{j=k}^{K_2-1} \mathbf{h}_\mathbf{bb}(k,j)\mathbf{b}(n-Mk)\mathbf{b}(n-Mj) + \sum_{k=0}^{K_2-1}\sum_{j=0}^{K_2-1} \mathbf{h}_\mathbf{ab}(k,j)\mathbf{a}(n-Mk)\mathbf{b}(n-Mj) \equiv \mathbf{y}_\mathbf{aa}(n) + \mathbf{y}_\mathbf{bb}(n) + \mathbf{y}_\mathbf{ab}(n).$$

By this means, the memory length of the first-order kernels is MK_1 units of the sampling increment $1/f_s$, while that of the second-order kernels is MK_2 units. Notice that the measured data $\mathbf{a}(n)$, $\mathbf{b}(n)$, and $\mathbf{z}(n)$ are *not* decimated, thereby retaining any harmonic and intermodulation products that might have been created by the cicada system itself.

2. Least-squares considerations

The details of a first-order fitting procedure will be presented; this formulation can then be extended to include all the terms in Eq. (9). The pertinent equation that governs the least-squares approach is to make

$$\mathbf{y}_\mathbf{a}(n) = \sum_{k=0}^{K_1-1} \mathbf{h}_\mathbf{a}(k)\mathbf{a}(n-Mk) = \mathbf{h}_\mathbf{a}(0)\mathbf{a}(n) + \mathbf{h}_\mathbf{a}(1)\mathbf{a}(n-M) + \cdots + \mathbf{h}_\mathbf{a}(K_1-1)\mathbf{a}(n-M(K_1-1)) \tag{10}$$

approximate $\mathbf{z}(n)$ for

$$N_1 \le n \le N_t, \quad N_1 \equiv M(K_1-1)+1, \tag{11}$$

where N_t is the common data length of the three available data sequences. The particular starting value N_1 for n arises so that the inherent buildup transient of the first-order kernel $\mathbf{h}_\mathbf{a}(k)$ will be excluded from the fitting procedure using Eq. (10). Trying to fit the transient can only degrade the procedure; confining the error minimization to the steady-state model output is the best approach.

Equations (7) and (9) can be put into a matrix formulation as follows:

$$\begin{bmatrix} a(N_1) & a(N_1-M) & \cdots & a(1) \\ a(N_1+1) & & & a(2) \\ \vdots & & & \vdots \\ a(N_t) & a(N_t-M) & \cdots & a(N_t-N_1+1) \end{bmatrix} \times \begin{bmatrix} h_a(0) \\ h_a(1) \\ \vdots \\ h_a(K_1-1) \end{bmatrix} \sim \begin{bmatrix} z(N_1) \\ z(N_1+1) \\ \vdots \\ z(N_t) \end{bmatrix}. \quad (12)$$

In matrix notation, this equation reads

$$\mathbf{Dh} \sim \mathbf{Z}, \quad (13)$$

where matrix $\mathbf{D}$ is $(N_t-N_1+1)\times K_1$, column vector $\mathbf{h}$ is $K_1 \times 1$, and column vector $\mathbf{Z}$ is $(N_t-N_1+1)\times 1$. Since data length N_t is generally a very large number, whereas the number K_1 of kernel coefficients is usually small, the attempted fit in Eqs. (12) and (13) cannot be achieved exactly. If an error sequence is defined as the difference between the left-hand and right-hand sides of Eq. (13), and the sum of squared errors is minimized, it can be shown that the optimum kernel $\mathbf{h_o}$ satisfies the equations

$$\mathbf{D'Dh_o} = \mathbf{D'Z}, \quad \mathbf{h_o} = (\mathbf{D'D})^{-1}\mathbf{D'Z} = (\mathbf{D'D}) \setminus (\mathbf{D'Z}). \quad (14)$$

Observe that both sides of Eq. (13) are premultiplied by $\mathbf{D'}$ and the approximation is replaced by an equals sign.

In MATLAB notation, the least-squares solution to Eq. (13) is obtained according to

$$\mathbf{h_o} = \mathbf{D} \setminus \mathbf{Z}. \quad (15)$$

The approach in Eq. (15) instead of Eq. (14) is more advantageous, in that the condition number of square matrix $\mathbf{D'D}$ is the square of the condition number of D, which makes approach Eq. (14) less reliable.

III. CICADA EXPERIMENTS

A. Cicada field experiment with microphone

Measured data from the Tibicen chloromera and Tibicen lyricen species were recorded in the field with a parabolic microphone and a Marantz PMD670 behind a high school in West Hartford, CT. Thus, some noise from crickets and other environmental sounds contaminated the signals, but this noise was about 40–60 dB below the cicada's distinct call. Also, only distinctive, strong cicada calls—not low-level idling songs—were used to determine the existence of nonlinear behavior in the signal.

B. Cicada field experiment with laser and microphone

Another field experiment was conducted on Magicicada septendecim and Magicicada cassini in Julibee State College Park in Peoria, IL by Hughes and Katz as well as the acknowledged support. The test site was specifically chosen to amass live specimens during the 17 year emergent cycle of periodical cicadas in brood XIII. In this test, the tymbal motion of the cicada was measured by laser while the acoustic output was recorded via a microphone. This simultaneous measurement was made with dual Polytec OFV-508 optical measurement heads controlled by the Polytec electronic signal processor OFV-2802. The dual optical sensor heads allowed simultaneous measurement of the motion of both tymbals and the tymbal-to-abdomen motion. Meanwhile, the parabolic microphone was used to continuously record either the output of the two tymbals or a tymbal-abdomen experimental setup. The measurements provide the quantitative velocity from the laser, which can be transformed into displacement (position versus time) information. The tymbal motion is proportional to acoustic mechanical vibration and thus can be used in conjunction with the microphone acoustic output data. Thus, the data gathered in this field experiment, in which the output and input signals were obtained *simultaneously*, allows a unique opportunity to gain further scientific understanding of the cicada sound production system. The input signal is from the tymbal, and the output signal is recorded via the microphone. These data sets allow the application of nonlinear mathematical techniques, such as the Volterra expansion, to real-world signals.

Two channels were designated for collecting dual laser measurements of the motion of both tymbals and the tymbal-abdomen simultaneously. Most collections consisted of live wingless insects, otherwise intact. A third channel was used to record the microphone output. A fourth channel was used to time tag all signals. Two multichannel recorders were used to record and back up the data.

The experimental setup used a multichannel digital recorder set to a 96 kHz sampling rate, a laser and its controller, a backup digital recorder, and a digital timekeeper. The digital recorder facilitated the simultaneous acquisition of the tymbal motion, a microphone, and time stamp. One advantage afforded by this experimental arrangement was the opportunity to analyze quantitatively both synchronous and asynchronous tymbal motion and how such motion might impact the vocalization output.

IV. NONLINEAR SIGNAL PROCESSING RESULTS

A. Nonlinear and linear modeling of cicada signal results

A linear representation of a simulated time series of the cicada vocalization based on the superposition of a Gaussian white-noise signal passed through three parallel independent narrowband filters and the power spectral representation of this simulated time series representation is shown in Fig. 1(a). Figure 1(b) shows the temporal power spectrum of the cicada vocalization measured in the field. Given the striking similarity between these two spectra, the immediate, but erroneous, conclusion might be that the modeled spectra are a good representation of the insect's actual acoustic signature. This is the classical error made by analysts using strictly linear techniques to model behavior in a physical system. In fact, linear techniques are insufficient for a correct analysis of the underlying signal processing mechanics of the cicada's sound production system.

The simulated signal's Mi (i.e., white Gaussian excitation passed through three parallel narrowband filters) and the Gaussian Mi model of the simulated signal [from Eq. (3)] are

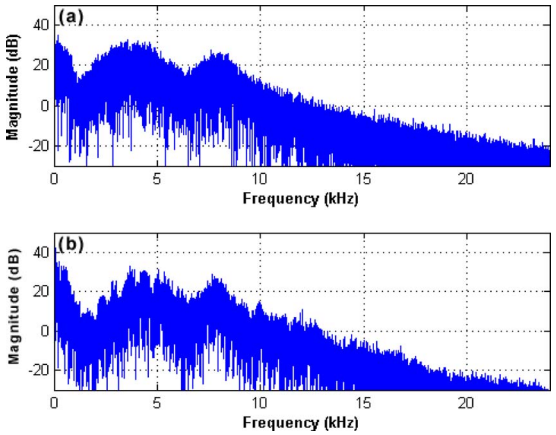


FIG. 1. (Color online) Power spectra of simulation vs field measurement. (a) Power spectrum of simulated cicada call estimated with linear model. (b) Temporal power spectrum of cicada call as measured in the field. The spectra appear similar, but the linear model does not accurately represent the measured acoustic signature.

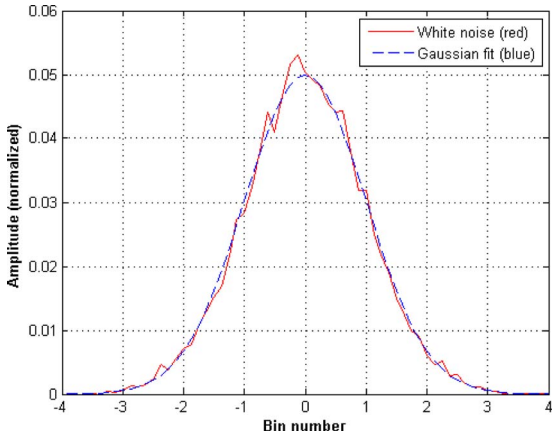


FIG. 3. (Color online) The Gaussian curve fit to the probability density function (dashed line) is very similar to the simulated data curve based on its histogram (solid line)—textbook example of results for a linear Gaussian process and baseline for the field measurements of cicada signals.

compared, which produce an identical plot to Fig. 2 where the solid line is the Mi and the "×" or cross symbol indicates the Gaussian rule. However, Fig. 2 compares a plot of the Mi obtained from the non-phase-randomized simulated date with the phase-randomized simulated data. As expected, the phase-randomized signal and the non-phase-randomized signal Mi generate the same values since phase randomization implies near Gaussianity (according to the central limit theorem), which in turn implies linearity. In Fig. 3, the Gaussian curve fit to the probability density function (PDF) is very similar to the simulated data curve based on its histogram. These graphs are like textbook examples of what the results should be for a linear Gaussian process. The graphs also serve as a baseline with which to compare the field measurements of cicada signals. A field recording of the cicada signal is displayed in Fig. 4(a), with a zoomed-in version of the same signal in Fig. 4(b). The expanded view illustrates the complexity of the cicada call. In Fig. 5, the cicada data (as Mi of the cicada vocalization) are compared with the surrogate data (the phase-randomized cicada data) to show how the phase-randomized cicada signal approaches the Gaussian rule. The separation between Mi plots of the cicada signal and the phase-randomized version of the same signal corroborate its non-Gaussian probability distribution. Also, the separation between the Mi of the cicada vocalization and the simulated data using the Gaussian rule is identical to Fig. 5, which is consistent with strong non-Gaussianity [i.e., Eq. (3)].

Additional evidence of the non-Gaussian behavior of the cicada vocalization is illustrated in Fig. 6. This figure compares the Gaussian fit with the curve based on the histogram associated with the cicada's non-Gaussian signal. Note that the non-Gaussian behavior in the cicada call was not reflected in the modeled spectra shown earlier in Fig. 1. Thus, Fig. 6 reinforces the point made earlier against relying solely on power spectral techniques to analyze cicada time waveforms.

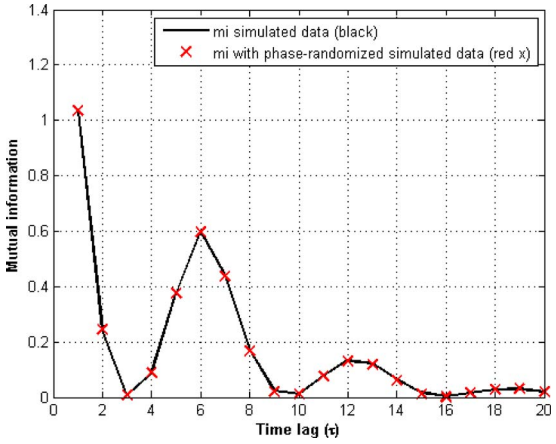


FIG. 2. (Color online) Mi obtained from the phase-randomized simulated signal as a function of the time lag τ (×) compared with the non-phase-randomized simulated data (solid line plot). The marker (×) is indistinguishable from the solid line, as expected, since phase randomization implies near Gaussianity (according to the central limit theorem), which in turn implies linearity. If plotted, the simulated signal's Mi—white Gaussian excitation passed through three narrowband filters—as a function of the time lag τ (solid line) is compared with the Gaussian rule model of the simulated signal from Eq. (3) (×). The plots are identical to what is shown.

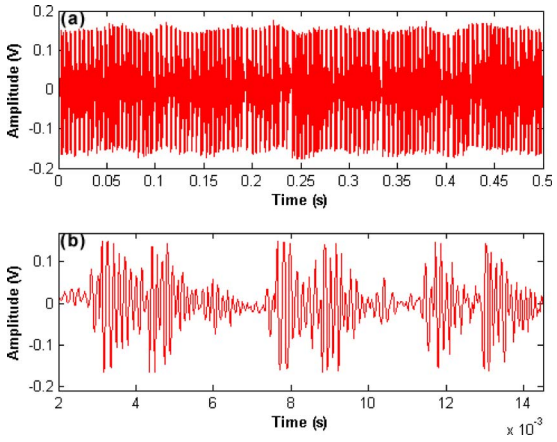


FIG. 4. (Color online) Time series of measured cicada vocalization: a 0.5 s field recording (a) and an expanded segment illustrating the complexity of the cicada call (b).

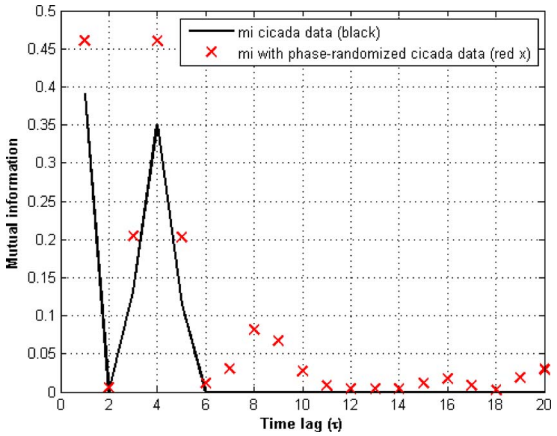


FIG. 5. (Color online) Mi of the cicada vocalization as a function of the time lag τ (solid line) compared with the surrogate data, i.e., Mi of the phase-randomized cicada data (×). The separation between Mi plots of the cicada signal and the phase-randomized version of the same signal confirm its non-Gaussian probability distribution. If plotted, the cicada signal's Mi as a function of the time lag τ (solid line) is compared with the Gaussian rule model of the signal from Eq. (3). (×) is identical to this plot.

B. Results of higher-order spectral technique using MSB

The results from Eq. (4) are shown in this section of the report. Namely, the bicoherence indicates whether the time series data is linear or nonlinear. The theoretical MSB $B(f_1,f_2)$ of a linear process is flat throughout the f_1-f_2 plane. Therefore, the example shown in Fig. 7 with an exact solution is utilized to ascertain if the bicoherence will always predict a flat surface for this linear process, where the center frequency (f_c) is zero. Here is a zero-mean, unit-variance Gaussian process. Consider the entire expression of Eq. (4) with the denominator terms containing the individual frequencies as well as the addition of both frequencies. Figure 8 illustrates that the denominator terms do not completely flatten the conical peak when the numerator (bispectrum) is divided by the denominator shown in Eq. (4). Note that the MSB plot is not flat suggesting that the process is not linear. Figure 9 illustrates that the PDF does have an effect on the random amplitudes generated in the bicoherence computation. However, the amplitudes of the sample MSB in Fig. 9 also suggest that averaging may help reduce the peaks and valleys of the plots. Figure 10 confirms that averaging does reduce the amplitude variation of the sample MSB. Such observations are consistent with the bias and variance reduction observed in coherence estimation.[33] Nonetheless, the amplitude variation is not completely removed from the estimate. Hence, averaging the MSB only scales down the problem of a fluctuating surface. Consequently, a flatness factor could yield a meaningful parameter to employ on the variations calculated for the sample MSB.

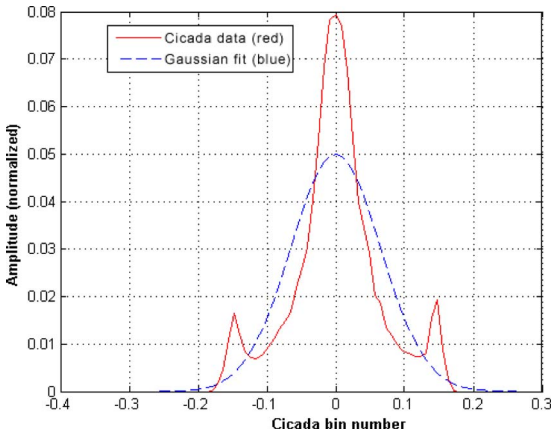


FIG. 6. (Color online) The PDF of the cicada signal (solid line, multiple peaks) compared with a Gaussian curve fit to the PDF (dashed line, single peak). The separation of the plots is additional evidence of the non-Gaussian behavior of the cicada signal.

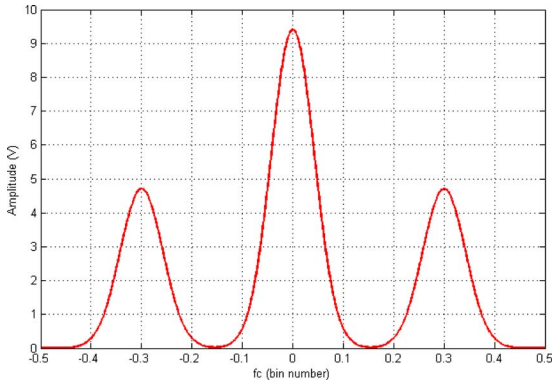


FIG. 7. (Color online) The spectrum of nonlinear function $g(t)^2-1$ provides an opportunity to assess the capability of the MSB to evaluate nonlinearity.

$$\bar{x} = \frac{1}{N}\sum_{i=1}^{N} x_i,$$

$$\sigma = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(x_i-\bar{x})^2}, \tag{16}$$

$$\sigma_{\mathrm{MSB}} = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(\sigma_1-\sigma_2)^2}.$$

Equation (16) defines the "flatness factor," which derives from the sample standard deviation of the fluctuations in the f_1-f_2 plane. The variable σ represents the calculation of the sample standard deviation for a known sample set $\{x_i\}$ with a total sample size N, while the parameter $\bar{x}$ is the sample

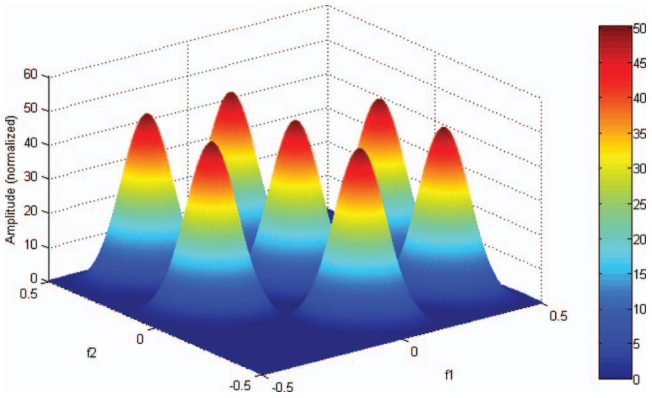


FIG. 8. The amplitude of the MSB for $g(t)^2-1$ is large despite the denominator attempting to scale the numerator (bispectrum).

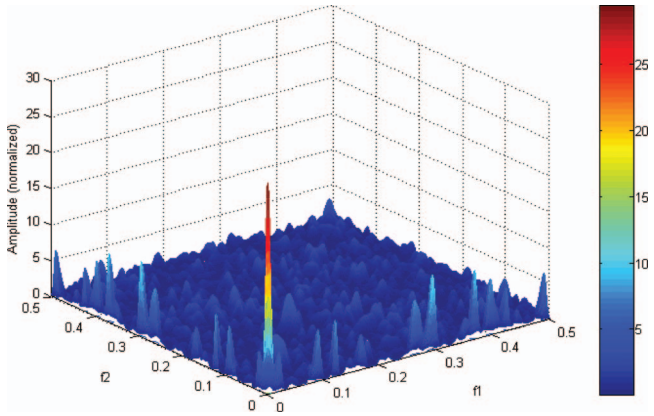


FIG. 9. The MSB is calculated for a normalized random noise signal and the mesh plot exposes that surface undulations are present in the f_1-f_2 plane, which would give the false conclusion that Gaussian random noise is not a linear process.

mean of the data set. Therefore, the sample standard deviation σ_{MSB} represents the variation for the entire plane.

Since the PDF of a given data set affects the standard derivation, a determination must be made to show how the PDF affects the MSB. A cumulative distribution function (CDF) is a convenient approach to determine if a PDF has Gaussian characteristics, which is the case for the exponential and normal distributions. If the exponential and the normal CDFs differ, the implication is that the PDF influences the MSB significantly. In Fig. 11, the CDF for the exponential signal is shown. Compared to Fig. 12, there is a difference in value of the accumulation of bins in the MSB for the exponential and normal distributions of 0.5–0.95, respectively. This difference is considerable and thus indicates that the PDF of the data contributes notably to the MSB. Consequently, a prerequisite algorithm must account for the effect of the PDF on the MSB, and PDFs could vary between individual cicadas and even more so with different species. Because the MSB significantly depends on the PDF, additional work beyond the scope of the present research is required for quantifying how statistically useful the sample MSB is for analyzing cicada signals.[35,36]

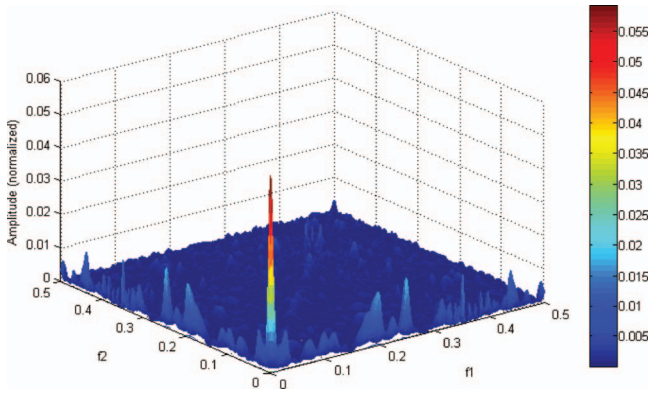


FIG. 10. The mesh plot of MSB for a normalized random noise averaged over 1000 trials still has ripples on the f_1-f_2 plane. However, this averaging scales the fluctuating surface but determining nonlinearity remains difficult to quantify.

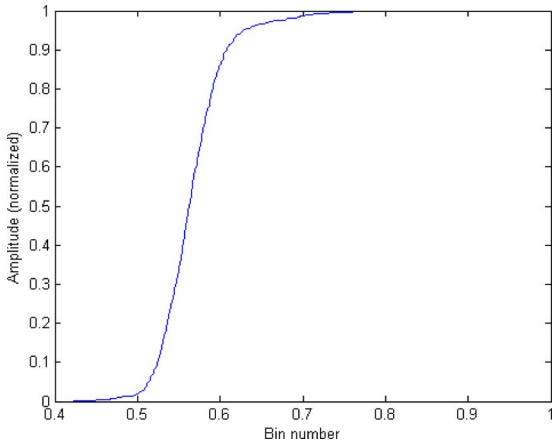


FIG. 11. (Color online) A CDF for an exponential random noise averaged over 1000 trials indicates the Gaussian characteristics of the MSB.

C. Volterra expansion method applied to cicada experimental results

A sample of 100,000 data points from the laser and microphone measurements is selected for fitting purposes, using a decimation factor $M=4$. The number of coefficients employed for the first-order fit is $K_1=100$, and the number used for the second order is $K_2=50$, which results in solving for a total of $K=5251$ coefficients. The execution time required simply to fill the $\mathbf{D}'\mathbf{D}$ matrix, using segment length $L=1000$, is 1670 s on a 2.4 GHz computer. The solution time for the optimum kernel is 22 s, and the time required to compute all five individual component waveforms $\mathbf{y_a}(n)$, $\mathbf{y_b}(n)$, $\mathbf{y_{aa}}(n)$, $\mathbf{y_{bb}}(n)$, and $\mathbf{y_{ab}}(n)$ in Eq. (9) is 220 s. The singular value decomposition of matrix $\mathbf{D}'\mathbf{D}$ took 720 s, and its condition number is 7.2×10^6. Thus, approximately 8 decimal digits (out of 15) of significance remain in the numerical results obtained. The total execution time is almost 44 min. The ratio of the power in the *total* model output, per Eq. (6), to the power in the measured microphone output $\mathbf{z}(n)$ is 0.41. Thus, the Volterra fitting procedure of the second order is capable of representing 41% of the power of the microphone waveform.

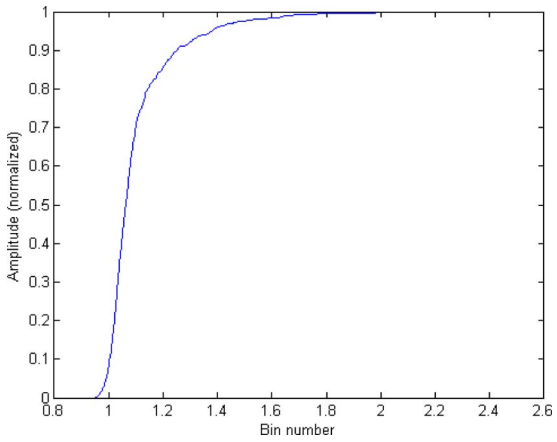


FIG. 12. (Color online) CDF for a normal random noise averaged over 1000 trials is also Gaussian but has a significant different PDF, which affects the MSB when compared to the exponential random noise in Fig. 11.

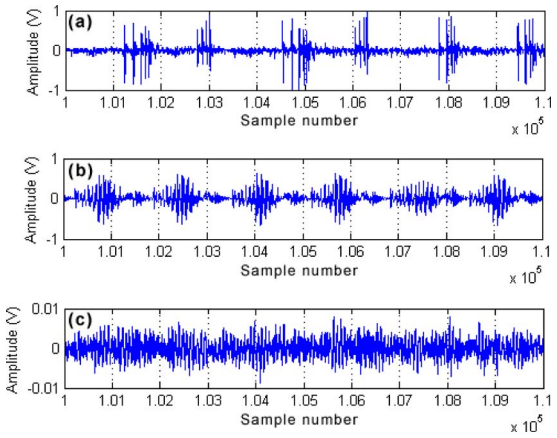


FIG. 13. (Color online) Short time segments of simultaneous laser measurements of the motion of the two tymbals and the sound output at the microphone in the field are shown. The repeated bursts of narrowband energy are clear in the laser A and B measurements in (a) and (b), respectively. The microphone output (c) tends to capture environmental noise, which blurs the individual bursts seen in (a) and (b).

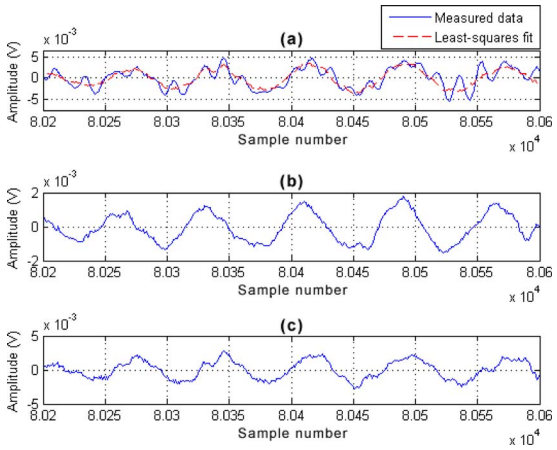


FIG. 15. (Color online) Least-squares fit for Volterra expansion. (a) A segment of the microphone data (solid line) [see Fig. 13(c)] is compared with the total fitted waveform (dashed line). The fit is rather good in some time intervals, but poorer in others—a manifestation of the fitted power ratio of 0.41. (b) and (c) are the individual total first- and second-order fits, respectively.

As a check case, an unrelated random white-noise process replaces the microphone output $\mathbf{z}(n)$, and the fitting procedure is repeated with identical parameters. The power ratio of the final fit is reduced to 0.053. An additional run with a different random sequence for $\mathbf{z}(n)$ yields a comparable value for the power ratio. Thus, the fitting ratio 0.41 that is actually attained is a significant value and indicates that nonlinearities are present in the cicada system between laser inputs and microphone output. In fact, the power in the second-order component $\mathbf{y}_2(n)$ in Eq. (9) is almost three times greater than the power in the linear component $\mathbf{y}_1(n)$. Also, the power in the cross component $\mathbf{y}_{\mathbf{ab}}(n)$ is *greater* than the powers in the two autocomponents $\mathbf{y}_{\mathbf{aa}}(n)$ and $\mathbf{y}_{\mathbf{bb}}(n)$.

D. Volterra graphical results

A short time segment of the three simultaneous field measurements is displayed in Fig 13. Each of the lasers—A and B—measured the narrowband buckling of a single cicada tymbal, shown in Figs. 13(a) and 13(b), respectively, while the microphone [see Fig. 13(c)] tended to capture all the acoustic noise in that segment of the field measurement. The corresponding (unsmoothed) spectral estimates of the complete data segments used in the Volterra fit are plotted in Fig. 14. A segment of the microphone data $\mathbf{z}(n)$ is compared with the total fitted waveform $\mathbf{y}(n)$ in Fig. 15(a). The fit is rather good in some time intervals but poorer in others—a manifestation of the actual fitted power ratio of 0.41. The plots in Figs. 15(b) and 15(c) show the individual total first-order and total second-order fits, respectively. The two first-order time-domain kernels are plotted in Fig. 16, while their complex frequency-domain transfer functions are displayed in Fig. 17. The decimation factor $M=4$ is the reason for the upper frequency limit of 12 kHz for these kernels.

An alterative Volterra expansion is computed that measures the nonlinearity in the cicada's mating call. Table I describes each tymbal's contribution to the first-and-second-order components in the cicada signal by using a Volterra expansion with $M=3$ and $K_2=70$. The Volterra technique

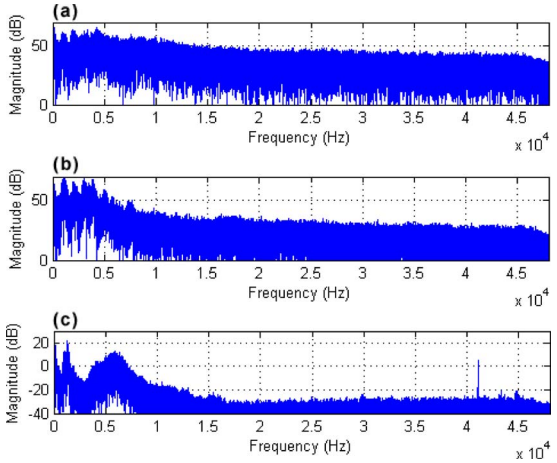


FIG. 14. (Color online) Unsmoothed spectral estimates of the data measured at lasers A (a) and B (b) and at the microphone (c). These complete data segments, corresponding to both tymbal and microphone outputs in Fig. 13, are used in the Volterra fit.

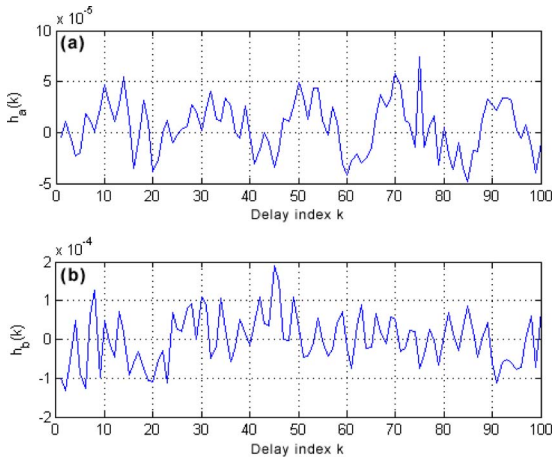


FIG. 16. (Color online) First-order time-domain kernels of tymbal motion measured by lasers A (a) and B (b).

requires both output and input functions: in this case, microphone data **zz** are the output, and the input functions are laser A and laser B. Variable **y** in Table I represents the Volterra expansion solution. The lowercase a and b for the Volterra expansion correspond to the measured vibration from the two tymbals. The first-order and second-order Volterra solutions are $\mathbf{y}_1$ and $\mathbf{y}_2$, respectively. Therefore, $\mathbf{y_a}$ is the first-order solution for laser A, $\mathbf{y_b}$ is the first-order solution for laser B, and $\mathbf{y_{aa}}$ represents the second-order solution for laser A, and so forth. The combination of both tymbals $\mathbf{y_{ab}}$ is a mixed second-order component that contributes approximately 36% of the second-order Volterra solution. This finding suggests that parametric generation may contribute to the cicada's sound propagation.[37] Also, the second-order solution contains 87% of the Volterra solution, which also indicates significant nonlinearity in the cicada call.

There are two reasons that the fitting fraction is not greater than 0.41 or 0.36 in the cases above. The common length $K_1=100$ of the two first-order kernels $\mathbf{h_a}(k)$ and $\mathbf{h_b}(k)$ is apparently adequate because both of the estimated first-order kernels decayed essentially to zero at both ends of the memory interval of length MK_1. Because the microphone data contained a considerable amount of background noise, the common length $K_2=50$ of the three second-order kernels $\mathbf{h_{aa}}(k,j)$, $\mathbf{h_{bb}}(k,j)$, and $\mathbf{h_{ab}}(k,j)$ is inadequate because the estimate does not decay sufficiently by the edges of the $k-j$ plane. The common length $K_2=70$ did reduce the decay in the kernel, but still more delay is required.

E. Implications of cicada signal results for underwater acoustics

In the commercial world, small active "fish-finding" sonars with wristband receivers are sold at a modest price to sports fishermen. Recent scientific advances in the study of small insects generating loud acoustics in air, by Bennet-Clark,[38] suggest opportunities to transfer this technology to small-sized active sonar applications. The cicada's tymbal mechanics include a tymbal plate, four ribs and resilin pad. "The four ribs of the Tymbal buckle inwards from posterior to anterior…. A train of four sound pulses [was produced], each corresponding to the inward buckling on one rib…."[38] The difference in the frequencies "suggests that the mass-to-stiffness ratio that determines the resonant frequencies of the various pulses differs from pulse to pulse…. Pulses produced later in the inward buckling sequence were less affected by the loading than earlier ones. This suggests that the effective mass determining the resonance in the later pulses is greater than that in the earlier pulses…. The Tymbal appears to act as an energy storage mechanism that releases energy as the tymbal ribs buckle inwards in sequence…. It [also] appears that the central part of each rib is decoupled from its predecessor as it buckles inwards…. observations suggest that the vibration is initially non-linear when the amplitude of vibration is largest but that it becomes more linear as the pulse decays."[38] In certain regions of interest, Bennet-Clark suggests that the resonance "is determined by the simple interaction between linear mass, compliance and damping elements."[38] Moreover, Young and Bennet-Clark suggest "that the inward movement of the tymbal plate and rib 1 will be constrained by the resilin sheet that couple it to the next, unbuckled, rib."[39] Bennet-Clark states that the "change in relative resonance frequency implies the ratio between the mass and the stiffness of the vibrating system alters as successive ribs buckle."[38] Young and Bennet-Clark go on to observe that "analogous mechanisms for producing long coherent waveforms by a succession of impulses have been modeled in bush crickets."[39] Part of the cicada's sound generation relies on the rubber-like resilin material which "acts as an energy store in which muscle work is stored comparatively slowly (perhaps during the first 2 to 4 ms of contraction) and is then released rapidly by the sudden buckling of the tymbal ribs over a period of one cycle or 230 microsec."[39] Note that 230 μs is 0.23 ms, and, thus, for 2.3 ms (between 2 and 4 ms), the discharge occurs in nominally ten times less time. This detailed description of the cidada's storage and release mechanism provides useful insight for construction of an underwater acoustic source.

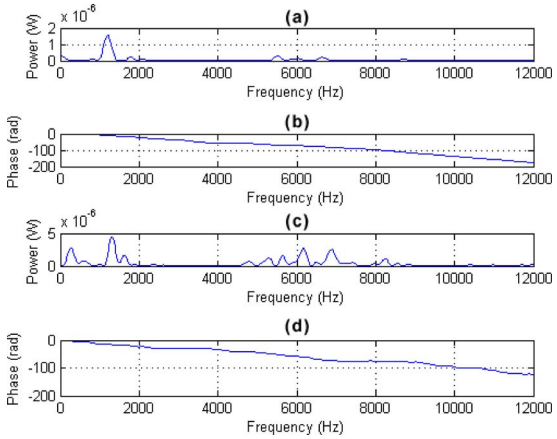


FIG. 17. (Color online) Transfer functions of first-order time-domain kernels of tymbal motion: power (a) and phase (b) for laser A data, and power (c) and phase (d) for laser B data.

V. CONCLUSIONS

This research utilizing advanced signal processing techniques to support that the cicada's loud mating call is produced as a non-Gaussian, nonlinear vocalization. Calculating the Mi to determine nonlinearity is often difficult because of the many degrees of freedom (i.e., large sample size) required to obtain a valid joint probability function [Eq. (1)]. To circumvent size limitations on the data sample, the Volterra expansion was used to discover that the second-order kernel possesses enough of the energy in the cicada signal to

TABLE I. Volterra expansion to quantify cicada signal nonlinearity for laser A (a), laser B (b), and the microphone (zz).

Ya	Yb	Yaa	ybb	yab	Y1	Y2	Y	zz	Error
6.4677×10^{-8}	3.2090×10^{-7}	2.5461×10^{-7}	1.0602×10^{-6}	6.0081×10^{-7}	3.9690×10^{-7}	1.6650×10^{-6}	1.9100×10^{-6}	4.8180×10^{-6}	2.9090×10^{-6}

indicate nonlinear behavior. Although the curse of dimensionality pertains to the Volterra expansion, the current second-order solution is satisfactory for initiating the quantification of the nonlinear parameters for modeling and simulating purposes.

Mi and surrogate hypothesis testing techniques can be combined with the Volterra or Wiener expansion to parametrize the cicada's abdominal cavity motion and tymbal excitation and create a model to simulate the insect's sound production mechanism. The combined and individual contributions of the motion of the abdomen and tymbals can be evaluated and quantified using the nonlinear least-squares technique combined with a Wiener expansion. The Wiener method allows the computation of second- and third-order solutions with fewer coefficients. Reducing the coefficient size for each order makes the computational limitation less of a concern, and consideration can be given to calculating higher-order kernels. Determining these parameters will aid in developing a device that generates sound propagation with the efficiency of the cicada vocalizations. Such efficient sound wave propagation, if viable in water, would markedly enhance source radiation efficiency for a variety of sonar applications.

ACKNOWLEDGMENTS

This study was funded by the Office of Naval Research, program manager Dr. Joel Davis. In addition, the Naval Undersea Warfare Center Division, Newport, RI provided the lasers and the initial signal processing analysis for the laser data. The authors are specifically grateful to Dr. Paul Lefebvre, Dr. Pierre Corriveau, and Mr. Joseph Monti for their support of the experiment. The authors are also grateful to Dr. John Cooley for providing his knowledge of several species of cicada as well as his expertise in producing mating calls, to Mr. Michael Neckermann and Mr. Gerry Bunker for helping in the laser field experiment, and to Dr. Lynn Antonelli for her technical consultations on effectively using the dual-headed laser to record this historic data.

[1]H. C. Bennet-Clark, "Tymbal mechanics and the control of song frequency in the cicada Cyclochila australasiae," J. Exp. Biol. **200**, 1681–1694 (1997).
[2]H. C. Bennet-Clark, "Size and scale effects as constraints in insect sound communication," Philos. Trans. R. Soc. London, Ser. B **353**, 407–419 (1998).
[3]H. C. Bennet-Clark, "Resonators in insect sound production: How insects produce loud pure-tone songs," J. Exp. Biol. **202**, 3347–3357 (1999).
[4]H. C. Bennet-Clark and A. G. Daws, "Transduction of mechanical energy into sound energy in the cicada Cyclochila australasiae," J. Exp. Biol. **202**, 1803–1817 (1999).
[5]H. C. Bennet-Clark and D. Young, "The scaling of song frequency in cicadas," J. Exp. Biol. **191**, 291–294 (1994).
[6]H. C. Bennet-Clark and D. Young, "Sound radiation by the bladder cicada Cystosoma saundersii," J. Exp. Biol. **201**, 701–715 (1998).
[7]R. MacNally and D. Young, "Song energetics of the bladder cicada, Cystosoma saundersii," J. Exp. Biol. **90**, 185–196 (1980).
[8]J. R. Cooley and D. C. Marshall, "Sexual signaling in periodical cicadas, Magicicada spp. (Hemiptera: Cicadidae)," Behaviour **138**, 827–855 (2001).
[9]J. S. Dugdale and C. A. Fleming, "Two New Zealand cicadas collected on Cook's Endeavour Voyage, with description of a new genus," New Zealand J. Sci. **12**, 929–957 (1969).
[10]D. H. Lane, "The recognition concept of species applied in an analysis of putative hybridization in New Zealand cicadas of the genus Kikihia (Insecta: Hemiptera: Tibicinidae)," in *Speciation and the Recognition Concept: Theory and Application*, edited by D. M. Lambert and H. G. Spencer (Johns Hopkins University Press, Baltimore, 1995), pp. 367–421.
[11]C. L. Marlatt, "The periodical cicada," U.S. Dept. Agric. Bur. Entomol. Bull. **71**, 1–181 (1907).
[12]R. D. Alexander and T. E. Moore, "The evolutionary relationships of 17-year and 13-year cicadas, and three new species (Homoptera, Cicadidae, Magicicada)," U. Mich. Zoo. Misc. Pub. **121**, 1–59 (1962).
[13]D. Young and R. Josephson, "Pure-tone songs in cicadas with special reference to the genus Magicicada," J. Comp. Physiol. **152**, 197–207 (1983).
[14]H. C. Bennet-Clark and D. Young, "A model of the mechanism of sound production in cicadas," J. Exp. Biol. **173**, 123–153 (1992).
[15]P. J. Fonseca and A. V. Popov, "Sound radiation in a cicada: The role of different structures," J. Comp. Physiol., A **175**, 349–361 (1994).
[16]D. Young and H. C. Bennet-Clark, "The role of the tymbals in cicada sound production," J. Exp. Biol. **198**, 1001–1019 (1995).
[17]P. J. Fonseca and R. M. Hennig, "Phasic action of the tensor muscle modulates the calling song in cicadas," J. Exp. Biol. **199**, 1535–1544 (1996).
[18]P. J. Fonseca and A. V. Popov, "Directionality of the tympanal vibrations in a cicada: A biophysical analysis," J. Comp. Physiol., A **180**, 417–427 (1997).
[19]P. J. Fonseca and H. C. Bennet-Clark, "Asymmetry of tymbal action and structure in a cicada: A possible role in the production of complex songs," J. Exp. Biol. **201**, 717–730 (1998).
[20]N. H. Fletcher, *Acoustic Systems in Biology* (Oxford University Press, New York, 1992).
[21]G. W. Pierce, *The Songs of Insects* (Harvard University Press, Cambridge, 1948).
[22]W. J. Bailey, *Acoustic Behavior of Insects: An Evolutionary Perspective* (Chapman and Hall, New York, 1991).
[23]V. B. Wigglesworth, *The Principles of Insect Physiology* (Chapman and Hall, London, 1972).
[24]H. C. Gerhardt and F. Huber, *Acoustic Communication in Insects and Anurans: Common Problems and Diverse Solutions* (University of Chicago Press, Chicago, 2002).
[25]A. W. Ewing, *Arthropod Bioacoustics: Neurobiology and Behaviour* (Comstock, Ithaca, NY, 1989).
[26]J. E. Treherne, *Insect Neurobiology* (Elsevier, New York, 1974).
[27]B. Lewis, *Bioacoustics: A Comparative Approach* (Academic, New York, 1983).
[28]M. D. Atkins, *Introduction to Insect Behavior* (Macmillan, New York, 1980).
[29]S. P. Timoshenko and J. M. Gere, *Mechanics of Materials*, 3rd ed. (PWS, Boston, MA, 1990).
[30]H. D. I. Abarbanel, *Analysis of Observed Chaotic Data* (Springer-Verlag, New York, 1996) p. 28.
[31]J. Theiler, S. Eubank, A. Longtin, B. Galdrikian, and J. D. Farmer, "Testing for nonlinearity in time series: The method of surrogate data," IUTAM Symposium and NATO Advanced Research Workshop on Interpretation of Time Series for Nonlinear Mechanical Systems, University of Warwick, Coventry, UK, August 1991.
[32]F. M. Reza, *An Introduction to Information Theory* (Dover, New York, 1994) p. 283.
[33]G. C. Carter, *Coherence and Time Delay Estimation: An Applied Tutorial for Research, Development, Test, and Evaluation Engineers* (IEEE, Piscataway, NJ, 1993).
[34]M. Schetzen, *The Volterra and Wiener Theories of Nonlinear Systems*, revised ed. (Krieger, Malabar, FL, 2006).
[35]M. J. Hinich, E. M. Mendes, and L. Stone "Detecting Nonlinearity in Time Series: Surrogate and Bootstrap Approaches," *Studies in Nonlinear Dynamics & Econometrics* (The Berkeley Electronic Press, Berkeley, CA, 2005), Vol. **9**, No. 4, Art. 3.
[36]D. M. Patterson and R. A. Ashley, *A Nonlinear Time Series Workshop: A Toolkit for Detecting and Identifying Nonlinear Serial Dependence* (Kluwer Academic, Boston, MA, 2000).
[37]M. B. Moffett and R. H. Mellen, "Model for Parametric Acoustic Sources," J. Acoust. Soc. Am. **61**, 325–337 (1977).
[38]H. C. Bennet-Clark, "Tymbal mechanics and the control of song frequency in the cicada Cyclochila australasiae," J. Exp. Biol. **200**, 1681–1694 (1997).
[39]D. Young and H. C. Bennet-Clark, "The role of the tymbal in cicada sound production," J. Exp. Biol. **198**, 1001–1019 (1995).

Ossicular resonance modes of the human middle ear for bone and air conduction

Kenji Homma[a] and Yu Du
Adaptive Technologies, Inc., 2020 Kraft Drive, Suite 3040, Blacksburg, Virginia 24060

Yoshitaka Shimizu
Department of Otolaryngology-HNS, Stanford University, Stanford, California 94305 and Palo Alto Veterans Administration, Palo Alto, California 94305

Sunil Puria
Department of Mechanical Engineering and Department of Otolaryngology-HNS, Stanford University, Stanford, California 94305 and Palo Alto Veterans Administration, Palo Alto, California 94305

(Received 1 April 2008; revised 3 December 2008; accepted 4 December 2008)

The mean resonance frequency of the human middle ear under air conduction (AC) excitation is known to be around 0.8–1.2 kHz. However, studies suggest that the mean resonance frequency under bone conduction (BC) excitation is at a higher frequency around 1.5–2 kHz. To identify the cause for this difference, middle-ear responses to both AC and BC excitations were measured at the umbo and lateral process of the malleus using five human cadaver temporal bones. The resonance modes identified from these measurements, along with finite element analysis results, indicate the presence of two ossicular modes below 2 kHz. The dominant mode under AC excitation is the first mode, which typically occurs around 1.2 kHz and is characterized by a "hinging" ossicular motion, whereas the dominant mode under BC excitation is the second mode, which typically occurs around 1.7 kHz and is characterized by a "pivoting" ossicular motion. The results indicate that this second mode is responsible for the translational component in the malleus handle motion. The finding is also consistent with the hypothesis that a middle-ear structural resonance is responsible for the prominent peak seen at 1.5–2 kHz in BC limit data.
 [DOI: 10.1121/1.3056564]

PACS number(s): 43.64.Ha, 43.64.Bt [WPS] Pages: 968–979

I. INTRODUCTION

There is an ongoing need to provide more effective and reliable hearing protection devices for people who work in extremely noisy environments, such as on the flight deck of an aircraft carrier. Typical hearing protectors, such as earmuffs and earplugs, reduce the risk of hearing damage by suppressing the sound that reaches the inner ear through the air conduction (AC) pathway. However, the maximum amount of hearing protection provided by a conventional hearing protector is limited by the presence of bone conduction (BC) pathways, such as acoustically induced skull vibrations, through which residual acoustic energy is transmitted to the cochlea. The hearing protection performance limit due to BC is commonly called the "BC limit" (or the "BC threshold") in the literature. Figure 1 shows several estimates of the BC limit (Zwislocki, 1957; Berger *et al.*, 2003; Reinfeldt *et al.*, 2007).

The most notable feature of the BC limit estimates shown in Fig. 1 is the presence of a prominent peak between 1.5 and 2 kHz. The maximum overall attenuation obtainable by a typical hearing protection device is significantly restricted by the presence of this peak since it limits the mean attenuation level to about 40 dB in the critical mid-frequency range. It has been hypothesized that this peak may be associated with a middle-ear structural resonance. This hypothesis stems from the current understanding of the primary mechanisms of BC sound transmission, which are typically classified into the following three types (Silman and Silverman, 1991; Stenfelt *et al.*, 2002): (a) compressional, (b) inertial-ossicular, and (c) external-canal. The compressional BC mechanism refers to the case where the skull vibration is transmitted directly to the cochlea via vibrational distortions of the bone enclosing the cochlea fluid. The inertial-ossicular mechanism refers to the case where the BC excitation is transmitted to the cochlea through vibrations of the middle-ear ossicles. This is called the "inertial-ossicular" mode since it is the inertia of the middle-ear ossicles that leads to relative motion between the ossicles and the vibrations of the surrounding bone, which in turn causes cochlear excitation through vibrations of the stapes footplate. The external-canal mechanism is the case where BC-induced vibrations of the cartilaginous ear canal and ear plug produce acoustic pressure in the ear canal that excites the tympanic membrane (TM).

There have been a number of studies that point to the inertial-ossicular mechanism as the dominant BC hearing mechanism in the mid-frequency range around 2 kHz. A study by Carhart (1971) has shown that BC hearing sensitivity especially decreases at 2 kHz on average for patients with a condition of stapes fixation ("otosclerosis"), a phenomenon

[a]Electronic mail: kenji@adaptivetechinc.com

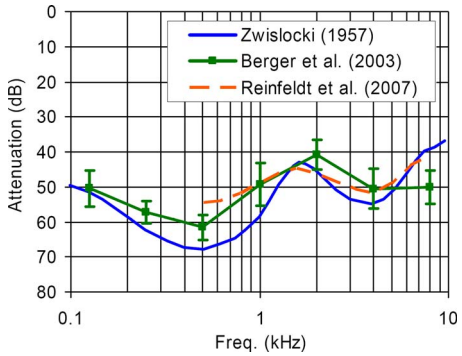


FIG. 1. (Color online) Estimates of the mean BC limit (i.e., the BC threshold): The data by Zwislocki (1957) and Reinfeldt *et al.* (2007) were obtained with a finer frequency resolutions than those by Berger *et al.* (2003) (error bars= ± 1S.D.). Note the prominent peak feature between 1.5 and 2 kHz, indicating a significant amount of BC sound transmission in this frequency range.

known as the "Carhart notch." Tonndorf (1972) attributed this effect to a loss of the middle-ear BC resonance contribution by observing correlation between the middle-ear resonance frequency and the frequency of BC hearing sensitivity loss for various animals. Linstrom *et al.* (2001) showed that BC hearing improved most significantly at 2 kHz after performing middle-ear reconstruction surgery on patients with middle-ear impairments.

Based on the evidence suggesting the dominance of the middle-ear BC mechanism around 2 kHz, it is logical to hypothesize that the prominent peaks occurring in the 1.5–2 kHz range in the BC limit data of Fig. 1 would most likely be due to the middle-ear BC mechanism, provided that BC is dominating over AC (which is considered to be the case as the terms BC limit and BC threshold suggest by definition). It is pointed out that the higher frequency-resolution data by Zwislocki (1957) and Reinfeldt *et al.* (2007) in Fig. 1 indicate the prominent BC limit peak to be at around 1.6–1.7 kHz, rather than exactly at 2 kHz as indicated by the BC limit data by Berger *et al.* (2003) with limited frequency resolution (which is typical of many hearing-related studies including the aforementioned studies indicating the main frequency of the middle-ear BC contribution to be exactly at 2 kHz.)

However, this appears to be inconsistent with knowledge that the mean middle-ear resonance frequency for AC is around 0.8–1.2 kHz (Silman and Silverman, 1991; Wada *et al.* 1998). Measurements carried out by Wada *et al.* (1998) on 275 ears from live subjects show the mean resonance frequency (± 1 S.D.) for AC to be 1.17(±0.27) kHz. This raises the following question: Why does the observed middle-ear resonance frequency appear in the 1.5–2 kHz range for BC rather than the 0.8–1.2 kHz range which is the commonly acknowledged middle-ear resonance frequency for AC? The reasons for this apparent difference in the middle-ear resonance frequencies between BC and AC have not been clarified in the past.

In order to identify the mechanisms responsible for this difference, measurements of both BC and AC middle-ear responses were conducted using five human temporal bone specimens. Furthermore, a finite element (FE) model of a middle-ear structural system was developed to gain insight

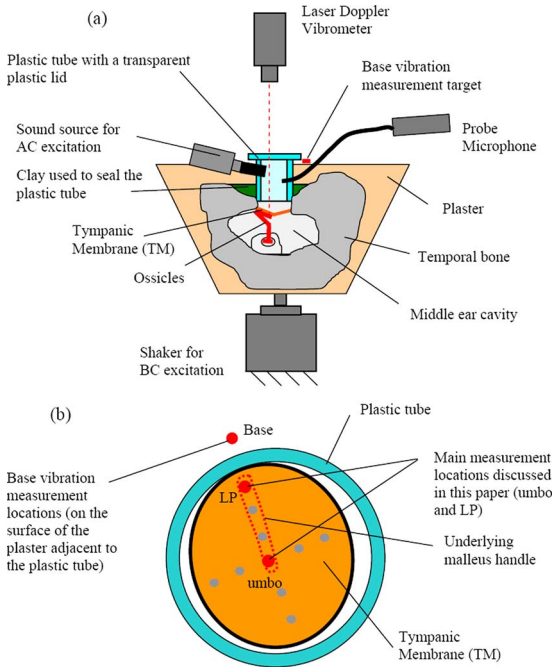


FIG. 2. (Color online) Temporal bone experiment setup for measuring the AC and BC middle-ear responses: (a) cross-sectional view; (b) laser measurement locations on the TM surface from the perspective of the otoscope integrated with the laser Doppler vibrometer. The measurement locations reported in this paper, which indicate the underlying malleus handle motion, are the LP and the umbo.

into the modal characteristics observed in the temporal bone measurement data. Understanding the fundamental modal characteristics, such as the natural frequencies and normal modes, is of critical importance in understanding middle-ear dynamics at low-to-mid frequencies. This study is a first step toward the goal of identifying the mechanisms behind the mid-frequency BC limit peak at 1.5–2 kHz, which could ultimately lead to further improvements in hearing protection devices used for extreme noise environments.

II. METHODS

A. Temporal bone measurements

Human temporal bone measurements were conducted to investigate the dynamic characteristics of the middle ear for both AC and BC excitations. The experimental setup is shown in Fig. 2.

1. Temporal bone preparations

Temporal bones were extracted from human cadavers within 48 h of death, at the time of autopsy. A total of five temporal bones (TBs) were used in this study (labeled as TB1, TB2, TB3, TB4, and TB5), which were extracted from an 84 year old male (left ear), a 61 year old male (right ear and left ear for TB2 and TB3), an 81 year old male (right ear), and an 83 year old female (right ear), respectively. For all preparations, the attached extraneous tissues were removed and the bony wall of the external ear canal was drilled down to 2 mm from the tympanic annulus. A 10-mm-long plastic tube with an 8.5 mm internal diameter was placed against the bony ear canal remnant such that the axis of the tube was approximately perpendicular to the

plane defined by the tympanic annulus around the edge of the tympanic membrane (TM). This plane is referred to as the "TM plane" in this study. The plastic tube was held in place with clay, and the majority of the temporal bone was encased in plaster. A transparent plastic disk was attached to the tube opening using beeswax in order to acoustically isolate the tube canal, and also to increase the efficiency of introducing sound for AC excitation. The effects of the resulting closed air volume were checked by measuring and comparing the BC responses with and without the plastic disk, and no significant differences were observed. The AC responses were not affected since they are normalized by the acoustic pressure in the air volume.

The middle-ear air cavity (tympanic cavity plus mastoid cavity) also formed a closed air volume since it was not vented to the outside. However, the contribution of the middle-ear air cavity to the overall middle-ear dynamics is considered to be insignificant (Zwislocki, 1962). It is not thought to have a recognizable effect unless the middle-ear air volume is reduced to the extreme, such that the whole mastoid cavity is sacrificed (McElveen *et al.*, 1982; Voss *et al.*, 2000). Based on qualitative inspections of the temporal bones used in this study, it is likely that a significant portion of the mastoid air volume was retained in each specimen. Therefore, it is considered unlikely that the middle-ear air cavities provided a sufficient acoustic stiffness to significantly affect the dynamics. This was also supported by the observation that the mean value of the measured AC resonance frequency among the temporal bones, which is proportional to the square root of the middle-ear stiffness, was largely consistent with the normal expected value as shown later in Sec. IV.

The state of the cochlear fluid was not inspected during the temporal bone preparations, on the assumption that the fluid would be mostly intact for fresh temporal bones. Also, it has been shown that the cochlear input impedance is predominantly resistive (Aibara *et al.*, 2001), which implies that leakage of cochlear fluid mostly has the effect of reducing the amount of damping in the system, and this is not of critical importance in this study.

2. AC/BC excitation methods

The plaster-encased temporal bone specimens were attached to a shaker (B&K type 4810, B&K, Nærum, Denmark) to provide BC stimulation. The temporal bone preparations were rigidly attached to the shaker using a screw and cement in such a way that the direction of vibration was approximately perpendicular to the TM plane. AC excitations were provided by a small acoustic transducer, which introduced acoustic pressure into the plastic tube positioned above the TM. The peak sound pressure level for the AC stimulation was 90–93 dB, which is well below the range where nonlinear distortions are expected to occur (120–130 dB) (Voss *et al.*, 2000). The middle-ear vibrations produced by the BC excitation were comparable to or slightly smaller than those produced by the AC excitation: the BC-induced umbo displacement magnitudes ranged from 1 to 40 μm, while the AC-induced magnitudes ranged from 2 to 80 μm. Therefore, the middle-ear structures were not overdriven during the measurements.

3. Response measurements

A probe tube microphone (ER-7C, Etymotic Research, Elk Grove Village, IL) was also installed to measure the acoustic pressure within the plastic tube. The tip of the probe tube was positioned at a distance of about 2 mm from the TM. A laser Doppler vibrometer sensor head (HLV-1000, Polytec, Tustin, CA) with a joystick-controlled mirror was mounted on an operating microscope to enable the laser beam to be aimed at the desired measurement points. Measurements were obtained at multiple locations on the surface of the TM, as shown in Fig. 2(b). Retroreflective microbeads were positioned at the measurement locations to increase the reflected signal. These microbeads were negligibly small in size (about 5–10 μm in diameter and 0.001 mg in mass), and thus provided no potential for mass loading. Although measurements were obtained at a number of TM locations during the experiments, the discussion in this paper is limited to the lateral process (LP) and umbo locations. Since the TM is firmly attached to the malleus bone at these two locations, they are suitable targets for measuring the malleus handle motion. As it will become clear in subsequent discussions, the rigid-body motions of the underlying malleus handle, observed from the two-point measurements, offer essential clues regarding the overall three-dimensional (3D) motion characteristics of the middle ear.

4. Response data format

AC responses were calculated by normalizing the umbo and LP velocities, v_{umbo} and v_{lp}, by the pressure in the plastic "ear-canal" tube, p_{ec}. BC responses were measured at the same locations along the malleus handle, but with the BC excitation introduced by the shaker. Since the surrounding bone structure encasing the middle ear was also vibrating in this case, the ossicular responses are expressed in terms of the differential velocity (Stenfelt *et al.*, 2002), Δv,

$$\Delta v/v_{\text{base}} = (v - v_{\text{base}})/v_{\text{base}}, \tag{1}$$

where v_{base} is the base vibration velocity introduced by the BC excitation. The base velocity was obtained at a point near the tube on the plastered surface of the temporal bone, as shown in Fig. 2. The differential velocity, Δv, in the BC response is equivalent to the absolute velocity, v, in the AC response, with the only difference being that the base velocity is zero for AC.

5. Identification of the primary resonance frequency

Primary resonance frequencies in AC and BC, which are designated as f_o^{ac} and f_o^{bc}, respectively, were identified for each individual temporal bone from the middle-ear response data. A system identification procedure, based on the invfreqs function in MATLAB (Mathworks, Natick, MA), was performed for this task. The invfreqs function accepts complex frequency response data and uses a curve-fitting algorithm to find an approximate representation of these data as a

rational transfer function, $H(s)$, that consists of the ratio of a numerator polynomial, $B(s)$, (of order m) and a denominator polynomial, $A(s)$, (of order n):

$$H(s) = \frac{B(s)}{A(s)} = \frac{b_m s^m + \cdots + b_1 s + b_0}{a_n s^n + \cdots + a_1 s + a_0}. \quad (2)$$

This transfer function expresses an input-output relationship of a dynamic system in the frequency domain (i.e., the s-domain). The polynomial orders, m and n, are selected iteratively by the user until a transfer function is found that produces the best fit to the frequency response data in the frequency range of interest. [That is, all the polynomial coefficients, b_m and a_n, in Eq. (2) are determined.] The natural frequencies, ω_n, and damping ratios, ξ, associated with the resonance peaks in the frequency response are then extracted from the complex-valued roots of the denominator polynomial, $A(s)$ [i.e. the roots of the characteristic equation, $A(s) = 0$], as

$$s = -\xi\omega_n \pm j\omega_n\sqrt{1-\xi^2}, \quad (3)$$

which are also called "system poles" and represent the complex-valued eigenvalues of the dynamic system. In the present case, the purpose of this system identification procedure was to identify the characteristic parameters, ω_n and ξ, associated with the primary resonance peak, which is the first major resonance peak visible in the mid-frequency range. Once a good fit was achieved, and the natural frequency, ω_n, associated with the primary resonance peak was determined, the primary resonance frequency in hertz was given as $f = \omega_n/2\pi$. In practice, there was some variability in the identified values of the natural frequency and damping ratio due to the dependence on user judgment for choosing the appropriate polynomial orders, m and n, and also due to the relatively high level of damping that existed in the middle-ear frequency responses, which also tended to increase the variability. In an effort to minimize this variability, the system identification procedure was performed on both umbo and LP responses, and then the mean of the two values was used. This is a valid technique since natural frequencies and damping ratios are properties inherent to the structural system, and therefore the same modal information should be contained in both frequency responses, which are taken from different locations of the same system.

It should be noted that, in this study, the "resonance frequency" is synonymous with the natural frequency associated with the resonance. Strictly speaking, the center frequency of a resonance peak can be shifted slightly from the natural frequency depending on the damping level. However, for the damping ratio, ξ, presently observed for the middle ear, which is about 0.2 on average, the resonance peak center frequency is approximately equal to the natural frequency, with the potential theoretical difference being only up to about 4% (James *et al.*, 1994). It should also be noted that the natural frequency is a property determined only by the mass and stiffness of the structure, and thus is independent of the damping.

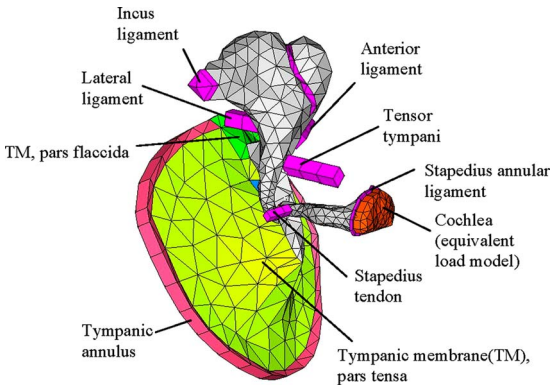


FIG. 3. (Color online) A FE model of a human middle-ear structure (left ear). The model consists of the TM, ossicles, ligaments, tendons, and a simplified model of the cochlea. The air volumes of the ear canal and the middle-ear cavity were not included in the model.

6. Frequency normalization

Further processing of the measured data was performed, in which the frequency axis of each individual frequency response was normalized by its primary resonance frequency, f_o^{ac} (or f_o^{bc}). This was done in order to compensate for the variability in resonance frequencies among the temporal bones, whose basic modal characteristics were otherwise highly similar aside from being shifted in frequency according to the resonance frequency values. In previous middle-ear dynamics studies, the frequency responses obtained from a number of temporal bone specimens were typically averaged to obtain mean response data, based on which middle-ear models were developed. However, this approach is not suitable in this case, since the averaging process tends to eliminate or "smear" the detailed modal characteristics that were otherwise present in individual responses. Since the current focus is on the middle-ear modal characteristics, it was necessary to observe the individual middle-ear responses without averaging.

B. Finite element analysis

A FE middle-ear model was developed in order to help reveal the 3D dynamic characteristics of the overall middle-ear structure, which underlies the malleus handle motions observed in the measured data.

1. Human middle-ear FE model

Figure 3 shows the FE model of a left middle ear. The middle-ear FE model consisted of the TM, ossicles (malleus, incus, and stapes), ligaments, and tendons. The cochlea was modeled as an equivalent mechanical load based on the cochlear input resistance value reported by Aibara *et al.* (2001). The geometries of the ossicles and the TM were based on micro-CT imaging data from an actual human middle-ear sample (Sim *et al.*, 2007). Then, a corresponding FE mesh model was created using the FE pre/post processing software HYPERMESH (Altair Engineering, Tory, MI). Tendons and ligaments were not directly based on the micro-CT imaging data, but approximated using columns with constant cross sections. Solid elements were used for most of the components except for the TM, which instead consisted of "solid-

TABLE I. Material properties of various components in the middle-ear FE model. The initial material values obtained from the literature are shown in curly brackets.

Component	Elastic modulus E_1 (N/m^2)	Density ρ (kg/m^3)	Loss factor η
Incus ossicle	1.41×10^{10} [a]	2.15×10^{3} [c]	0.01 (constant)
Malleus ossicle	1.41×10^{10} [a]	2.39×10^{3} [c]	0.01 (constant)
Stapes ossicle	1.41×10^{10} [a]	2.20×10^{3} [a]	0.01 (constant)
Malleus/incus joint	1.41×10^{10} [a]	2.39×10^{3}	0.01 (constant)
Incus/stapes joint	4.4×10^{5} {6×10^{5} [a]}	1.2×10^{3} [a]	0.15 at 1 kHz
Tensor tympani	1.9×10^{7} {7.0×10^{7} [a]}	1.2×10^{3}	0.15 at 1 kHz
Anterior ligament	1.5×10^{7} {2.1×10^{7} [a]}	1.2×10^{3}	0.15 at 1 kHz
Lateral ligament	5.0×10^{5} {6.7×10^{6} [a]}	1.2×10^{3}	0.15 at 1 kHz
Stapes tendon	3.8×10^{5} {5.2×10^{7} [a]}	1.2×10^{3}	0.15 at 1 kHz
Incus ligament	4.8×10^{6} {6.5×10^{6} [a]}	1.2×10^{3}	0.15 at 1 kHz
Tympanic membrane, Pars tensa	3×10^{7} [b]	1.2×10^{3} [d]	0.15 at 1 kHz
Tympanic membrane, Pars flaccida	0.7×10^{7} {$1/3E_{\text{pars tensa}}$}	1.2×10^{3} [d]	0.15 at 1 kHz
Tympanic annulus	6×10^{5} [d]	1.2×10^{3}	0.15 at 1 kHz
Stapes annular ligament	4.12×10^{5} [e]	1.2×10^{3}	0.25 at 1 kHz[e]
Cochlear load	Effective resistance: 30 GΩ[f] (mass and stiffness are assumed to be insignificant)		

[a]Gan *et al.* (2004).
[b]Fay *et al.* (2006).
[c]Sim *et al.* (2007).
[d]Koike *et al.* (2002).
[e]Shin *et al.* (2008).
[f]Aibara *et al.* (2001) [with 1.41 correction factor applied; see O'Connor and Puria (2008)].

shell" elements (based on the shell equations, but with 3D solid element topologies).

The material properties of various components were initially obtained from the existing literature (Fay *et al.*, 2006; Sim *et al.*, 2007), including past studies of FE middle-ear modeling (Koike *et al.*, 2002; Gan *et al.*, 2004). In the previous FE middle-ear modeling studies, material property values for the ligaments and tendons were typically adjusted by matching the simulated responses to the experimental dynamic responses. The material property values in the present model were also tuned by comparing against the temporal bone data. However, the approach in this study differed from the previous approaches in some aspects. The elastic modulus values of the ligaments and tendons in the present FE model were calibrated by performing the normal mode analysis first, so that natural frequencies were in agreement with the average values observed in the temporal bone measurements. This is in contrast to previous approaches where the FE model was calibrated against the mean AC response data, whose detailed modal characteristics were no longer identifiable due to the averaging process across a number of temporal bones. Having determined the elastic modulus values of the components, the damping in the model was then adjusted so that the simulated responses (for both AC and BC) exhibited the level of damping that was consistent with that observed in corresponding temporal bone responses. The damping was adjusted through varying the material loss factors associated with the components. Table I shows the final material property values of the middle-ear components as a result of this model tuning process, along with the initial reference values obtained from the literature.

The current FE model did not include air volumes associated with the ear canal and the middle-ear cavities. This is because the impedance magnitudes associated with these air cavities were expected to be significantly smaller than those associated with the structural part of the middle ear and consequently the effects of the air cavities on the dynamics of the middle ear were considered to be insignificant (Zwislocki, 1962). Therefore, the current FE model excluded the air cavities, which allowed concentration on mechanically the most significant part of the middle-ear structures.

2. Forced response analysis

Simulations were performed using the FE simulation software ACTRAN (Free Field Technologies, Belgium), which was developed specifically for vibroacoustic problems. Two kinds of FE analysis were performed: one was the forced response analysis, and the other was the normal mode analysis. In the forced response analysis, the middle-ear responses to external excitations (either AC or BC) are simulated by solving the following matrix equation of motion (EOM):

$$\mathbf{K}\cdot\mathbf{x}-\omega^2\mathbf{M}\cdot\mathbf{x}=\mathbf{f}, \tag{4}$$

where ω is the radian frequency and $\mathbf{x}$ is the displacement vector to be solved as a response to the forcing vector, $\mathbf{f}$. The matrices, $\mathbf{K}$ and $\mathbf{M}$, are the stiffness and mass matrices. The stiffness and damping properties associated with the structural components are represented by the frequency-dependent complex-valued material modulus:

$$E(\omega)=E_1(\omega)+jE_2(\omega), \tag{5}$$

where E_1, is called the "storage" modulus which represents the stiffness, and the imaginary part, E_2, is called the "loss"

modulus which defines the damping. As a result, the stiffness matrix, $\mathbf{K}$, in the EOM of Eq. (4) is complex valued and frequency dependent:

$$\mathbf{K}(\omega) = \mathbf{K}_1(\omega) + j\mathbf{K}_2(\omega), \tag{6}$$

where $\mathbf{K}_1$ and $\mathbf{K}_2$ represent the overall stiffness and damping of the system, respectively.

The complex material modulus of Eq. (5) above can be alternatively written as

$$E(\omega) = E_1(\omega)\left(1 + j\frac{E_2(\omega)}{E_1(\omega)}\right) = E_1(\omega)(1 + j\,\eta(\omega)), \tag{7}$$

where the loss factor, η, specifies the material damping. In the current middle-ear FE model, the storage modulus, E_1, is assumed to be constant, which means that the matrix $\mathbf{K}_1$ in Eq. (6) is also constant. The loss factor for the middle-ear ligaments and tendons are assumed to be proportional to frequency, that is,

$$\eta(\omega) = c\omega, \tag{8}$$

where c is a constant. In Table I, the proportionality constants, c, for components are specified indirectly by a reference loss factor value at 1 kHz. This material damping model is essentially equivalent to the Rayleigh damping model (with only the stiffness-proportional constant β being used, and the mass-proportional constant α being set to 0), which has been used in previous middle-ear FE model studies. It can be shown that the constant β of the Rayleigh damping model is equivalent to the constant c in the loss factor model of Eq. (8) (James *et al.*, 1994). It should be noted that the loss factor, η, which describes the material damping in individual components, should be distinguished from the damping ratio, ξ, of Eq. (3), which describes the total damping in the system resulting from the cumulative effects of the component-level damping.

AC excitations were simulated by assigning a uniformly distributed dynamic pressure over the TM surface on the ear-canal side. This is considered reasonable for the current study since the acoustic wavelength is still fairly large compared to the TM dimensions for the frequencies of interest (up to 4–5 kHz). The BC excitations were simulated by assigning uniform displacement vectors (both magnitude and phase) at the boundaries of the structure, such as at the ends of the ligaments and tendons, and the edge of the tympanic annulus. This essentially simulated the rigid-body vibration of the base temporal bone structure. The direction of the BC excitation was perpendicular to the TM plane (i.e., the plane of the tympanic annulus), which is consistent with the temporal bone measurements.

3. Normal mode analysis

The other type of FE analysis performed in this study was the normal mode analysis, in which the following equation is solved:

$$\mathbf{M}^{-1}\mathbf{K}_1\mathbf{x} = \omega^2\mathbf{x}. \tag{9}$$

This equation is obtained from the EOM of Eq. (4) above with the excitation vector, $\mathbf{f}$, set to zero. It should be noted that the stiffness matrix is, $\mathbf{K}_1$, which is only the real part of the original stiffness matrix, $\mathbf{K}$, and does not include the damping part, $\mathbf{K}_2$. This renders Eq. (9) a real-valued eigenproblem, which is solved to obtain the natural frequencies (eigenfrequencies), ω_n, and the mode shapes (eigenvectors), $\mathbf{x}$. The removal of the system damping was achieved in the present model by eliminating the imaginary parts of the elastic modulus values for all the components, including the cochlear fluid load. In a normal mode analysis, it is typical to solve this real-valued eigenproblem without damping, since the natural frequencies and mode shapes are properties that depend solely on the mass and stiffness characteristics of a dynamic structure, and are independent of the damping and excitation.

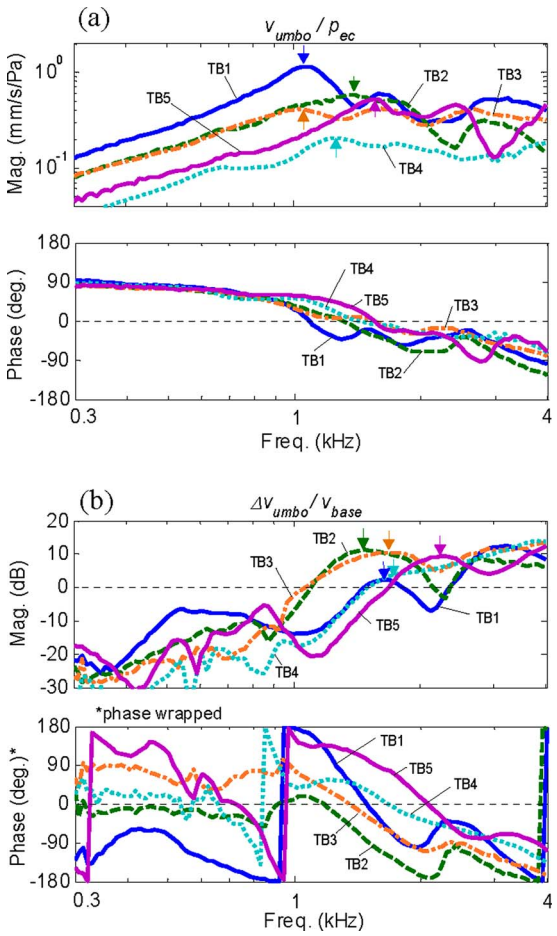


FIG. 4. (Color online) Measured umbo responses from the five temporal bones: (a) AC responses—ratio of the umbo velocity (mm/s), v_{umbo}, to the ear-canal acoustic pressure (Pa), p_{ec}; (b) BC responses—ratio of the differential umbo velocity, Δv_{umbo} ($=v_{umbo}-v_{base}$), to the base excitation velocity, v_{base}. The arrows indicate the primary resonance peaks.

III. RESULTS

This section presents the results obtained from the temporal bone measurements, followed by the simulation results obtained by the FE analyses.

A. Temporal bone measurements

1. Measured middle-ear responses

Figure 4 shows the middle-ear responses measured at

TABLE II. Primary AC and BC resonance frequencies, f_o^{ac} and f_o^{bc}, along with the associated damping ratios identified from the measured temporal bone responses.

	Primary AC resonance f_o^{ac}		Primary BC resonance f_o^{bc}	
Temporal bone	Frequency (kHz)	Damping ratio, ξ	Frequency (kHz)	Damping ratio, ξ
TB1	1.06	0.18	1.66	0.18
TB2	1.30	0.21	1.44	0.18
TB3	1.13	0.23	1.62	0.29
TB4	1.23	0.15	1.59	0.21
TB5	1.57	0.17	2.31	0.14
Mean(±S.D.)	1.26(±0.20)	0.19(±0.03)	1.72(±0.34)	0.20(±0.06)

the umbo for five temporal bones, with the AC responses, $v_{\text{umbo}}/p_{\text{ec}}$, shown in Fig. 4(a), and the BC responses, $\Delta v_{\text{umbo}}/v_{\text{base}}$, shown in Fig. 4(b). The arrows indicate the primary resonance peaks. It can be observed from the two plots that the primary resonance peaks for BC occur at higher frequencies than for AC.

2. Identified primary resonance frequencies

Table II summarizes the values of the primary resonance frequencies, f_o^{ac} and f_o^{bc}, and associated damping ratios, ξ, identified for the five temporal bones. As shown in the table, the mean AC resonance frequency, f_o^{ac}, is 1.26(±0.20) kHz and the mean BC resonance frequency, f_o^{bc}, is 1.72(±0.34) kHz, which is clearly higher than the mean AC resonance frequency, f_o^{ac}. The BC resonance frequency is observed to be higher than the AC resonance frequency in every temporal bone, typically by a factor of around 1.4–1.5. The mean damping ratios, ξ, identified for the AC and BC resonances are 0.19(±0.03) and 0.20(±0.06), respectively.

3. Frequency-normalized middle-ear responses

Figure 5 shows the frequency-normalized middle-ear responses. Figures 5(a) and 5(c) show the frequency-normalized umbo responses for AC, $v_{\text{umbo}}/p_{\text{ec}}$, and for BC, $\Delta v_{\text{umbo}}/v_{\text{base}}$, respectively. The magnitude of each umbo AC response has also been normalized by its mean magnitude calculated over the frequency range shown. Figures 5(b) and 5(d) show the velocity ratios between the umbo and LP, which indicate the two-dimensional motion of the malleus handle for AC, $v_{\text{umbo}}/v_{\text{lp}}$, and for BC, $\Delta v_{\text{umbo}}/\Delta v_{\text{lp}}$.

As shown in Fig. 5, the normalization clearly reveals common characteristics, which are otherwise difficult to distinguish, among the temporal bone data. First, in the umbo AC responses shown in Fig. 5(a), the primary resonances occur at $f/f_o^{ac}=1$ by definition, as a result of the normalization. The phase responses tend to cross or approach 0° around $f/f_o^{ac}=1$. There also tend to be two additional resonance peaks at higher frequencies at $f/f_o^{ac}\approx 1.5$ and 2.3. The umbo/LP velocity ratio plot in Fig. 5(b) also reveals common characteristics among the temporal bones. At low frequen-

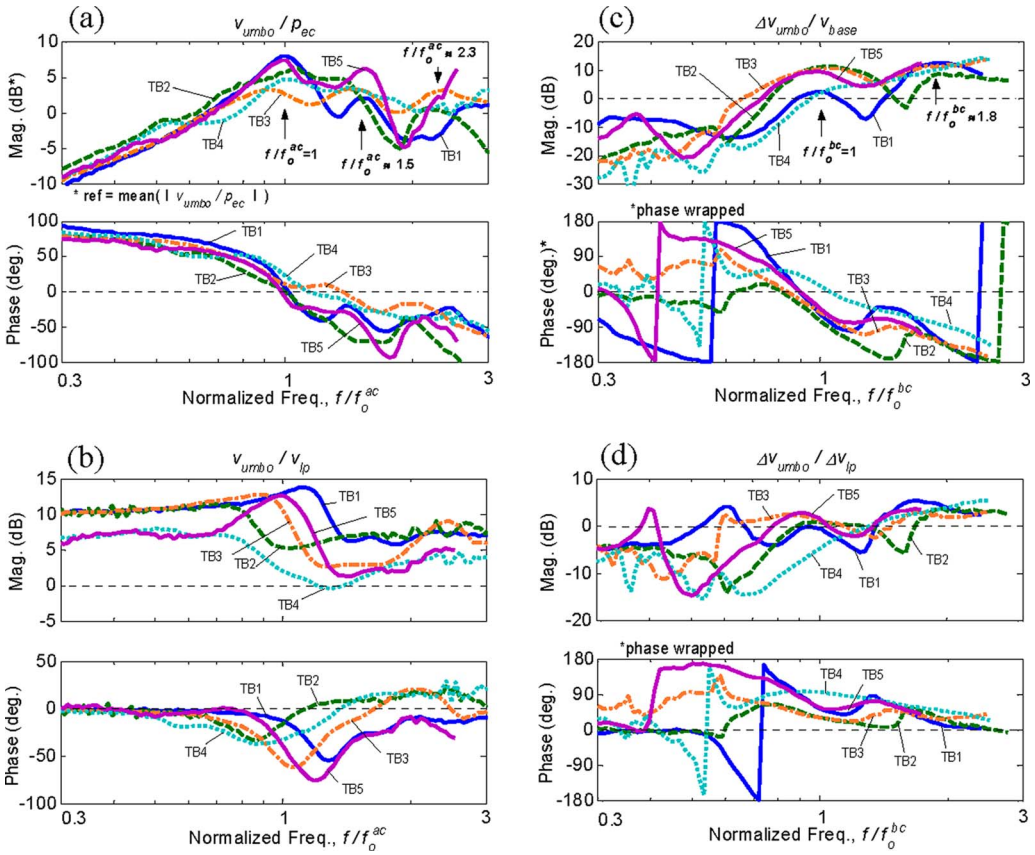


FIG. 5. (Color online) Frequency-normalized middle-ear responses of the five temporal bones: (a) umbo AC response, $v_{\text{umbo}}/p_{\text{ec}}$; (b) umbo/LP AC velocity ratio, $v_{\text{umbo}}/v_{\text{lp}}$; (c) umbo BC response, $\Delta v_{\text{umbo}}/v_{\text{base}}$; (d) umbo/LP BC velocity ratio, $\Delta v_{\text{umbo}}/\Delta v_{\text{lp}}$. The frequency axes are normalized by their respective resonance frequencies, f_o^{ac} and f_o^{bc}. Each umbo AC response in (a) are also normalized by its mean magnitude calculated over the frequency range shown.

cies, the umbo and LP motions are in phase, but the magnitudes are significantly higher for the umbo than the LP. Then, around the primary AC resonance frequency, f_o^{ac}, there is a transitional region where the magnitude of the umbo/LP response ratio shifts to a smaller value. This is accompanied by a dip in the phase around this frequency.

The frequency-normalized umbo BC responses in Fig. 5(c) show the primary BC resonance at $f/f_o^{bc}=1$, by definition, and the phase responses are mostly around −90° at this frequency. An additional resonance peak is also recognizable at around $f/f_o^{bc}\approx 1.8$, and the magnitudes become significantly smaller for frequencies below $f/f_o^{bc}\approx 0.8$. Figure 5(d) shows the umbo/LP velocity ratio for BC, $\Delta v_{umbo}/\Delta v_{lp}$, which exhibits significant similarity among the temporal bones at frequencies above $f/f_o^{bc}\approx 0.7$. The figure shows that the magnitude and the phase of the umbo/LP response ratios tend to approach zero at the primary BC resonance frequency, $f/f_o^{bc}=1$. At frequencies above $f/f_o^{bc}=1$, there tends to be a slight dip in magnitude, but then it recovers to about 2–5 dB at higher frequencies.

It is observed in Figs. 5(c) and 5(d) that the BC responses show significant variability, especially in phase responses, at low frequencies below $f/f_o^{bc}\approx 0.7$. The likely reason for this variability is that the BC response is obtained by taking the difference between two velocity measurements (e.g., $v_{umbo}=v_{umbo}-v_{base}$), such that when the two velocities become more similar at low frequencies, the phase of their difference becomes more sensitive to small variations between the two velocities. Furthermore, some temporal bone data are seen to contain low-frequency peaks, for example, at $f/f_o^{bc}=0.3–0.4$ in TB1 and $f/f_o^{bc}=0.35$ in TB5 in Fig. 5(c), which are not found in the other temporal bones. These extra peaks may be the result of minor low-frequency rocking motions of the temporal bone/shaker assembly, caused by an imperfect alignment of the temporal bone's center of gravity with the shaker's axis of vibration. However, the inconsistency in the BC response data at low frequencies does not affect the current discussion since the main frequency range of interest for BC is $f/f_o^{bc}\approx 0.7–1.5$, where the BC response magnitudes become significant due to the middle-ear BC resonance at f_o^{bc}.

B. Finite element analysis

1. Normal mode analysis

The normal mode analysis identified a total of three modes in the frequency range below 3 kHz whose natural frequencies, designated as f_0, f_1, and f_2, are 1.15, 1.69, and 2.76 kHz, respectively. Although the analysis identified additional modes at frequencies above 3 kHz, these higher-order modes are not discussed in this study since the focus is on the characteristics of the low-to-mid frequencies. The first two modes, found at 1.15 and 1.69 kHz, are attributed to rigid-body motions of the ossicles, which are supported by flexible ligaments and tendons, and are therefore referred to as "ossicular modes" in this study. Figure 6 shows qualitative illustrations of the characteristic middle-ear motions (i.e., mode shapes) associated with these two ossicular modes. Figure 6(a) shows the first mode, at $f_0=1.15$ kHz, which is characterized by a hinging motion of the malleus-incus complex about an axis that connects the incus ligament and a point near the LP. On the other hand, the second mode, at $f_1=1.69$ kHz, shown in Fig. 6(b) involves an ossicular motion that is significantly different from that of the first mode at f_0. The characteristic motion of the second mode can be described as the malleus-incus assembly "pivoting" about an axis running through the incus ligament in the superior-inferior direction, which is clearly seen by the superior view in Fig. 6(b). The posterior view of this mode in Fig. 6(b) shows a translational movement of the malleus handle where the umbo and the LP are moving in phase and in parallel. The stapes motions associated with both modes are observed to be rather "piston-like," but with some amount of rocking component also observable. The third mode, at $f_2=2.76$ kHz, is observed to be mainly associated with the TM, and is characterized by a prominent displacement in the posterior region of the TM consistent with the measurements by Tonndorf and Khanna (1972). This TM resonance mode was not of primary interest in this study, therefore and was not illustrated in Fig. 6.

2. Forced response analysis

Figure 7 shows the simulated AC and BC middle-ear responses in comparison to the measured responses of one individual temporal bone, TB3. This bone was chosen for comparison with the model since its natural frequencies are close to the mean values from the five temporal bones that were used to tune the FE middle-ear model (Table II). This bone also exhibits response characteristics that reflect the overall trends of the four other temporal bones in the normalization analysis of Fig. 5. As discussed earlier, it was important to compare the model with the response from a representative individual ear rather than with the mean response, since in the latter case the detailed response characteristics tend to be lost due to averaging. The figure also shows simulated middle-ear responses with a negligible level of damping, in order to highlight the resonance characteristics. This was done by reducing the loss factors, η, from 0.15–0.25 (Table I) to 0.01, for all tendons and ligaments and also by removing the resistive cochlear load.

Figures 7(a) and 7(c) show the simulated and measured umbo responses for AC, v_{umbo}/p_{ec}, and for BC, $\Delta v_{umbo}/v_{base}$, respectively. Figures 7(b) and 7(d) show the velocity ratios between the umbo and LP for AC, v_{umbo}/v_{lp}, and for BC, $\Delta v_{umbo}/\Delta v_{lp}$, respectively. The figures also show simulated responses for the stapes footplate, which were calculated by taking the ratio of the normal velocity at the center of the stapes footplate, v_{stapes}, to the acoustic pressure applied at TM, p_{ec}. The simulated umbo AC response in Fig. 7(a) shows a high degree of correspondence with the temporal bone measurement data. The result shows the presence of three major resonance peaks, which are associated with the three modes at f_0, f_1, and f_2 identified earlier in the normal mode analysis. These resonance peaks become especially clear in the simulated response with reduced damping. Comparing Fig. 7(a) with Fig. 5(a) suggests that these three middle-ear modes, at f_0, f_1, and f_2, correspond to the three

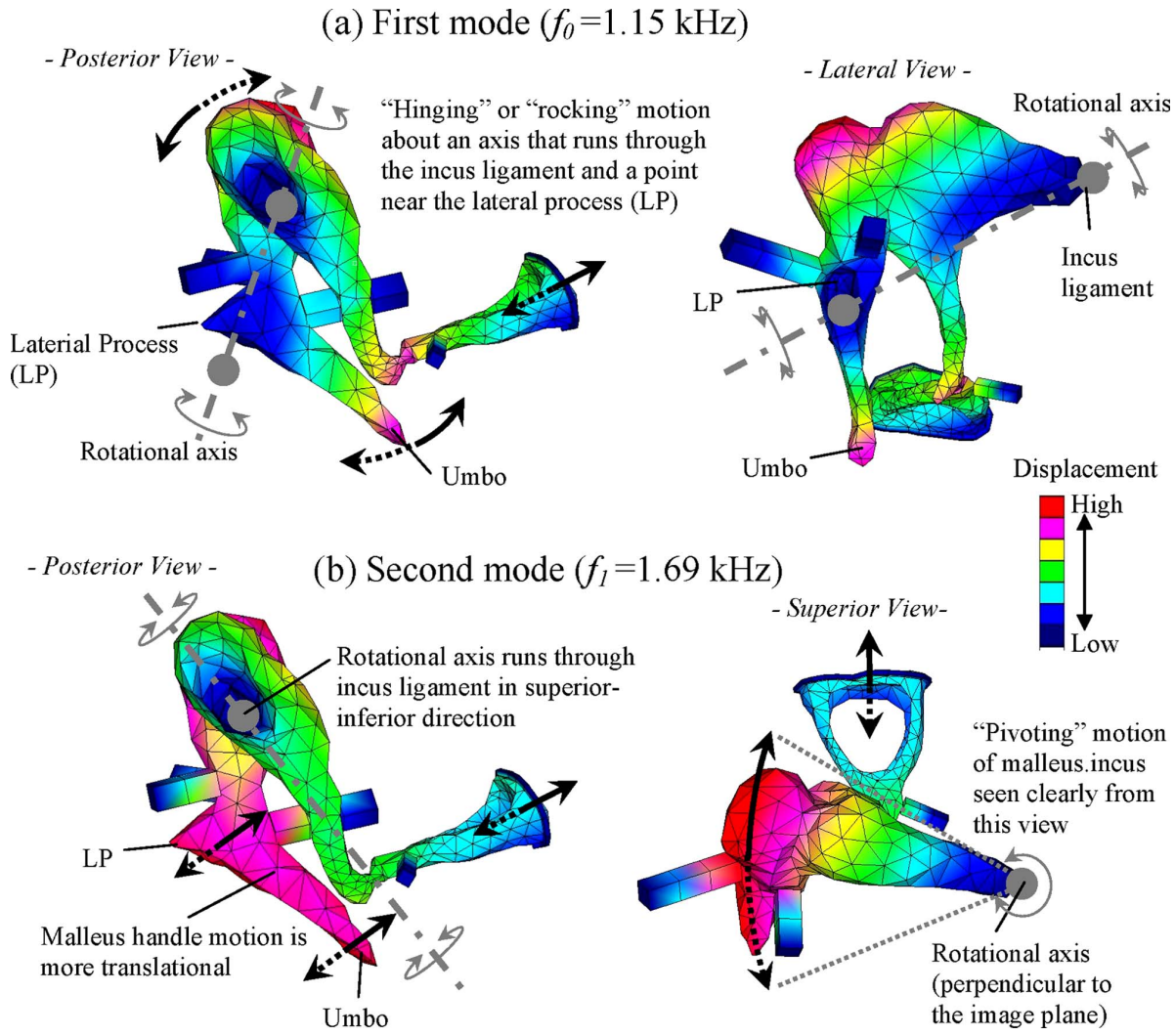


FIG. 6. (Color online) Characteristic motions (mode shapes) associated with the two middle-ear modes identified by the FE normal mode analysis below 2 kHz: (a) The first mode ("hinging mode") at f_0=1.15 kHz; (b) The second mode ("pivoting mode") at f_1=1.69 kHz. The color map indicates relative displacement amplitude and the arrows suggest general motional directions.

resonance peaks observed earlier in the frequency-normalized AC responses at $f/f_o^{ac} \approx 1$, 1.5, and 2.3.

Figure 7(a) also shows the simulated AC response at the stapes footplate, v_{stapes}/p_{ec}. It can be seen that the stapes response is also mainly characterized by the three middle-ear modes, f_0, f_1, and f_2. In other words, the modal response characteristics of the middle ear observed at umbo are also observable at stapes. This is reasonable since the overall motion of the middle-ear structural system, characterized by the three normal modes, is also what drives the stapes. This in turn reinforces the validity of the present approach, in which the modal characteristics of the middle ear are investigated primarily through the motions of the malleus handle (i.e., the umbo and LP), rather than the stapes.

Figure 7(b) shows the umbo/LP velocity ratio, v_{umbo}/v_{lp}, in response to AC excitation. Again, the response characteristics exhibited by the measured data are well captured by the FE model. At frequencies below the first mode at f_0, the umbo response magnitude is significantly higher than that of the LP, which is characteristic of the hinging motion associated with the first mode shown in Fig. 6(a). Above f_0, both the magnitude and phase transition to approximately zero at f_1, indicating a translational motion of the malleus handle, which is associated with the characteristic pivoting motion of the second mode shown in Fig. 6(b). These characteristics can also be seen in the reduced damping case where the malleus handle motion at f_0 becomes essentially that of the first mode (rotational malleus handle motion) and the motion at f_1 becomes that of the second mode (translational malleus handle motion).

The simulated umbo BC response shown in Fig. 7(c), $\Delta v_{umbo}/v_{base}$, also exhibits good agreement with the overall characteristics of the measured response. Again, the BC response in this frequency range is also characterized mainly by the three normal modes at f_0, f_1, and f_2. This is also observed for the simulated stapes BC responses, which exhibit similar modal response characteristics. [The stapes BC response is calculated using Eq. (1), but in the normal direction of the stapes footplate. Accordingly, the base velocity for the stapes, $v_{base,stapes}$, is the component of the base velocity, v_{base}, in the normal direction of the stapes footplate.] By comparing the BC responses in Fig. 7(c) to the AC responses in Fig. 7(a), however, it can be observed that the resonance peak of the second mode at f_1, compared to that of the first

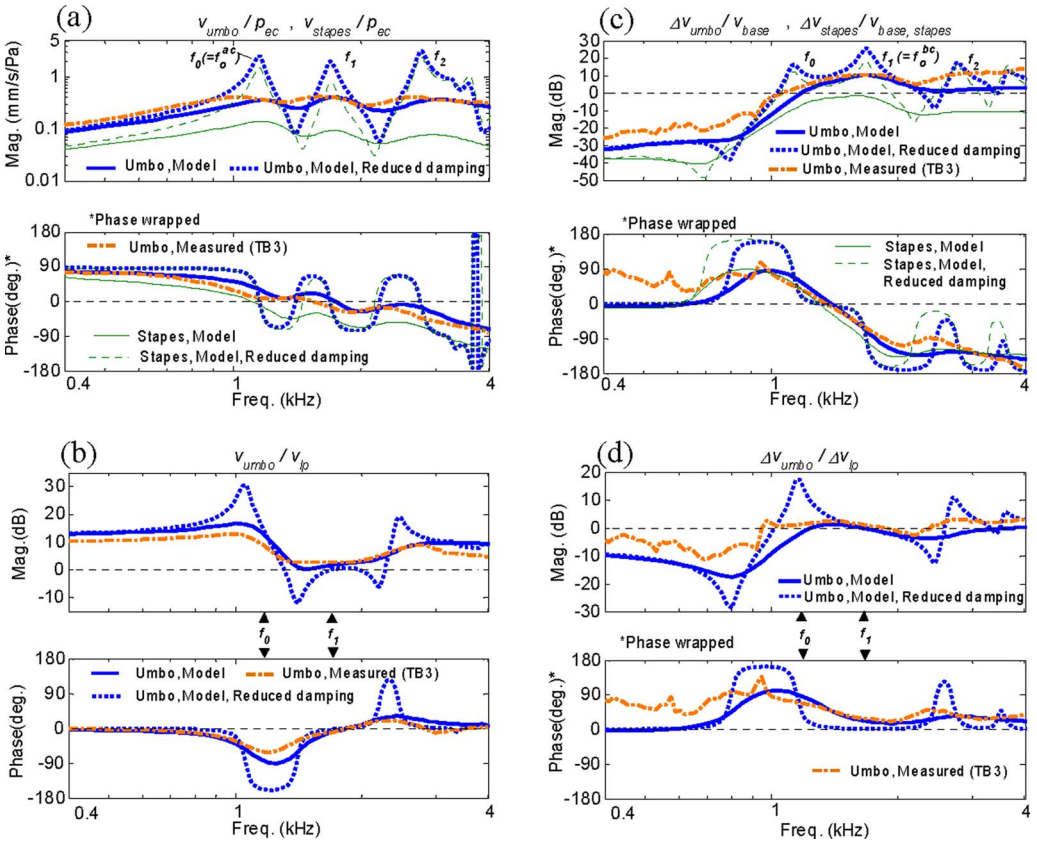


FIG. 7. (Color online) Comparison of simulated AC and BC responses (with nominal and reduced damping) and representative measured responses from a temporal bone (TB3): (a) umbo AC response, $v_{\mathrm{umbo}}/p_{\mathrm{ec}}$; (b) Umbo/LP AC velocity ratio, $v_{\mathrm{umbo}}/v_{\mathrm{lp}}$; (c) umbo BC response, $\Delta v_{\mathrm{umbo}}/v_{\mathrm{base}}$; (d) umbo/LP BC velocity ratio, $\Delta v_{\mathrm{umbo}}/\Delta v_{\mathrm{lp}}$. In addition, FE-simulated stapes responses (nominal and reduced damping) are also shown in (a) and (c).

mode at f_0, is more prominent in BC than in AC. This is particularly recognizable in the reduced damping case.

Figure 7(d) shows the umbo/LP velocity ratio for BC, $\Delta v_{\mathrm{umbo}}/\Delta v_{\mathrm{lp}}$, which also exhibits a good level of agreement between the model and the measured data. At the natural frequency of the second mode, f_1, the magnitude and phase are close to zero, which indicates the translational motion of the malleus handle associated with the second mode in Fig. 6(b). However, in contrast to the AC response case in Fig. 7(b), the umbo/LP magnitude ratio at f_0 for the BC case indicates that the umbo and LP are moving by comparable amounts, and therefore are not exhibiting the rotational malleus handle motion associated with the first mode shown in Fig. 6(a). Only when the damping is reduced, as shown by the simulation curve, does the malleus handle motion at f_0 match what is expected of the first mode, in which the umbo moves significantly more than the LP. This indicates that, for BC, the contribution of the first mode to the malleus handle motion is relatively weak, compared to that of the second mode.

IV. DISCUSSION

A. Comparisons with published middle-ear response data

Figure 8 compares the middle-ear responses obtained from the FE model with mean response data from previous temporal bone measurements from the literature. Figure 8(a) shows the magnitude and phase of the FE-simulated AC responses obtained at both the umbo and the stapes, compared with the measured mean data from Asai *et al.* (1999) with sample size $N=22$, and Gan *et al.* (2004) with $N=8$. [Magnitude responses are only presented for Asai *et al.* (1999) since the phase data were not reported in their study.] The figure also shows the mean of the umbo response data from the current temporal bone measurement ($N=5$). As shown in Fig. 8(a), the FE-simulated results follow the general trends exhibited by the published mean data, except that the simulated results contain detailed resonance features that are missing from the mean data. The umbo responses show especially good agreement with the mean response data for both the magnitude and the phase, in the frequency range shown. The stapes responses are also in generally good agreement, especially at frequencies below 2 kHz. However, the rates of magnitude and phase roll-off at frequencies above 2 kHz for the mean stapes response data are seemingly higher than those exhibited by the simulated result, although the overall trend is similar. The cause of this difference is currently unknown.

Figure 8(b) shows the comparison between the FE-simulated BC responses and the mean temporal bone measurement data by Stenfelt *et al.* (2002), at both the umbo and the stapes. Only magnitudes are shown in the figure since Stenfelt *et al.* (2002) did not report the corresponding phase data. As shown by the figure, the overall responses at both the umbo and the stapes are in general agreement between the FE-simulated responses and the mean data, except again that the mean data are much smoother than the FE model

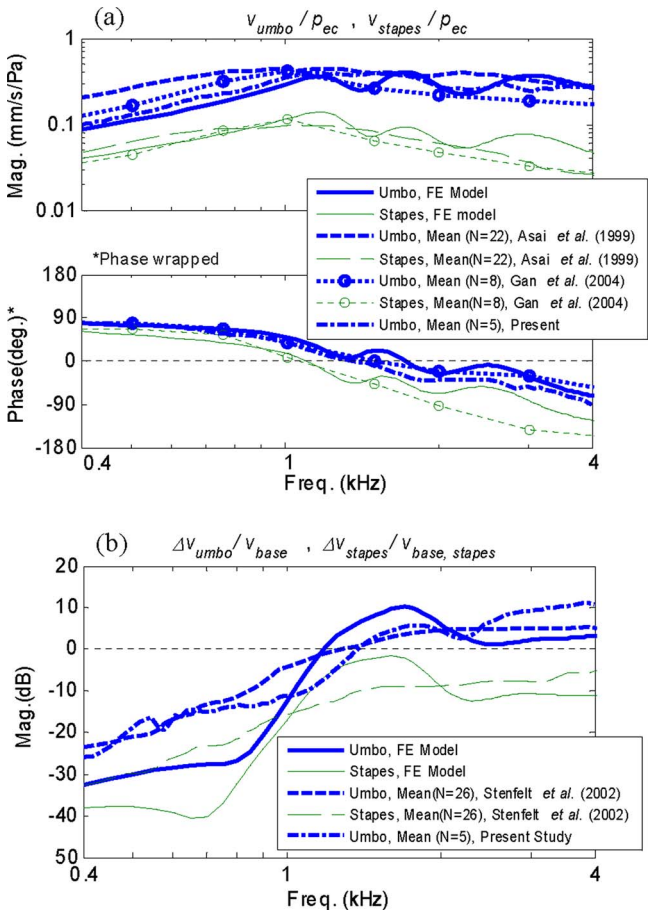


FIG. 8. (Color online) Comparison of the FE-simulated AC and BC responses with published mean temporal bone data (Asai *et al.*, 1999; Gan *et al.*, 2004; Stenfelt *et al.*, 2002): (a) AC responses at umbo and stapes (magnitude and phase); (b) BC responses at umbo and stapes (magnitude only). Measured mean umbo response data (N=5) from the present study are also shown.

response, due to averaging (N=26). The current mean temporal bone data are also seen to be consistent with the data from Stenfelt *et al.* (2002).

B. Primary resonance frequency differences between AC and BC

The mean primary BC resonance frequency, f_o^{bc}, was 1.72(±0.34) kHz for the five temporal bones, while the mean AC primary resonance frequency, f_o^{ac}, was 1.26(±0.20) kHz (Table II). The AC resonance frequency of 1.26(±0.20) kHz in this study is comparable to the general expected range of 0.8–1.2 kHz (Silman and Silverman, 1991) and is also consistent with the 1.17(±0.27) kHz value from Wada *et al.* (1998). For the mean BC resonance frequency, no previous studies are known that explicitly identify this value. However, the characteristics of the BC response data from Stenfelt *et al.* (2002), in Fig. 8, suggest a BC resonance frequency in the 1.5–2.0 kHz range.

The results obtained in this study suggest that difference in the primary resonance frequencies between AC and BC, which has not been clearly explained in the past, can be attributed to the modal characteristics of the middle ear. The temporal bone measurement data, together with the FE analysis, reveal the presence of two distinct middle-ear structural modes within the frequency range below 2 kHz. The first mode, whose natural frequency f_0 is typically at about 1.2 kHz, is recognizable as the primary AC resonance mode, and so $f_o^{ac}=f_0$. The most dominant mode in the BC case is not the first mode, but rather the second mode, whose natural frequency f_1 is typically around 1.7 kHz, so $f_o^{bc}=f_1$. Therefore, the difference in the primary resonance frequency between AC and BC can be said to result from a difference in the characteristic vibration mode that is dominantly excited by the two forms of excitation.

The reason why the second mode at f_1 is dominant in BC rather than the first mode at f_0 is not completely clear at this point and requires further investigation. It is speculated that this may be due to a difference in the degree of coupling of the BC excitation to each of the two modes. Generally speaking, a specific vibration mode is excited most effectively when the excitation vector is aligned with the direction of the eigenvector, or mode shape, associated with the mode. For the human middle-ear structure investigated in this study, the preferential vector direction of the second mode may be better matched to the vector direction of the BC excitation than the first mode. It appears intuitively reasonable that the second mode, which is characterized by a translation-like, pivoting ossicular motion, would be excited efficiently by the BC excitation given in the direction that is generally aligned with the motional direction of the mode.

Now, what if the BC excitation were given in directions other than this particular direction? It is possible that from some excitation directions the second mode may not be as efficiently excited, such as if the BC excitation were given in a direction orthogonal to the characteristic motion of the second mode. In that case, one might expect the first mode at f_0=1.2 kHz to increase in prominence relative to the second mode at f_1=1.7 kHz. However, this contradicts available evidence (Carhart, 1971; Tonndorf, 1972; Linstrom *et al.*, 2001) that indicate the peak BC resonance effect to be at around 2 kHz, which is more in line with the second mode at f_1 =1.7 kHz than the first mode at f_0=1.2 kHz. Therefore, it is unlikely that *in situ* BC excitations occur primarily in these alternate directions. However, the issues surrounding in situ BC excitation in live subjects are beyond the scope of this study and would require future investigation. In any case, the basic finding that the human middle ear exhibits a dual ossicular mode structure (i.e., the presence of the two ossicular modes) is independent of the particular type of excitation given to the system.

C. Presence of the second ossicular mode

One of the key findings in this study is the presence of the second ossicular mode at f_1, which is described as the pivoting mode in this study. In addition to being the primary resonance mode for BC, the results indicate that this mode is also excited in the AC case, resulting in the second resonance peak at $f=f_1$, just above the first peak at $f=f_0$. The presence of this mode and its role in AC and BC hearing mechanics have not been recognized in the past. Decraemer and Khanna (1994) observed that the measured malleus handle motion contains a significant translational component in addition to

a rotational component at some frequencies. Goode *et al.* (1994) also observed in their temporal bone measurements that the magnitude of the LP response approaches that of the umbo at around 1.6 kHz, indicating a translational motion of the malleus handle. The current findings suggest that the translational malleus handle motions observed in these previous studies are likely attributed to the motional contribution of the second mode.

D. Relationship to the mid-frequency BC limit peak

This study was originally motivated by a desire to understand the origin of the prominent peak feature at 1.5–2 kHz in the BC limit data (Fig. 1). The working hypothesis has been that this feature can be attributed to a structural resonance of the middle ear. The findings from the present study support this hypothesis since the middle-ear structural system is observed to resonate at around 1.7 kHz on average in response to a BC excitation, due to the presence of the second ossicular mode whose natural frequency, f_1, occurs at that frequency. However, further studies are perhaps desirable before it is possible to conclusively link the middle-ear BC resonance to the mid-frequency BC limit peak.

V. CONCLUSION

The modal dynamic characteristics of the middle ear were investigated through temporal bone measurements and FE analysis. The results indicate that the apparent difference in the primary resonance frequency between AC and BC is due to the presence of two distinct ossicular resonance modes whose natural frequencies are in the 1–2 kHz range. Although the first middle-ear mode, whose natural frequency is typically at around 1.2 kHz, is the primary resonance in AC, it is the second mode, near 1.7 kHz, that is dominantly excited in the BC response. The second mode is also excited in AC, resulting in a second resonance peak in the AC response. The FE simulation shows that the ossicular motion associated with the second mode occurs as a pivoting motion of the malleus-incus complex, with its rotational axis running through the incus ligament in an approximately superior-inferior orientation. This motion is fundamentally different from the classical ossicular hinging motion associated with the first mode. The present results indicate that this little-recognized second ossicular mode is likely the source of the translational component of the malleus handle motion that has been observed in some previous studies. The current findings also further support the hypothesis that a middle-ear structural resonance is responsible for the prominent peak feature seen at 1.5–2 kHz in the BC limit data.

ACKNOWLEDGMENTS

The authors would like to thank Dr. Stefan Stenfelt for his advice on temporal bone experiments. This work was sponsored by the Air Force Office of Scientific Research (AFOSR) under a STTR funding (Contract No: FA9550-06-C-0039).

Aibara, R., Welsh, J. T., Puria, S., and Goode, R. L. (**2001**). "Human middle-ear sound transfer function and cochlear input impedance," Hear. Res. **152**, 100–109.

Asai, M., Huber, A. M., and Goode, R. L. (**1999**). "Analysis of the best site on the stapes footplate for ossicular chain reconstruction," Acta Oto-Laryngol. **119**, 356–361.

Berger, E. H., Kieper, R. W., and Gauger, D. (**2003**). "Hearing protection: surpassing the limits to attenuation imposed by the bone-conduction pathways," J. Acoust. Soc. Am. **114**, 1955–1967.

Carhart, R. (**1971**). "Effects of stapes fixation on bone-conduction response," in *Hearing Measurement: A Book of Readings*, edited by I. M. Ventry, J. B. Chailkin, and R. F. Dixon (Appleton-Century-Crofts, New York), pp. 116–129.

Decraemer, W. F., and Khanna, S. M. (**1994**). "Modeling the malleus vibration as a rigid body motion with one rotational and one translational degree of freedom," Hear. Res. **72**, 1–18.

Fay, J. P., Puria, S., and Steele, C. R. (**2006**). "The discordant eardrum," Proc. Natl. Acad. Sci. U.S.A. **103**, 19743–19748.

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.

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.

James, M. L., Smith, G. M., Wolford, J. C., and Whaley, P. W. (**1994**). *Vibration of Mechanical and Structural Systems* (HarperCollins, New York).

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.

Linstrom, C. J., Silverman, C. A., Rosen, A., and Meiteles, L. Z. (**2001**). "Bone conduction impairment in chronic ear disease," Ann. Otol. Rhinol. Laryngol. **110**, 437–441.

McElveen, J. T., Goode, R. L., Miller, C., and Falk, S. A. (**1982**). "Effect of mastoid cavity modification on middle ear sound transmission," Ann. Otol. Rhinol. Laryngol. **91**, 526–532.

O'Connor, K. N., and Puria, S. (**2008**). "Middle-ear circuit model parameters based on a population of human ears," J. Acoust. Soc. Am. **123**, 197–211.

Reinfeldt, S., Stenfelt, S., and Håkansson, B. (**2007**). "Examination of bone-conducted transmission from sound field excitation measured by thresholds, ear-canal sound pressure, and skull vibrations," J. Acoust. Soc. Am. **121**, 1576–1587.

Shin, M., Baek, J. D., Steele, C. R., and Puria, S. (**2008**). "Stapes biomechanics: Is there an optimal stimulation axis?," Association for Research in Oto-Laryngology, Mid-Winter Meeting, Phoenix, AZ.

Silman, S., and Silverman, C. A. (**1991**). *Auditory Diagnosis* (Academic, San Diego), Chap. 3, p. 79.

Sim, J. H., Puria, S., and Steele, C. R. (**2007**). "Calculation of inertial properties of the malleus-incus complex from micro-CT imaging," J. Mech. Mater. Struct. **2**, 1515–1524.

Stenfelt, S., Hato, N., and Goode, R. L. (**2002**). "Factors contributing to bone conduction: The middle ear," J. Acoust. Soc. Am. **111**, 947–959.

Tonndorf, J. (**1972**). "Bone conduction," in *Foundations of Modern Auditory Theory*, edited by J. V. Tobias (Academic, New York), Vol. **II**, pp. 197–237.

Tonndorf, J., and Khanna, S. M. (**1972**). "Tympanic-membrane vibrations in human cadaver ears studied by time-averaged holography," J. Acoust. Soc. Am. **52**, 1221–1233.

Voss, S. E., Rosowski, J. J., Merchant, S. N., and Peake, W. T. (**2000**). "Acoustic responses of the human middle ear," Hear. Res. **150**, 43–69.

Wada, H., Koike, T., and Kobayashi, T. (**1998**). "Clinical applicability of the sweep frequency measuring apparatus for diagnosis of middle ear diseases," Ear Hear. **19**, 240–249.

Zwislocki, J. (**1957**). "In search of the bone-conduction threshold in a free sound field," J. Acoust. Soc. Am. **29**, 795–804.

Zwislocki, J. (**1962**). "Analysis of the middle-ear function. Part I: Input impedance," J. Acoust. Soc. Am. **34**, 1514–1523.

Postnatal development of sound pressure transformations by the head and pinnae of the cat: Monaural characteristics

Daniel J. Tollin[a] and Kanthaiah Koka
Department of Physiology and Biophysics, University of Colorado Health Sciences Center, Aurora, Colorado 80045

(Received 26 August 2008; revised 4 December 2008; accepted 5 December 2008)

Although there have been many anatomical, physiological, and psychophysical studies of auditory development in cat, there have been no comparable studies of the development of the sound pressured transformations by the cat head and pinnae. Because the physical dimensions of the head and pinnae determine the spectral and temporal transformations of sound, as head and pinnae size increase during development, the magnitude and frequency ranges of these transformations are hypothesized to systematically change. This hypothesis was tested by measuring directional transfer functions (DTFs), the directional components of head-related transfer functions, and the linear dimensions of the head and pinnae in cats from the onset of hearing (~1.5 weeks) through adulthood. Head and pinnae dimensions increased by factors of ~2 and ~2.5, respectively, reaching adult values by ~23 and ~16 weeks, respectively. The development of the spectral notch cues to source location, the spatial- and frequency-dependent distributions of DTF amplitude gain (acoustic directionality), maximum gain, and the acoustic axis, and the resonance frequency and associated gain of the ear canal and concha were systematically related to the dimensions of the head and pinnae. These monaural acoustical properties of the head and pinnae in the cat are mature by 16 weeks. © *2009 Acoustical Society of America.* [DOI: 10.1121/1.3058630]

PACS number(s): 43.64.Ha, 43.66.Pn, 43.66.Qp [JCM] Pages: 980–994

I. INTRODUCTION

The head and the pinnae are fundamental in shaping the spatial-location dependence of the spectral and temporal aspects of sounds that ultimately arrive at the tympanic membrane (Ruggero and Temchin, 2002; Kuhn, 1987). An important consequence of the acoustic directionality of the head and pinnae is their role in establishing the cues to sound source location. The three primary cues for location are generated by the spatial- and frequency-dependent reflection and diffraction of the propagating sound waves by the head and pinnae. Interaural time differences (ITDs) arise because the distance of the path of sound to the two ears differs. Interaural level differences (ILDs) result jointly from the amplification effects of the pinna ipsilateral to the source and the acoustic shadowing effect of the head and contralateral pinna that occurs primarily for high-frequency sounds. Finally, monaural spectral shape cues arise from differential reflection and diffraction of pressure waveforms from sounds originating from different directions by the head, torso, and pinnae.

The spatial and frequency dependences of the monaural and binaural cues to location are well documented in the adult cat (Wiener *et al.*, 1966; Middlebrooks and Pettigrew, 1981; Calford and Pettigrew, 1984; Irvine, 1987; Martin and Webster, 1989; Musicant *et al.*, 1990; Rice *et al.*, 1992; Young *et al.*, 1996; Xu and Middlebrooks, 2000; Phillips *et al.*, 1982). Moreover, the magnitudes of the cues to location and the manner in which they change with location are dependent on the physical size and dimensions of the head and pinnae (Shaw, 1974; Middlebrooks, 1999; Xu and Middlebrooks, 2000; Schnupp *et al.*, 2003; Maki and Furukawa, 2005). Interindividual differences in head and pinnae size and morphology are the basis for individual differences in the cues to location. These facts also create a challenge during development where the growing size of the head and pinnae in mammals increases dramatically from birth, thus also changing not only the magnitude of the acoustical transformations but also the relationship between the cues and sound location.

Cats have been a common model for anatomical, physiological, and behavioral studies of auditory system development [see reviews by Kitzes (1990) and Walsh and McGee (1986)]. Their auditory system is relatively immature at birth and their physical size relative to other species (e.g., rat, mouse, and gerbil) permits good access to the neural structures of interest. A wealth of knowledge exists on the anatomy, physiology, and behavior of the adult cat binaural auditory system to which developmental data can be compared (Irvine, 1986). However, aside from some spatially and spectrally sparse measurements of the development of the ILD cues in kittens by Moore and Irvine (1979) there has been no systematic study of the development of the complete complement of acoustical cues and their relationship to the development of the linear dimensions of the head and pinnae. In this paper we investigate the development of the physical dimensions of the head and pinnae in the cat from the onset of hearing through adulthood and the concomitant changes in the monaural acoustical transformations of sound pressure at the ear.

[a]Author to whom correspondence should be addressed. Electronic mail: daniel.tollin@uchsc.edu

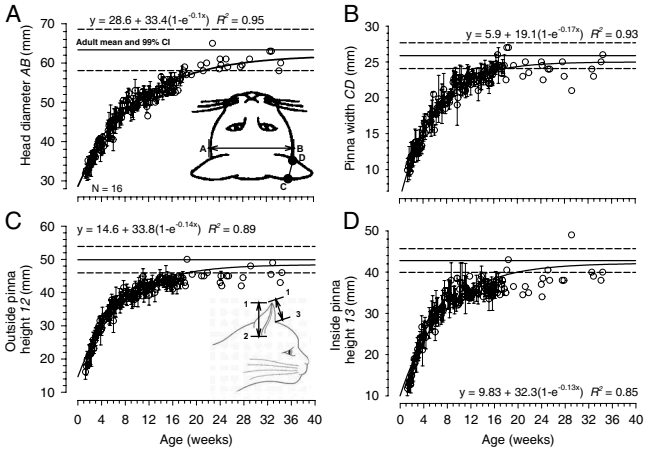


FIG. 1. Developmental growth of the head and pinnae of the cat. The four measured dimensions are shown in the insets of (A) and (C). (A) Head diameter *AB*. (B) Pinna width at half-height *CD*. (C) Outside pinna height 1-2. (D) Inside pinna height 1-3. The measured data are from 16 animals. Data points with error bars indicate the across-animal mean ± SD of the measured dimension at that age. Data points without error bars indicate single animal measurements at that age. In each panel, the solid and dashed horizontal lines indicate the mean 99% confidence interval, respectively, of the measured dimension taken from nine adult animals. The parameters of the best-fitting growth curve for each measured dimension are displayed in each panel along with the coefficient of determination (R^2) for the fit.

II. METHODOLOGY

A. Animal preparation

Twenty nine domestic short-hair cats (Liberty Research, Waverly, NY) were used in this study. Most animals were female (5/29 were male). All animals had clean external ears and ear canals. The weight and linear measurements of head diameter and pinnae height and width of each animal were taken immediately before each experiment [see insets, Figs. 1(A) and 1(C)]. A digital micrometer was used for the linear measurements. Head diameter was taken at the widest part of the head, occurring a few millimeters rostral to the meatus at the points of the zygomatic process (the bizygomatic breadth) of the temporal bone (Gilbert, 1981); this is the maximum transverse diameter of the cat head (Latimer, 1931). Pinnae heights were measured on the inside (medial, 1-3 in Fig. 1(C) inset) and outside (lateral, 1-2 in Fig. 1(C) inset) from the tip of the pinna to where the pinna edges intersected the skin covering the skull; this intersection is well defined in the cat pinna. Finally, pinna width *CD* was measured as the linear distance across the pinna opening from one edge to the other at one-half the pinna height (this does not include the depth of the pinna). In 16 (12 of which acoustic data were collected) animals, these dimensions were measured nearly every day beginning at 1 week of age up through 16–20 weeks in order to construct growth curves for these structures (Fig. 1).

The acoustic measurement procedure detailed below lasted approximately 1–2 h during which the animals remained anaesthetized and kept in areflexia. In preparation for acoustical measurements, cats were anaesthetized with ketamine hydrochloride (20 mg/kg) along with acepromazine (0.1 mg/kg). Atropine sulfate (0.05 mg/kg) was also given to reduce mucous secretions, and a tracheal cannula was inserted. Supplemental doses of sodium pentobarbital (3–5 mg/kg) were administered intravenously into the femoral vein as needed to maintain areflexia. Heart rate was continuously monitored as was core body temperature (with a rectal probe), the latter maintained with a heating pad at 37 °C (model TC 100, CWE, Inc., Ardmore, PA). Additionally, blood-oxygen levels, respiratory rate, and end-tidal CO_2 were measured continuously via a capnograph (Surgivet V90040, Waukesha, WI). A small hole was made in the wall of the posterior and ventral aspect of ear canal as near as possible to where the ear canal enters the skull (just medial to where the ear canal makes a nearly 90° turn into the skull) by advancing a slightly curved 14-gauge needle through the canal and skin from inside the canal to outside the skin so that the sharp tip of the needle is outside. A 50-mm-long (Bruel and Kjaer, part no. AF-0555, 1.65 mm outer diameter) or 76-mm-long (Etymotic ER7-14C probe tube) flexible probe tube was then inserted into the needle. The needle was then extracted so as to leave the tip of the probe tube well within the ear canal. Using an otoscope the location of the probe tube tip was adjusted by slightly extracting the probe tube until the desired location of the tip was achieved. The probe tube was fixed in place with cyanoacrylate glue.

Animals were placed on a platform in the center of the sound-attenuating room (see below) and secured in place via a custom bite bar (modeled directly after that used by Musicant *et al.*, 1990) with its interaural axis aligned in the boom of loudspeakers using three lasers, two at the poles and one at (0°,0°). No part of the bite bar/head holder interfered with the direct propagation of sound to the pinnae. However, for some source locations behind and below the animal the platform impeded the propagation of sound to the contralateral ear; data from these sources were not used. This platform and bite bar ensured the orientation of the head so that the Horsley–Clarke plane was nearly horizontal in all animals. A scaled version of the bite bar was used for the infant animals. To examine the role of the pinna, after taking the first set of acoustic measurements, the pinnae of one animal was completely removed and the measurements repeated. Removal of the pinnae did not alter the position of the probe tube microphone in the ear canal and animals remained centered in the loudspeaker boom. All surgical and experimental procedures complied with the guidelines of the University of Colorado Health Sciences Center Animal Care and Use Committees and the National Institutes of Health.

B. Experimental setup

All experiments were performed in an $\sim 3 \times 3 \times 3$ m^3 (interior dimensions) double-walled sound-attenuating room (IAC, Bronx, NY); the walls and equipment were lined with 4-in.-thick reticulated wedged acoustic foam (Sonex Classic). Stimuli were presented in one of two different experimental setups. Although we could not make any direct comparisons between the two setups, results from adult animals measured in the different setups were consistent and comparable to previous literature. In one setup stimuli were presented from 25 loudspeakers (Morel MDT-20) attached to a custom-built horizontally oriented [i.e., "single-pole" coordi-

nate system (Middlebrooks and Pettigrew, 1981; Leong and Carlile, 1998)] semicircular boom. The 25 loudspeakers were spaced in azimuth along the boom at 7.5° spacing, from −90° (left) to +90° (right). The axis of rotation of the boom was aligned with the interaural axis of the animal (i.e., through the ears). The radius of the boom was 1 m. The 25 loudspeakers were selected from a larger set (~100) on the basis of best-matching frequency responses. A stepper motor (Advanced Micro Systems AMH34-1303-3) and motor controller/power supply (Advanced Micro Systems CMAX-810) under computer control could position the boom in elevation with a precision of <1°. The semicircular boom was moved in steps of 7.5° along the elevation using the stepper motor controlled via personal computer (PC) by custom written software in MATLAB (Mathworks, Natick, MA). Stimuli were presented from a total of 625 different locations selected to evenly sample azimuth and elevation (i.e., the pole locations were not overly sampled). The elevation spanned from −45° to +225°. In some experiments, a second setup was used where stimuli were presented from nine frequency-response matched loudspeakers (Morel MDT-20) spaced in elevation from −30° to +90° in 15° steps on a fixed boom (1 m radius), the ends of which were attached to the chamber directly above and directly below the animals' head. In this setup, instead of moving the loudspeaker boom, a stepper motor was used to rotate the platform holding the animal about the interaural axis in steps of 10°. 289 different locations were sampled (8 elevations × 36 azimuths, plus the overhead position). For this latter setup, the interaural axis of the cat was placed in the center of the boom supporting the speakers using the system of three lasers as described earlier. This centering was important as the cat was then rotated about this point. As one test of the validity of the recordings at each rotational angle, the responses to the speaker placed directly overhead were compared. If the cat was perfectly aligned the acoustic responses should not depend on the rotation angle since the two ears would always be equidistant from the overhead speaker. For animals tested in this setup, the acoustic responses measured from this position differed little with azimuth; the standard deviation in acoustic gain at each frequency was <1.0 dB versus 10–25 dB at 0° elevation. In order to treat the data collected in each of these two setups in the same way, all data were expressed in a vertical pole coordinate system in which "azimuth" is the angle around the vertical axis and "elevation" is the angle above or below the horizontal plane.

The general measurement stimuli consisted of 11th order maximum length sequences (MLSs) (Rife and Vanderkooy, 1989) repeated without interruption 128 times from each loudspeaker. The MLS was presented at full 24 bit resolution at a rate of 97 656.25 Hz [Tucker-Davis Technologies (TDT), RP2.1, Alachua, FL]. In some of the very first experiments using the second setup described above, the stimuli consisted of trains of 10.24 μs clicks presented 300–1000 times (e.g., Musicant *et al.*, 1990; Rice *et al.*, 1992). In both setups, loudspeakers for stimulus presentation were selected via two daisy-chained TDT multiplexers (TDT PM2R power multiplexer) and the stimulus amplified (TDT SA1 stereo power amplifier) before being presented to the loudspeaker. The resulting acoustic waveforms in the ear canals of the left and right ears were simultaneously recorded through two probe tube microphones (Bruel and Kjaer, type 4182), amplified (TDT MA2), and collected at 24 bits using two analog to digital converter channels at 97 656.25 Hz (TDT RP2.1). All the stimulus presentation, acquisition, analysis, and movement of the speaker boom were controlled by custom written software in MATLAB. The recorded signals were stored on a PC hard disk for later processing. In all experiments a calibration measurement was also made in the absence of the animal by placing the tips of the probe tubes so that they corresponded to the location where the center of the head of the cats would be located. The calibration measurements capture the spectral characteristics of the loudspeakers and microphones for later processing.

C. Data processing and data analysis

For data collected in the first experimental setup, the impulse response for each ear and each location was calculated by circular cross-correlation of the original 11th order MLS stimulus and the in-ear recording from the probe tube microphone (Rife and Vanderkooy, 1989). For data collected in the second setup, the impulse response for each ear and location was computed from the average of the responses to the train of clicks. In both cases, the impulse responses were then truncated to 512 points (5.12 ms duration) by a 512-point Hanning window where the center of the window was set to coincide approximately with the point of maximum amplitude in the impulse response. This windowing procedure removes the small-amplitude reflections that may be contained in the impulse response. Next, the head related transfer functions (HRTFs) were derived by dividing the frequency response of the in-ear recording by that of appropriate loudspeaker calibration measurement. This procedure removes the loudspeaker and microphone frequency response from each in-ear measurement. The resulting function is referred to as the HRTF, as it represents the acoustical gain and delay introduced by the head and the pinnae. However, the resulting HRTF can be highly dependent on the exact placement of the tip of the probe tube microphone in the ear canal relative to the tympanic membrane (Middlebrooks *et al.*, 1989). To reduce the confounding effects of the probe tube placement in the ear canal, for each ear the directional transfer functions (DTFs) were then calculated from the HRTFs by dividing the HRTF made at each spatial location by the geometrical mean of all the measured HRTFs across all measurement locations for that ear. The spectral features resulting from the exact placement of the probe tube microphone in the ear canal are expected to be similar for all measurement locations (i.e., they are not dependent on spatial location), so this "common" spectral feature is removed from the HRTFs, resulting in the DTFs (Middlebrooks and Green, 1990). In essence, the DTFs are the sound source direction-dependent components of HRTFs.

The amplitude spectra of the DTFs were calculated after the spectra were passed through a bank of 351 bandpass filters; the center frequencies of which were spaced at intervals of 0.0143 octave spanning from 1 to 32 kHz. The 3 dB

TABLE I. Pearson correlation coefficients. $N=18$ adult cats. Bold and italic values indicate correlation coefficients that were significant at $p<0.01$ and $p<0.05$, respectively. Values for *AB*, 1-3, 1-2, and *CD* correspond to the head and pinna dimensions, as shown in Fig. 1.

	Weight	Head *AB*	Pinna 1-3	Pinna 1-2	Pinna *CD*
Weight	1.00				
Head *AB*	0.41	1.00			
Pinna 1-3	0.39	*0.48*	1.00		
Pinna 1-2	**0.62**	0.45	**0.79**	1.00	
Pinna *CD*	**0.66**	0.36	**0.66**	**0.70**	1.00

bandwidth of filters was held constant across all frequencies at 0.12 octaves, and the upper and lower slopes of the filters fell off at ~105 dB/octave. These filters have properties similar to the bank of bandpass filters that have been used elsewhere to filter DTFs (Middlebrooks, 1999; Xu and Middlebrooks, 2000; Schnupp *et al.*, 2003). Only data up to 32 kHz were used here as the signal-to-noise ratio was poor for higher frequencies in some animals.

For spatial plotting purposes, the data were displayed as Mollweide projections. In each of these projections the nose of the animal is considered to be projecting out of the page at 0° azimuth and 0° elevation, as if the animal were looking at the reader. The Mollweide projections were plotted for elevations from −30° to +90° and all azimuths from −180° to +180°.

III. RESULTS

Results are based on data from 29 animals (5/29 were male). Acoustical measurements were obtained in 20 of these animals, 9 of which were adults and 11 were animals at different ages ranging from 1.3 to 22.1 weeks (for convenience, the age in days was converted to weeks by dividing by 7 and the quotient rounded to the nearest 1/10th of a week. For example, 9 days divided by 7 is equivalent to 1.3 weeks). The latter 11 animals came from five different litters. Nine additional adults were used only for measurements of head and pinnae dimensions. Adult acoustic measurements were obtained in conjunction with physiological experiments that utilized the acoustical measurements for virtual space stimulus presentation. In this paper we show detailed data from three animals from different age groups spanning development: 1.3 weeks (K009), 5 weeks (K008), and adult (Adult). Summary data, when shown, were computed from all animals.

A. Growth of the head and pinnae

Figure 1 shows linear measurements of head diameter *AB*, pinna width *CD*, and inside 1-3 and outside 1-2 pinnae height as a function of age in weeks for 16 cats starting at 1.3 weeks. The horizontal lines represent the mean and 99% confidence interval for these values in 12 adults (>52 weeks). To quantify the growth rate a three-parameter exponential rise to maximum function was fitted to the data of the form $y=y_o+a(1-e^{-bx})$, where x is the age in weeks, y_o is the extrapolated dimension at birth (0 week), a is the amount by which that dimension increases during development, (y_o+a) is the asymptotic value at full development, and b is the rate of growth. This equation accurately characterized the growth of each dimension (based on *F -test* $P<0.0001$ for all fitted equations; coefficients of determination R^2 are reported on the figure). Solving the growth equation for the age parameter x, we can compute the age at which the curve reaches any arbitrary proportion of the maximum value; given a proportion (e.g., 0.9), the equation reduces to $-\ln(1\text{-proportion})/b$. We use a proportion of 0.9 (90%) to indicate "adult" values because this proportion produced values of the dimensions that first fell within the 99% confidence intervals for adult dimensions. The fitted parameters of the equations are shown in each panel in Fig. 1.

Based on the fitted growth curve, head diameter increases from 28.6 mm at birth to 62 mm, reaching 90% of adult value by 23 weeks. In contrast, the growth of the pinnae was much more rapid. Inside pinnae height 1-3 increased from 9.8 to 42.1 mm, reaching 90% at 18 weeks. Outside pinnae height 1-2 increased from 14.6 to 48.4 mm, reaching 90% at 16 weeks. And pinnae width *CD* increased from 5.9 to 25.0 mm, reaching 90% at 14 weeks. Bodyweight (not shown) increased from 0.2 ± 0.06 kg at ~1.5 weeks and asymptotes at 2.6 ± 1.1 kg by ~18 weeks. Although not shown as a figure, we also measured the interocular distance (i.e., the distance between the center of the pupils) resulting in the equation $y=13.3+20.4(1-e^{-0.11x})$ ($R^2=0.83$, $n=16$). Interocular distance reached 90% of adult size by 21 weeks, which was comparable to the growth of the head.

During development, the dimensions of the head and pinnae were highly and significantly correlated with the weight and age of the animal (mean $R^2=0.91\pm0.05$, $n=15$ pairwise comparisons). These correlations were not surprising and not particularly useful. However, as shown in the correlation matrix in Table I, even in a group of 18 adults we found significant correlations: outside pinna height 1-2 and pinna width *CD* were significantly correlated with weight ($p<0.01$), but the head diameter *AB* was not. We did not track gender differences in development because the five males in the study were utilized for the acoustical measurements at very young ages.

B. Frequency range and spatial-location dependence of broadband spectral notches

We observed a systematic change in the frequency of the first (i.e., lowest frequency) broadband spectral notch [i.e., first notch frequency (FNF), Rice *et al.* (1992)] with changes in source location in the frontal hemisphere in animals of all

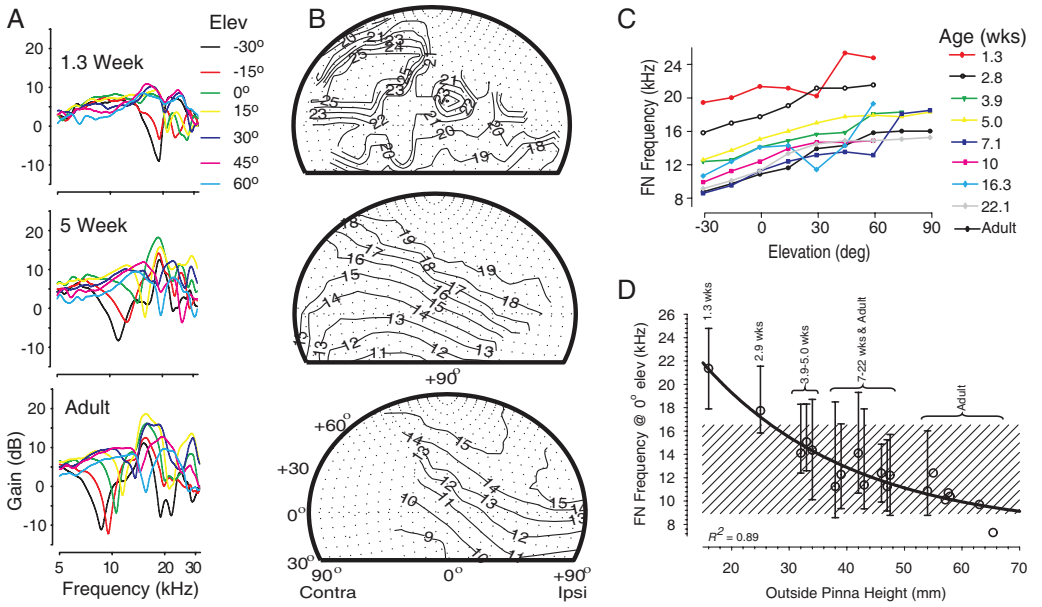


FIG. 2. Development of the monaural broadband spectral notch cues. (A) DTF gain for seven different elevations from −30° to 60° at 0° azimuth for cats of three different ages (upper left in each panel). (B) Plots of the isofrequency contours of the first (lowest frequency) notch frequencies, or FNF, for sources in the frontal hemispheres for the same three animals as in (A). (C) FNFs plotted as a function of elevation at 0° azimuth for animals of nine different ages. (D) Development of the first notch frequency range as a function of the development of the outside pinnae height 1-2 (n=18 animals). Symbols indicate FNF at (0°,0°) while the error bars indicate the range of FNFs observed in the frontal hemisphere. Solid line indicates the best-fitting function relating FNFs at (0°, 0°) to pinna height 1-2. Hatched region indicates the range of FNFs observed in the frontal hemisphere in adult animals.

ages. The depths of the spectral notches were ~10–15 dB in the youngest animals increasing to 15–20 dB in adults. Figure 2(A) shows the DTFs for the right ear (ipsilateral ear, +90°) of three animals of different ages, 1.3 weeks, 5 weeks, and adult, for elevations ranging from −30° to 90° in 15° steps at 0° azimuth. Prominent broadband spectral notches were observed for most source locations in the frontal hemisphere particularly toward the ipsilateral ear but for different frequency ranges depending on source location and age [Fig. 2(B)]. The FNF was easily detectable and moved systematically with elevation and azimuth for sources in the frontal hemisphere [e.g., Figs. 2(A) and 2(B)], except in the youngest animal (1.3 weeks). The iso-FNF contours [Fig. 2(B)] reveal that as the source was moved up in elevation and in azimuth toward the ipsilateral ear, the FNFs generally increased. However, for a given source location, FNF decreased with age [Figs. 2(B) and 2(C)]. For example, for a source at (0°,0°), FNFs were 20, 14, and 10.5 kHz at 1.3 weeks, 5 weeks, and adult. This developmental aspect of FNF is illustrated in Fig. 2(C) where FNFs at 0° azimuth and changing elevation for animals at nine ages are shown. At all ages, FNF increases with elevation, but the range of FNFs shifts progressively lower during development.

Given the developmental changes in the linear dimensions of the pinnae [Figs. 1(B)–1(D)] and the finding that the FNFs for a given spatial location decrease systematically with age, we hypothesized that the developmental change in FNF range is determined by the development of pinnae size. To test this hypothesis, Fig. 2(D) shows the FNF for a single source location (0°,0°) as a function of the outside pinnae height 1-2 in 19 animals (8 were adults); the error bars show the overall range of detectable FNFs observed in the frontal hemisphere for 13 of the animals used in the developmental phase of the studies. The remaining data points for adults had FNFs in encompassing a range from ~8 to 16 kHz [hatched area, Fig. 2(D)], which is typical of FNF ranges reported for adult cats (Musicant *et al.*, 1990; Rice *et al.*, 1992; Xu and Middlebrooks, 2000). The data show that FNFs did not fully encompass the adult range in all animals until sometime after 10 weeks (Figs. 2(C) and 2(D)). Recall in Fig. 1(C) that the outside pinnae dimension 1-2 reached 90% of adult size, or ~44 mm, by ~16 weeks. Figure 2(D) shows that the ranges of FNFs observed in animals with pinnae dimensions >44 mm were all adultlike while those <44 mm were not. A three-parameter exponential decay function accurately describes the development of the FNF at 0° as a function of the outside pinnae height 1-2 (R^2=0.89, P<0.0001, and y =7.42+25.82$e^{-0.039x}$) supporting the hypothesis that the linear dimensions of the pinnae determine the FNFs. In further support, in one animal (K013, 2.86 weeks) the pinnae on both sides were removed and the acoustic measurements repeated. Pinnae removal eliminated the spectral notches in both ears (not shown). The spectral notch cues are adult by 16 weeks consistent with the development of the pinnae.

C. Spatial distribution of DTF amplitude gain

The DTF gain at a given frequency varied with source direction and with the age of the animal. Figure 3 shows the distribution of DTF gains for 1, 2, 4, 8, 12, 16, and 20 kHz for sources in the frontal hemisphere for right ears (i.e., −180° is ipsilateral) for the same three animals as in Figs. 2(A) and 2(B). Here DTF gain is plotted only for the right ear as it was almost mirror symmetric with that of the left ear in most of the animals. Each of the gain plots was normalized by the maximum DTF gain (upper left side of each

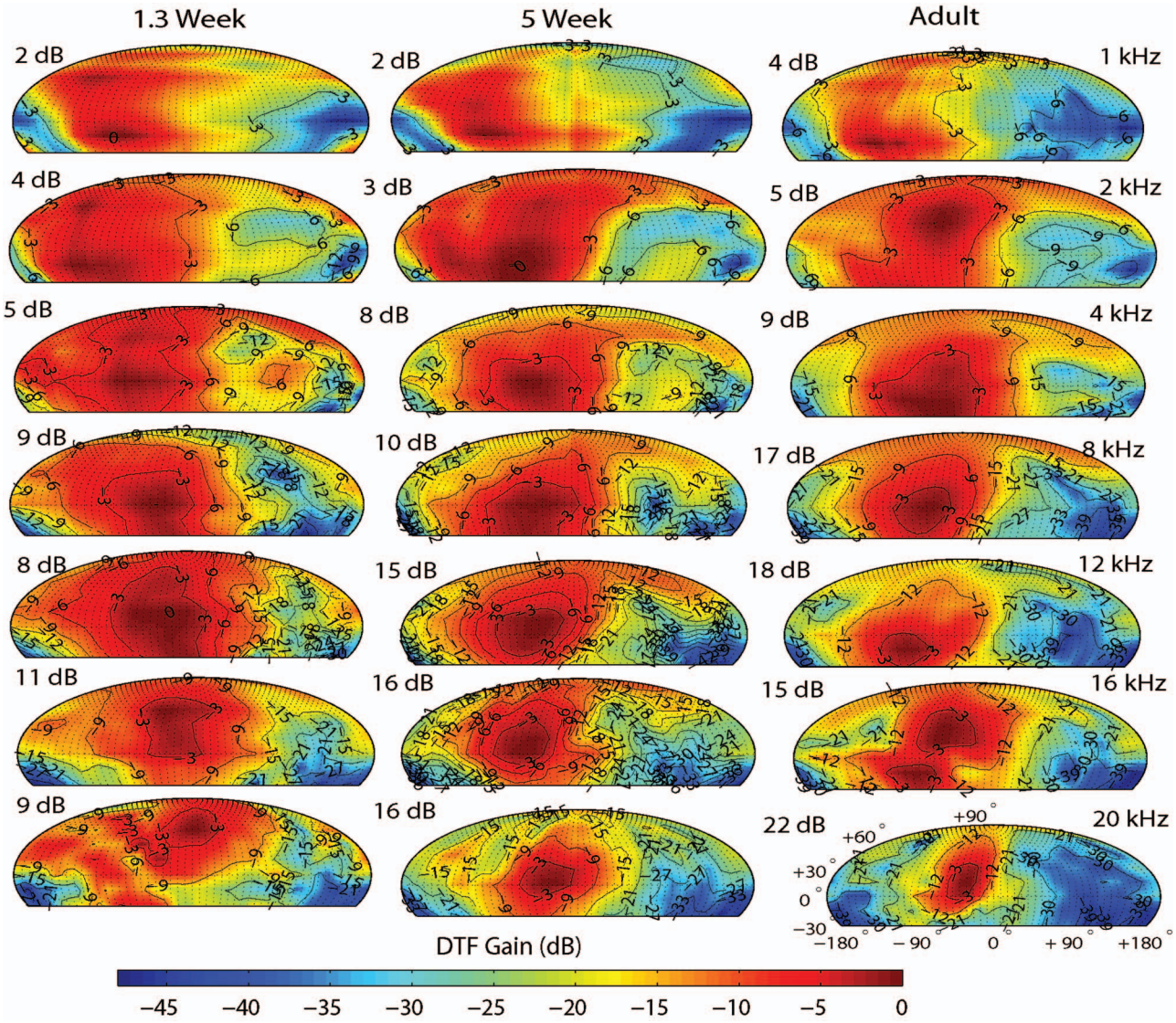


FIG. 3. Spatial distribution of DTF gains for seven frequencies for the right ears (e.g., −90° is ipsilateral) of three animals aged 1.3 weeks, 5 weeks, and adult. DTF gains for each animal have been normalized by the maximum gain (indicated at the upper left of each panel) at the indicated frequency (upper right of panels in last column). The contours are plotted at −3 dB intervals from the maximum gain. Color bar (bottom) indicates the relative gain with respect to the maximum gain.

panel) observed at that frequency and −3 dB contours were plotted. For a given frequency, the maximum gain tended to increase and the area of gain (i.e., the area bound by the −3 dB contour) tended to decrease with the age of the animal.

Figure 4(A) plots the maximum gain in the DTFs (which does not include the canal resonance gain) as a function of frequency for animals of three different ages from Fig. 3. Maximum gain occurs at the acoustic axis for a given frequency. In all animals, the maximum gain tended to increase with frequency, at least up to ~20–30 kHz. However, for frequencies >5 kHz there was a systematic increase in the maximum gain with age. In adults, maximum gains of 20–25 dB were achieved and occurred at higher frequencies. In the youngest animals, maximum gains only achieved 10–15 dB over this same range of frequencies. The gain curves appear to be simply shifted toward lower frequencies with age, consistent with the increase in the linear dimensions of the pinnae and head (Fig. 1). Since DTFs were computed in this paper, any acoustical gains (e.g., ear canal resonance) that were nondirectional were removed from the data. Estimates of canal and concha gain and resonance frequency and their development are discussed in Sec. III F.

1. Acoustic directionality of the head and pinnae and relationship to pinnae dimensions

The spatial distributions of DTF gain (Fig. 3) were quantified in two ways. The first measured the acoustic axis, the spatial location for a given frequency of highest acoustical gain, which will be reported in Sec. III D. Second, the directionality of the pinnae and head was obtained by computing the solid angle contained by the −3 dB contour at frequencies from 2 to 32 kHz in 1 kHz frequency steps. Figure 4(B) shows the −3 dB contour area (in π sr; here 2π sr equals $\frac{1}{2}$ hemisphere) with frequency for left and right ears for the three animals of different ages in Figs. 2 and 3. For frequencies <~4–5 kHz, the area covered more than

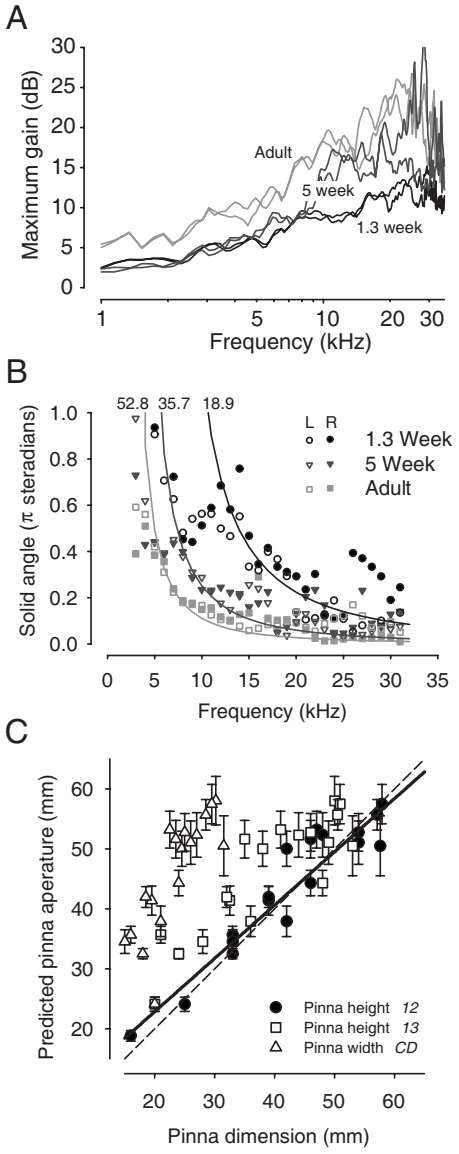


FIG. 4. (A) Maximum acoustical gain of the head and pinnae as a function of frequency for animals aged 1.3 weeks, 5 weeks, and adult. (B) Solid angle area (in units of π sr) enclosed by the -3 dB contour from the DTF gain plots (Fig. 3) as a function of frequency. Data are shown for the left (open) and right (filled) ears of the same animals as in (A). Solid lines indicate the best-fitting circular aperture model (see text) to the data corresponding to each animal; the aperture diameter from the best-fitting function is indicated at the top of each line. (C) Scatter plot of the predicted aperture diameter from the circular aperture model fitted to the data as a function of the pinnae dimension in 19 animals. Data are shown for the three pinnae dimensions: *CD*, 1-2, and 1-3 in Figs. 1(B)–1(D), respectively. Error bars show 95% confidence intervals. Solid line shows linear regression of predicted aperture and pinnae height 1-2. Dashed line indicates line of equality.

one hemisphere (2π sr, not shown in figures). At higher frequencies, the -3 dB contour area systematically decreased with frequency for all animals.

We hypothesized that for a given frequency the solid area enclosed by the -3 dB contour should decrease with the age of the animal because the linear dimensions of the external ear increase during development [Figs. 1(B)–1(D)]. To test this hypothesis the equation describing the frequency dependence of the solid angle for a -3 dB contour for acoustic diffraction through a circular aperture for a given aperture diameter [equation derived in Calford and Pettigrew (1984) and Coles and Guppy (1986)] was fitted to the acoustical data (e.g., Fig. 4) for 19 animals (8 adults and 11 at different ages). In the fitting the aperture diameter was the only free parameter (MATLAB Version 7.1 robust nonlinear least-squares trust-region method). We noticed that the empirical data at lower frequencies were poorly fitted by the model, so only data for solid angles $<\sim 0.5$ were used for the fitting resulting in frequencies $>\sim 7$–15 kHz (depending on the age of the animal) although all empirical data are shown in Fig. 4(B). The best-fitting functions describing the solid angle are shown for three animals in Fig. 4(B); the predicted diameters of the circular aperture for these animals were 18.9, 35.7, and 52.8 mm, respectively (R^2 equals 0.63, 0.5, and 0.5, respectively). Coefficients of determination (R^2) for the population averaged 0.52 ± 0.19 with a median of 0.5 and a range of 0.18–0.85. In cases where the fit was poor, the fitting error tended to occur at the lower frequency end of the curve [e.g., Fig. 4(B), 1.3 week].

There were some data points in Fig. 4(B) that deviated substantially from the predictions. The deviations occurred at low frequencies for all animals, at high frequencies for the 1.3 week animal, and $\sim$15 and 25 kHz in adults. These deviations occur due to a splitting of the spatial area of highest gain, the acoustical axis. Examples of this splitting are beginning to be apparent at 20 kHz in the 1.3 week animal and at 16 kHz in the adult in Fig. 3. This phenomenon has been reported previously in adult cats for frequencies around 14 and around 28 kHz, consistent with our findings here (Musicant *et al.*, 1990).

To test the hypothesis that the increasing linear dimensions of the pinnae during development determine the directionality as computed by the -3 dB area as a function of frequency [e.g., Fig. 4(B)], the predicted pinnae aperture from the fitted function was plotted in Fig. 4(C) as a function of the three empirically measured linear dimensions of the pinnae [Figs. 1(B)–1(D)]. The pinna dimension that described most of the variance in the predicted aperture diameter was the outside pinnae height 1-2; the linear regression of predicted pinnae aperture diameter on the empirical pinna height 1-2 was significant ($R^2=0.92$, $P<0.0001$, $y=5.1+0.89x$, and $n=19$). The linear regressions with the other pinnae dimensions were also significant ($P<0.0001$), but the coefficients of determination and the slopes of the regression functions were not as close to 1.0 (R^2 and slope 0.87 and 0.85, and 0.77 and 1.75 for pinnae dimensions 1-3 and *CD*, respectively). These data support the hypothesis that the linear dimensions of the pinnae, in particular, the outer height 1-2, determine the development of the frequency dependence of the -3 dB area of the acoustical directivity.

2. Broadband head and pinnae directionality

Figure 5 shows the directionality of the head and pinnae computed across the entire stimulus bandwidth considered here (1–30 kHz). Note that because DTFs were computed here, the gain at frequencies $<\sim 2$ kHz was quite small (<5 dB), as can be seen in Fig. 3. Thus, the broadband gain shown here is largely due to the gains for frequencies >2 kHz. Figure 5(A) shows the spatial distribution of broadband directionality and the associated -3 dB area (in

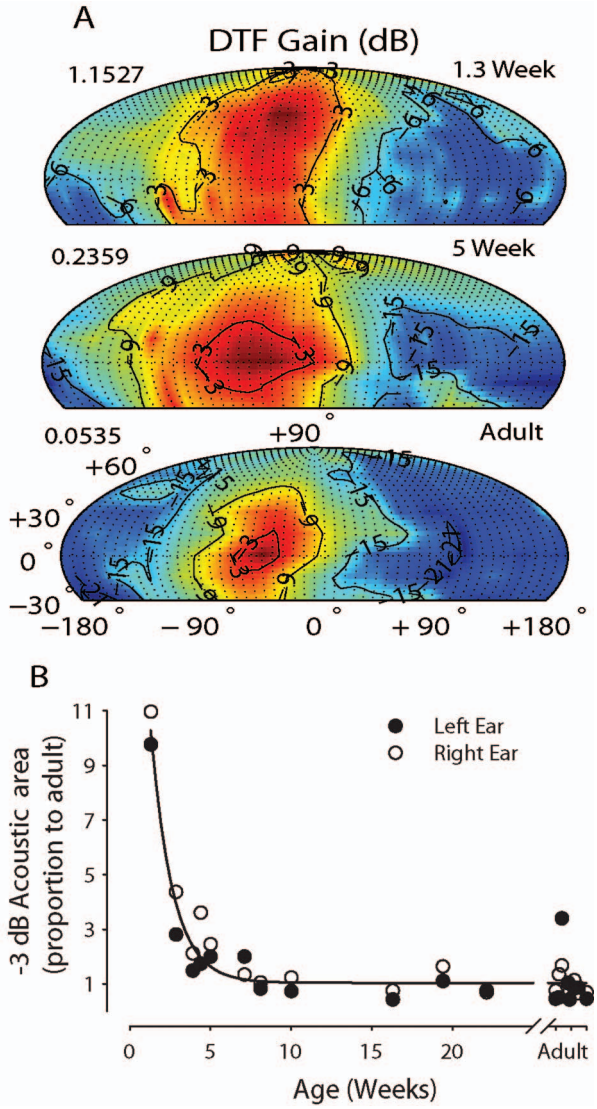


FIG. 5. (A) Broadband DTF gain for animals aged 1.3 weeks, 5 weeks, and adult. Each plot has been normalized by the maximum DTF gain and −3 dB contours with respect to the maximum gain are plotted. The area (in units of π sr) enclosed by the −3 dB contour is indicated in the upper left of each panel. (B) The development of the −3 dB area as a function of the age of cats for the left (filled symbols) and right (open symbols) ears (n=19 animals). The −3 dB area has been normalized by the average broadband gain area in 8 (8/19) adult animals. Solid line indicates best-fitting function to all of the data (see text).

π sr, upper left) for animals of three different ages corresponding to those in Figs. 2–4. Figure 5(B) summarizes the −3 dB spatial area for the left and right ears for 20 (9/20 were adults) animals as a function age. In Fig. 5(B) the −3 dB areas were normalized by the average area obtained in nine adults. These data can be thought of as a general characterization of the development of head and pinnae acoustic directionality with age (see Mrsic-Flogel *et al.*, 2003). A three-parameter exponential was fitted to the data (R^2=0.91, $P<0.0001$, and $y=1.04+24.3e^{-0.74x}$). The −3 dB area decreased from $\sim\frac{1}{4}$ of the frontal hemisphere ($\sim 1\pi$ sr) at 1.3 weeks by a factor of 10 during development and fully asymptotes to adult values (i.e., 1.0) by ∼14 weeks. Recall that by 16 weeks, the outside pinnae height 1-2 was 90% of adult size. The computation of the broadband acoustic directionality was motivated in part by some experimental data on the development of neural coding of acoustical space in the cat and other species, which will be detailed in Sec. IV.

D. The acoustical axis

The spatial location of the DTF gain maximum (and minimum) varied as a function of frequency. The direction of maximum acoustical gain at a given frequency is known as the acoustical axis (Middlebrooks and Pettigrew, 1981; Phillips *et al.*, 1982). Figure 6 shows the spatial location of the acoustic axis for azimuth (top) and elevation (bottom) as a function frequency for one ear in animals of four different ages (age indicated in upper left). Two general patterns emerged upon examination of the acoustic axis for these animals as well as the others. First, in all except the youngest animal (1.3 weeks, Fig. 6), with increasing frequency the location of the axis in azimuth revealed patterns that tended to move from medial to lateral locations interspersed by discrete jumps back toward the midline. Vertical dashed lines mark the approximate frequency of these transitions in Fig. 6. In the adult, the azimuth of the acoustic axis begins at ∼40°–50° at 3 kHz, then makes an abrupt transition toward the midline at ∼14 kHz, moves laterally again to ∼50° by 25 kHz, with yet another transition toward midline at 28 kHz. Qualitatively similar patterns were seen in each of the younger animals. For example, the data for the 7.1 and 5 week animals [see also data for 3.9 weeks in Fig. 8(A)] show similar movements and transitions of the azimuthal acoustic axis, but the transitions occurred at systematically higher frequencies than the adult. The movements of the acoustic axis in the youngest animal (1.3 weeks) were different than for the older animals; here the azimuthal acoustic axis moved from lateral to medial and with abrupt transitions back toward lateral locations.

The acoustic axis in elevation in all animals tended to move from lower to higher elevations with increases in frequency that were interrupted by transitions again to lower elevations. For example, in the adult, the elevation axis increases from −30° to 30° for frequencies from 4 to 12 kHz, and then shifts back down to 0° by 14 kHz increasing to +30° or more by 20 kHz. In the 5 week animal, the first downward transition occurs at ∼17 kHz while in the youngest animal (1.3 weeks) this transition does not appear to occur until ∼28 kHz. By 7.1 weeks, the first transition occurs at ∼14 kHz, which was comparable to adult values. These transitions are indicated with vertical dashed lines. The elevation acoustical axis was more difficult to interpret in the youngest animals.

The frequencies where the transitions occurred in the azimuth and elevation acoustic axes revealed a rough correspondence with the frequency ranges where the first spectral notch cues occurred (Fig. 2). Recall that the ranges of FNF move systematically toward lower frequencies during development [Fig. 2(D)] and so too do the acoustic axes transition frequencies (Fig. 6).

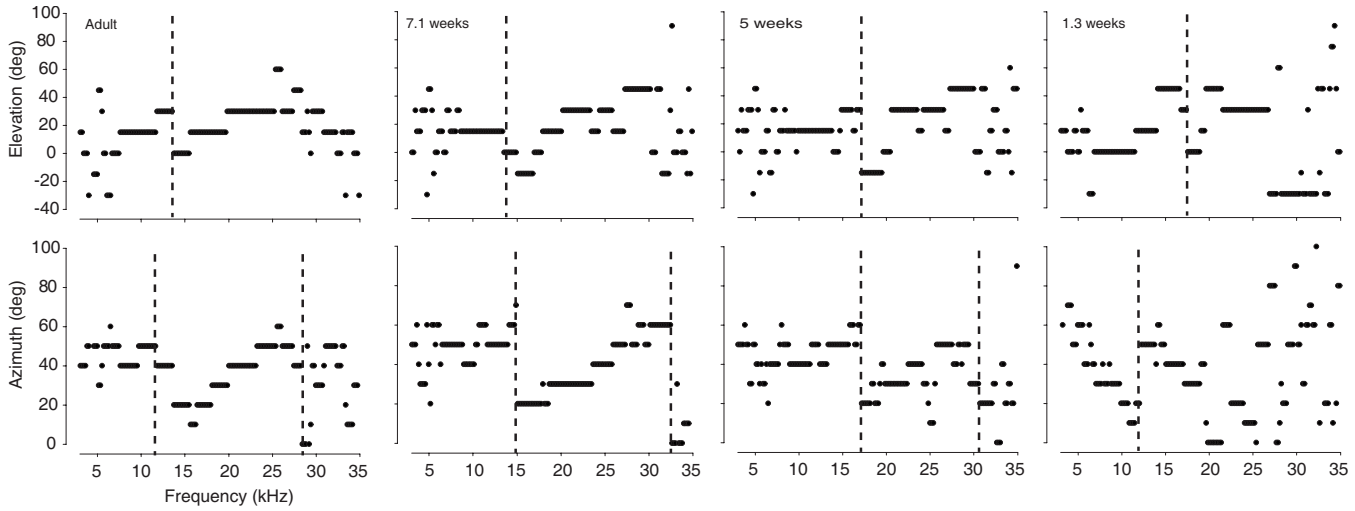


FIG. 6. The elevation (top panels) and azimuth (bottom panels) corresponding to the acoustic axis as a function of frequency in four animals aged 1.3, 5, and 7.1 weeks as well as adult (age indicated in upper left of each panel). The acoustic axis is the spatial location corresponding to the maximum DTF gain [Fig. 4(A)] at a particular frequency (see Fig. 3). Vertical dashed lines indicate frequencies where discrete transitions appear to occur in the acoustic axis.

E. Contribution of the pinnae to DTF gain, directivity, and acoustic axis

1. Spatial directivity

While it is typically assumed that the pinnae are a major determinant of the spatial and frequency dependences of acoustical gain, directivity, and the acoustic axis, few experiments have actually tested this hypothesis directly (see Koka *et al.*, 2008). None of these studies has examined the role of the pinnae in developing animals. Here, in one animal (K013, 2.9 weeks) we assessed the contribution of the pinnae in the developing animal to the DTF gain and the associated directivity by measuring DTFs before and after the removal of both pinnae. Figure 7(A) shows for the right ear (−180° corresponds to the ipsilateral ear) the spatial pattern of DTF gain for three frequencies with (left) and without (right) pinnae. The maximum gain for these conditions is listed at the top left of each panel. Two major findings were apparent. The pinnae increased the overall acoustical gain by as much as 6 dB for the example frequencies and also vastly increased the acoustical directivity. This latter finding is illustrated in Fig. 7(B) where the solid angle encompassing the −3 dB contour for the left and right ears is plotted as a function of frequency. Predictions of −3 dB area based on the circular aperture model (see Sec. III C) are plotted for aperture diameters of 40, 15, and 10 mm. The −3 dB area prediction corresponding to a 24.13 mm aperture is the best-fitting function to the data in the intact animal (R^2=0.7; although note the substantial deviations from the prediction < ~10 kHz). The 24.13 mm predicted value compares favorably with the 25 mm empirical measurement of the outer pinna height 1-2 for this animal [e.g., Fig. 4(C)]. For virtually all frequencies, the solid angle was substantially increased upon removal of the pinnae [Fig. 7(A)]. Data from the pinnaeless animal could not be accounted for by the circular aperture model. This result shows that the pinnae in developing cats are critical for increasing the acoustical gain and the overall acoustic directivity.

2. The acoustic axis

The pinnae are a major determinant of the acoustic axis, particularly at high frequencies. Figure 7(C) shows the acoustic axis in elevation (top) and azimuth (bottom) with and without the pinnae for the left ear. Results for the right ear were comparable. Changes in the elevation axis were not readily apparent. The axis in azimuth was substantially altered by pinnae removal with the axis located predominantly toward the far lateral azimuths. Moreover the pattern of acoustic axis movement with frequency was altered. After pinnae removal the axis moved from lateral to medial azimuths, beginning ~17 kHz with an abrupt transition again to lateral azimuths at ~23 kHz. In the pinnae intact condition, the azimuth axis moved from medial to lateral with a transition back to medial azimuths at ~25 kHz consistent with general trends shown in Fig. 6 for developing animals.

3. Acoustical gain due to the pinnae

The pinnae contribute substantially to the overall acoustical gain. Figure 7(D) shows the maximum gain for the left and right ears as a function of frequency both with and without the pinnae. The maximum gain occurs at the acoustic axis. With the pinnae, the maximum gain increased systematically with frequency, approaching 25 dB at high frequencies. After pinnae removal, the gain also increased with frequency, but much less so than with the pinnae. The gain due to the pinnae was computed by taking the difference of the gains produced with and without the pinnae. In this young animal, the pinnae do not become directional until ~8 kHz. Above this, the maximum gain increased with frequency reaching a maximum 9.6 dB at 16.4 kHz, then falling somewhat for higher frequencies.

F. Development of ear canal and concha resonance and gain

Because DTFs were computed in this paper, any acoustical gains that were nondirectional are not present in the

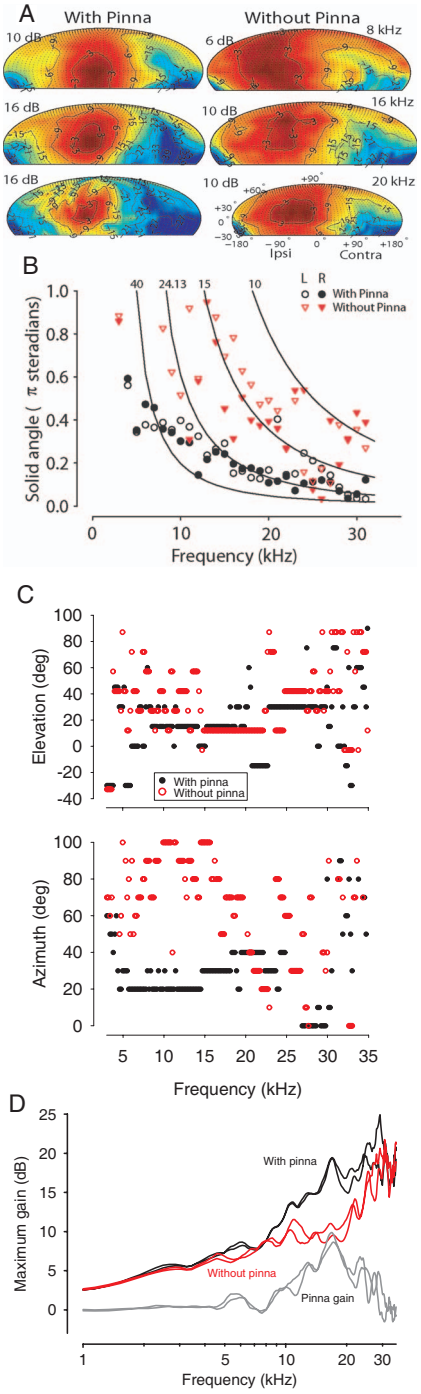


FIG. 7. The contribution of the pinnae to the spatial distributions of DTF gain (A), the frequency dependence of the −3 dB area of the DTF gain (B), the azimuth and elevation of the acoustic axis (C), and the maximum gain as a function of frequency (D). All data are from one animal aged 2.9 weeks. (A) DTF gain for three frequencies (upper right, left panels) with (left column) and without (right column) the pinnae. Maximum gain indicated in upper left of each panel. Axes and lines as in Fig. 3. (B) Solid angle area (in π sr) enclosing the −3 dB contours as a function of frequency for the left (open) and right (closed) ears with (black symbols) and without (red symbols) the pinnae. Solid lines indicate the predicted areas based on the circular aperture model (see text) with the diameter given at the top of each respective line. A diameter of 24.13 mm fitted the intact data the best. (C) The elevation (top) and azimuth (bottom) of the acoustic axis for the right ear with (black symbols) and without (red symbols) the pinnae. The values of the axis without pinnae have been shifted by −3° to prevent overlap. (D) Maximum gain as a function of frequency with (black) and without (red) the pinnae. The gain due to the pinnae (gray line) was computed from the difference in the gains with and without the pinnae. The two different lines for each condition in (D) correspond to the left and right ears.

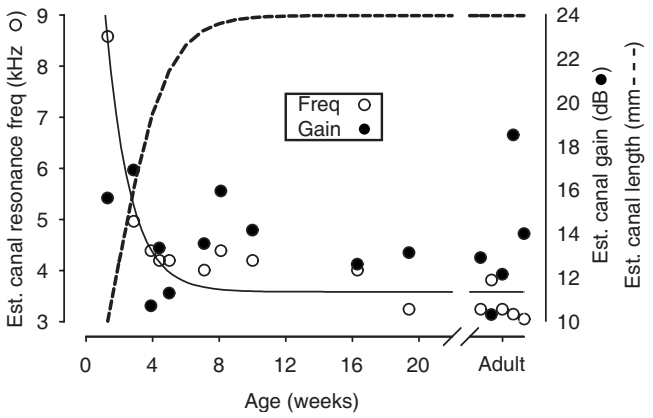


FIG. 8. Estimated development of the resonance frequency, acoustic gain, and the length of the ear canal and concha. The gain (filled) and resonance frequency (open) of the ear canal/concha (labeled simply "canal" in the figure) were estimated from the nondirectional common components of the DTFs (see Sec. II) in 16 animals (5/16 were adults). Solid line shows the best-fitting three-parameter exponential decay function relating the canal/concha resonance frequency to age in weeks. Dashed line indicates the canal/concha length (in millimeters) estimated from the fitted resonance frequency function and assuming that the canal/concha can be modeled as a simple cylinder of a given length and closed at one end (see text).

data. However, a major source of nondirectional gain is the resonance of the ear canal and concha (Rosowski *et al.*, 1988). The resonance of the canal and concha is a large source of the overall acoustical gain of the auditory periphery. During development the dimensions of the ear canal and concha (length and diameter) increase, which would be expected to lower the frequency of the resonance. Unfortunately, we did not measure ear canal or concha dimensions. However, the nondirectional gain and resonance frequency of the canal/concha can be estimated from the common components of the HRTFs [see Sec. II and Koka *et al.* (2008) and Rosowski *et al.* (1988)]. The common component had a gain of ~0 dB as frequency approached 0 Hz, but exhibited a prominent peak in the gain at midfrequencies consistent with expected resonance frequency of the canal and concha. The peak gains and frequencies were remarkably consistent in the left and right ears of each animal, so the results here represent the average gain and corresponding frequency across the left and right ears. The gain and frequency associated with this peak were computed and are plotted in Fig. 8 as a function of age for 16 animals (5/16 were adults). In the youngest animal (1.3 weeks) the estimated resonance frequency was 8.58 kHz producing a gain of 15.6 dB. With development, the resonance frequency decreased rapidly averaging 3.3 ± 0.3 kHz in five adults. To quantify the development of the canal resonance frequency an exponential decay function was fitted to the data. The fitted function $y = 3.58 + 12.04e^{-0.67x}$ accurately described the data ($R^2 = 0.91$, $P < 0.0001$, and $n = 16$). There was no systematic change in the gain with development, averaging 13.6 ± 3 dB in the five adults and 13.6 ± 2 dB in the 11 developing animals. The linear regression relating canal gain to age was not significant ($R = 0.08$, $P = 0.77$). In terms of the resonance frequency, the ear canal of the cat appears to be adultlike in its acoustical properties by 14 weeks when the fitted function asymptotes fully.

IV. DISCUSSION

Cats have been a common model system for general studies of the development of auditory system function including anatomy, physiology, and behavior. Although there has been one brief report of the development of the acoustics of the head and external ear in cats, that study was limited to measurements of the ILD cue to sound location (Moore and Irvine, 1979) in animals older than 5 weeks. To fill this void, we measured here the DTFs in cats from the onset of hearing (~1.5 weeks) from which the monaural acoustical spectral transformations could be computed. Our data reveal that there is considerable development of both the dimensions of the head and pinnae and the resultant acoustical transformations before 5 weeks in the cat and that the monaural acoustical cues are adult by ~16 weeks.

A. Development of the linear dimensions of the head and pinnae

We studied the development of head and pinnae dimensions in 16 animals beginning at 1.3 weeks after birth. Growth curves were fitted to the head and pinnae measurements to provide a quantitative measure of the general growth of these structures in a population of cats. The head and the pinnae of the cats increased in size substantially during development from birth to adult. Based on the growth curves, head diameter increased by a factor of 2.17 from 28.6 mm at birth to 62 mm in adults and reached 90% of adult value by 23 weeks. These compare favorably with values of 29 and 63.02 mm measured in 35 newborn (Latimer, 1931) and 54 adult female cats (Latimer, 1936), respectively. Head diameter in adult males averages 5.5% larger (Latimer, 1936). Interocular (or interpupillary) distance increased by a factor of 2.53 from 13.3 to 33.6 mm, reaching 90% of adult value by 21 weeks. While our measured growth rate for interocular distance was comparable to that reported by Timney (1988) our asymptotic value was ~15% less. The growth of the pinnae dimensions was much more rapid. Inside pinnae height 1-3 increased by a factor of 4.3 from 9.8 to 42.1 mm, reaching 90% at 17.7 weeks. Outside pinnae height 1-2 increased by a factor of 3.31 from 14.6 to 48.4 mm, reaching 90% at 16.4 weeks. And pinnae width *CD* increased by a factor of 4.23 from 5.9 to 25.0 mm, reaching 90% at 13.5 weeks. Bodyweight (not shown) increased from 0.2 ± 0.06 kg at ~1.5 weeks and asymptotes at 2.6 ± 1.1 kg by ~18 weeks. Newborn weights in kittens average 0.15 ± 0.03 kg ($n=35$, Latimer, 1967).

As expected, weight was positively and significantly correlated with all head and pinnae measurements during development, but some correlations remained even in adults (see Table I). The significant correlation between pinnae dimension and weight in adults found here provides a potential explanation for the observation by Xu and Middlebrooks (2000) that optimal frequency scaling factors for DTFs between different cats were significantly correlated with their weight.

Because the behavioral onset of hearing in cats is ~1.5 weeks (Ehret and Romand, 1981; Villablanca and Olmstead, 1979), the acoustical consequences of the increasing size of the head and pinnae are functionally relevant only for ~1.5 weeks and older. Beginning at 1.5 weeks instead of birth the average head diameter increases by a factor of 1.9 and pinnae dimensions *CD*, 1-3 and 1-2 increased by factors of 2.5, 2.7, and 2.3, respectively. Thus, from the onset of hearing in the cat the dimensions of the head and pinnae change substantially.

The rate of growth of the pinnae (given by parameter *b* in the equations in Fig. 1) was 30%–70% greater than that of the head. As such, the pinnae reach adult dimensions much sooner than the head diameter. The major dimension of the pinnae, the outside height 1-2, reached 90% of adult size by 16 weeks while head diameter took 23 weeks. One implication of the rapid development of the pinnae is that the acoustic transformations that are heavily determined by the pinnae, such as the spectral notches (Fig. 2), the acoustic gain (Figs. 3, 5, 7, and 8), and the acoustic axis (Figs. 3, 6, and 7), would be expected to become adultlike before the acoustical transformations that depend on head diameter, such as the binaural cues to sound location, ITDs, and ILDs. (The development of the binaural cues to location will be detailed in a separate paper.) The empirical acoustical measurements discussed below support this hypothesis.

B. Development of the broadband spectral notch cues to location

The spectral notch cues were found to be present in the cat primarily for source locations in the frontal hemisphere. First notch frequencies (e.g., Rice *et al.*, 1992) in animals of all ages were found to increase systematically from lower to higher frequencies as source elevation increased and as source azimuth moved toward the ipsilateral ear [Figs. 2(A)–2(C)]. However, the frequency ranges of FNFs in the frontal hemisphere were highly dependent on age, with high FNFs ranging from 18 to 25 kHz in the youngest animals to lower FNFs of ~8–16 kHz in adults. In other words, for a given source location, the FNF associated with that position systematically decreased with age. Removal of the pinnae in one animal eliminated the spectral notch cues, supporting the hypothesis that even in developing animals with smaller pinnae these cues are created exclusively by the pinnae, as has been demonstrated in other species (rat: Koka *et al.*, 2008; bat: Wotton *et al.*, 1995 and Aytekin *et al.*, 2004; cat: Musicant *et al.*, 1990; and ferret: Parsons *et al.*, 1999). Our data on the spatial dependence of FNFs in adult cats are comparable to that reported in prior studies on cats (Musicant *et al.*, 1990; Rice *et al.*, 1992; Young *et al.*, 1996; Xu and Middlebrooks, 2000).

The developmental changes in FNFs and their ranges were determined by the linear dimensions of the pinnae [Fig. 2(D)]. As the outside height 1-2 of the pinnae increased, the FNF range systematically decreased. Although the ranges of FNFs began to overlap with adult range by 3.9–5 weeks, it was not until >10 weeks that the range fully encompassed the adult range. Outside pinnae height 1-2 was essentially adult by 16 weeks with a value of 44 mm. FNF ranges were fully adult for pinnae heights >44 mm but were not adult

for values less than that. The pinnae dimensions determine directly the functional range of FNFs during development and their final range as adults.

C. Development of the spatial distribution of DTF amplitude gain

1. Development of acoustical gain and directionality

The maximum acoustical gain produced by the head and pinnae increased with frequency and age. Regardless of age, a gain of ~3–5 dB was observed at low frequencies <~3 kHz and far lateral angles [e.g., Figs. 4(A) and 6] and was likely due to the so-called obstacle effect of the head (Kinsler *et al.*, 1982). However, at higher frequencies the maximum gains increased considerably, reaching values of 10–15 dB in the youngest animals increasing to 20–25 dB in adults [Fig. 4(A)], at least over the range of frequencies examined here. The adult values of gain are comparable with the ~25 dB maximum gain at higher frequencies reported previously in cat (Musicant *et al.*, 1990). The increase in gain with age is likely due to the developing amplification capabilities of the pinnae as it increases in size (Coles and Guppy, 1986). Already by 2.9 weeks, the pinnae, by themselves, in the cat can produce nearly 10 dB of gain [Fig. 7(D)]. This is consistent with prior reports of frequency-dependent pinna-only gains of 3–18 dB, with the higher gains occurring at higher frequencies in cat (Phillips *et al.*, 1982), ferret (Carlile and King, 1994 and Schnupp *et al.*, 1998), wallaby (Coles and Guppy, 1986), guinea pig (Palmer and King, 1985), bat (Jen and Chen, 1988 and Obrist and Wenstrup, 1998), and rat (Koka *et al.*, 2008). In a behavioral study by Flynn and Elliot (1965) thresholds for detection in cats were increased when the pinnae were removed by 11 dB at 4 kHz increasing to 17 dB at 16 kHz.

DTF gain directionality and its development were quantified by computing the area of the −3 dB (with respect to maximum gain) contour. In animals of all ages, with increasing frequency the DTF amplitude gain became more directional in that the −3 dB area systematically decreased [Figs. 3 and 4(B)]. The hypothesis that the developmental increase in acoustic directionality was determined by the increasing dimensions of the pinnae was tested here by modeling the pinnae as a circular aperture (see Calford and Pettigrew, 1984 and Coles and Guppy, 1986) and comparing the outputs of the model to our empirical measurements. The frequency dependence of the diffraction of sound into a circular aperture has been shown to account qualitatively for the area of acoustic directionality in a variety of species (Calford and Pettigrew, 1984; Coles and Guppy, 1986; Carlile and Pettigrew, 1987; Musicant *et al.*, 1990; Carlile, 1990; Obrist and Wenstrup, 1998). Here we fitted the circular aperture model to the −3 dB area data for each animal, where the diameter of the aperture was the only free parameter. While the pinnae of the cat are certainly not circular, this simple model did provide an adequate description of the directionality, accounting on average for 50% of the variance in the data. Fits were very poor at low frequencies [Figs. 4(B) and 7(B)] and at a few discrete frequencies near 14 and 28 kHz [see Musicant *et al.* (1990) and Carlile (1990) for a discussion of this phenomenon].

A plot of the best-fitting circular aperture diameter as a function of the three empirically measured pinnae dimensions (Fig. 1) revealed significant linear relationships. However, only the outside pinnae height 1-2 accounted for the most variance (92%) and yielded a regression slope nearest 1.0 (0.89). That the empirical acoustical data can be accounted for by the simple aperture model and that the best-fitting aperture diameter is significantly correlated with the empirical pinnae dimensions (with a slope near 1.0) support the hypothesis that the increasing pinnae dimensions during development are responsible for the increasing acoustical directionality. The pinnae themselves were a critical determinant to directionality as demonstrated in the experiment where both pinnae were removed [Figs. 7(A)–7(D)], supporting an earlier finding by Phillips *et al.* (1982) in cat where pinnae removal abolished the pinna directionality at high frequencies.

2. Development of the acoustic axis

The acoustic axis is the spatial direction for each frequency that produces the largest gain relative to all other directions (Middlebrooks and Pettigrew, 1981; Phillips *et al.*, 1982). In animals of all ages, the acoustic axis in azimuth exhibited a general dependence on frequency where the axis began to migrate from medial to lateral azimuths with increasing frequency (Fig. 6). This trend was interrupted with jumps back toward the midline at particular frequencies followed again by migrations toward lateral azimuths. Sometimes another jump back toward the midline would occur at higher frequencies. Similar trends were observed in elevation where initially the axis migrates from low to high elevations with increasing frequency, transitions back to lower elevations, and then continues to increase in elevation. While the qualitative patterns of acoustic axis movements with frequency were similar in all animals (expect the youngest, 1.3 weeks), the frequencies of the transitional jumps decreased with increasing age. The values of the frequencies of transitions in a given animal were comparable to the ranges of FNFs observed in that animal [Fig. 2(D)]. Given that the FNFs were shown to be dependent on the linear dimensions of the pinnae, which increase with age, we hypothesize that the discrete frequency transitions in the movements of the acoustic axis are determined by the size of the pinnae. While our data are qualitatively consistent with this hypothesis, we did not formally test it. That the pattern of acoustic axis migration with frequency was similar in animals of all ages (and sizes) but that the frequency ranges of these patterns were systematically shifted toward higher frequencies in the younger, smaller animals support our hypothesis. In further support, when the pinnae were removed in one animal, these transitions were substantially disrupted [see also Carlile and Pettigrew (1987)]. These general patterns in the acoustic axis are similar to those reported in other species [cat: Musicant *et al.* (1990), this paper, wallaby: Coles and Guppy (1986), ferret: Carlile (1990), and rat: Koka *et al.* (2008)].

We should note that the acoustic axes in the youngest animal (1.3 weeks, Fig. 6) displayed trends that were different than observed in the older animals, particularly in azimuth. In this animal, the azimuthal axis shifted from lateral to medial azimuths, with discrete jumps back to lateral azimuths, with increasing frequency. This pattern was observed in both ears of this animal and was not observed in any of the other animals (or the other species listed above). Rather, these data were similar to that seen in the guinea pig (Carlile and Pettigrew, 1987). We do not know the exact explanation for this, but the morphology of the pinnae in the youngest cats (at least <2.9 weeks) was rounded, stumpy, and protruded directly away from the meatus and was quite different than that in juveniles or adults where the pinnae take a more triangular and upright shape. The infant cat pinnae that produce the anomalous acoustic axes may be more similar to that in guinea pig.

D. Development of the ear canal and concha resonance frequency and gain

A major factor in the acoustical transformation of sound by the external ear is the resonance created by the ear canal and concha. In our measurements we could not separate the different effects of canal and concha (see Rosowski *et al.*, 1988). For each animal we estimated the canal/concha resonance frequency and gain from the common components of the DTFs. Figure 8 shows that the resonance frequency decreased from 8.58 kHz at 1.3 weeks to an average of 3.3 kHz in adult cats; the asymptote of the fitted curve in Fig. 8 was 3.6 kHz. The resonance frequency was adult by ~ 14 weeks. The acoustical gain associated with this frequency was essentially constant over development, averaging 13.6 dB. Our estimate of resonance frequency and gain of 3.6 kHz and 13.6 dB is comparable to other measurements of canal/concha resonance in adult cats, which occur between 3 and 4 kHz with gains of ~ 15 dB (Wiener *et al.*, 1966; Phillips *et al.*, 1982; Rosowski *et al.*, 1988; Musicant *et al.*, 1990; Rice *et al.*, 1992). The acoustic effects of the ear canal are often modeled as a cylinder closed at one end by the tympanic membrane (Shaw, 1974; Rosowski *et al.*, 1988). In this model the wavelength of the resonance is equal to $\frac{1}{4}$ of the length of the canal. Using this equation, we estimate that the length of the ear canal and concha increased from ~ 10.2 mm at 1.3 weeks and asymptotes to a value of 23.6 mm in adults. Unfortunately we did not measure canal or concha lengths. The latter value of 23.6 mm compares favorably with the concha length of 25.4 mm measured in cats by Rosowski *et al.* (1988). In their study of the development of auditory capabilities in kittens, Olmstead and Villablanca (1980) measured canal and concha length from birth through ~ 2 weeks. At 1 week the length was 10 mm increasing to 15 mm by 2 weeks. The 10 mm empirical measurement is comparable to our 10.2 mm estimate based purely on acoustical measurements. From the fitted function to our data, at 2 weeks the length is estimated to be 13.3 mm, comparable to the 15 mm measured by Olmstead and Villablanca (1980). Given the correspondence between our estimated canal/concha lengths in these latter young ages and in adults, we believe that the length of the developing ear canal and concha in cats can be accurately described using the fitted function describing the resonance frequency and the model of the canal [e.g., $\text{length}=(343\ \text{m/s})/(4\times(3.58+12.04e^{-0.67x}))$] which is plotted also in Fig. 8.

E. Implications for physiological and behavioral development

There have been many anatomical, physiological, and behavioral studies in the cat for which our data may be relevant [see reviews by Kitzes (1990) and Walsh and McGee (1986)]. Because the acoustical properties of the outer and middle ears are major determinants in establishing the frequency range of hearing (Ruggero and Temchin, 2002), we suggest that the general development of the physiology of the ascending auditory pathway and of behavior in general will also be influenced to a large degree by the development of the outer and middle ears. For example, the development of behavioral (Ehret and Romand, 1981) and physiological (Litovsky, 1998) absolute auditory thresholds measured in the free field requires knowledge of the spatial- and frequency-dependent peripheral acoustical transformations reported here, such as the gain and resonance frequency of the ear canal (Fig. 8) and the frequency dependence of maximum acoustical gain of the head and pinnae [Fig. 4(A)]. These acoustical data along with the development of the middle ear transfer function will ultimately determine the effective input to the cochlea [see review by Ehret (1990)].

These data also have implications for the concomitant development of the acoustical cues to sound source location, the neural encoding of these cues, and their ultimate use by the animal for the perception of source location. We shall discuss at length the development of the primary binaural cues to location, ITDs, and ILDs, and how they relate to the growing dimensions of the head and pinnae in a separate paper. Behaviorally, adult cats localize sounds quickly and accurately with performance nearing that of humans (Moore *et al.*, 2008; Tollin *et al.*, 2005; Huang and May, 1996; Populin and Yin, 1998). And even kittens can approach sounds by around 24 days of age, although with much less precision (Clements and Kelly, 1978; Olmstead and Villablanca, 1980; Villablanca and Olmstead, 1979; Norton, 1974). The ability of kittens to make overt orienting responses to sounds suggests that the basic organization of the binaural system may be established early in development. But physiological (Pujol and Hilding, 1973) and simple behavioral (Ehret and Romand, 1981) responses to sound are seen much earlier, a few days after birth. The apparent delay in directional responding might be related to a slower rate of development of the binaural hearing mechanism, the specific cues for location, or simply motor control.

Another contributing factor for poor localization in infant cats may be that the acoustical cues to location are not yet mature. And this would be reflected in the neural coding and subsequent perceptual interpretation of the cues. There is only one study known to the authors that potentially addresses this issue. Wallace and Stein (1997) reported that the size of the auditory spatial receptive fields measured for

broadband stimuli in the superior colliculus in the developing cat decreased in size substantially, by more than a factor of 5 beginning from 2 weeks of age, with adult sizes being reached by ~15–16 weeks [see Fig. 4A in Wallace and Stein (1997)]. The rate of development of the physiological spatial receptive fields they reported was virtually identical to the rate of development of the broadband −3 dB acoustical areas we measured here in Fig. 5. We compared our broadband acoustical data because many superior colliculus neurons in cat exhibit fairly broad selectivity for sound frequency (Hirsch *et al.*, 1985). We fitted a three-parameter exponential function to the receptive field size data of Wallace and Stein (1997) which yielded $y=1.26+28.3e^{-0.78x}$ ($R^2=0.95$, $P<0.0001$). The 95% confidence intervals for the three fitted parameters encompassed all three of the parameters of the function that was fitted to our acoustical data in Fig. 5(B) ($y=1.04+24.3e^{-0.74x}$). In other words, the two functions were not significantly different ($P<0.05$). That the physiological spatial receptive field areas and the −3 dB acoustical areas measured here developed at the same rate strongly suggest that it was the development of the acoustics of the cat head and pinnae that determined the spatial receptive field sizes observed by Wallace and Stein (1997) and not necessarily a development of the physiological receptive field properties of the neurons themselves. Thus, a parsimonious explanation for the development of the spatial receptive fields observed by Wallace and Stein (1997) is that the neural receptive fields in the developing cat may be already adultlike in their physiological properties and that the changing spatial receptive field size with age was a simple consequence of the development of the acoustical properties of the head and external ears. A similar finding was shown directly in the auditory cortex of ferret (Mrsic-Flogel *et al.*, 2003).

We demonstrated here that one monaural cue to source location, the spectral notch (Fig. 2), develops considerably from the onset of hearing through ~16 weeks. These spectral notch cues have been shown to be used by cats for the localization of sounds varying primarily in elevation (Huang and May, 1996; Tollin and Yin, 2003). The range of FNFs decreases by 1–1.5 octaves over ~16 weeks raising interesting, but as yet unanswered, questions as to how cats compensate for these changes. Is there plasticity in the central auditory system during this period that effectively recalibrates mapping of the spectral notch cues (the FNFs) and source location [see Moore and King (2004)]? Or is the internal mapping in the developing animals fixed genetically at or near the onset of hearing with an adultlike mapping?

Studies in the barn owl have revealed a sensitive period early in development where normal acoustical input to the two ears, and thus normal cues to source location, must be present for normal sound localization behavior to develop (Knudsen *et al.*, 1984a, 1984b). The duration of this sensitive period was shown to be correlated with the time course over which the head and facial ruff (like pinnae) dimensions reach maturity, ~8 weeks (Knudsen *et al.*, 1984a). These studies revealed that owls reared with altered acoustical cues (e.g., ear plug) prior to 8 weeks were able to adapt and regain normal sound localization abilities despite the altered cues; however, when the cues were altered in owls after 8 weeks no adaptation was found. Thus, for a period of ~8 weeks, the internal mapping of the ensemble of acoustical cues to location and spatial location itself remains plastic. To the extent to which a similar sensitive period for the development of sound localization in cats exists, our present data detail the developmental constraints on when the peripheral acoustical transformations reach maturity. Here, the monaural spectral transformations are mature by 16 weeks, in line with the development of the linear dimensions of the head and pinnae. We hypothesize that a critical or sensitive period for the consolidation of sound localization in the cat for the monaural cues to location will occur within 16 weeks. Because the head dimensions and the associated binaural cues to location do not reach maturity until ~23 weeks, the sensitive period may be somewhat longer.

ACKNOWLEDGMENTS

We thank Heath Jones and Jennifer Thornton for comments on the manuscript and Janet Ruhland and Mike Wells for assistance in some of the experiments. This work was supported by National Institutes of Deafness and Other Communicative Disorders Grant No. DC-006865 to D.J.T.

Aytekin, M., Grassi, E., Sahota, M., and Moss, C. F. (**2004**). "The bat head-related transfer function reveals binaural cues for sound localization in azimuth and elevation," J. Acoust. Soc. Am. **116**, 3594–3605.
Calford, M. B., and Pettigrew, J. D. (**1984**). "Frequency dependence of directional amplification at the cat's pinna," Hear. Res. **14**, 13–19.
Carlile, S. (**1990**). "The auditory periphery of the ferret. I: Directional response properties and the pattern of interaural level differences," J. Acoust. Soc. Am. **88**, 2180–2195.
Carlile, S., and King, A. J. (**1994**). "Monaural and binaural spectrum level cues in the ferret: Acoustics and the neural representation of auditory space," J. Neurophysiol. **71**, 785–801.
Carlile, S., and Pettigrew, A. G. (**1987**). "Directional properties of the auditory periphery in the guinea pig," Hear. Res. **31**, 111–122.
Clements, M., and Kelly, J. B. (**1978**). "Directional responses by kittens to an auditory stimulus," Dev. Psychobiol. **11**, 505–511.
Coles, R. B., and Guppy, A. (**1986**). "Biophysical aspects of directional hearing in the tammar wallaby, *Macropus eugenii*," J. Exp. Biol. **121**, 371–394.
Ehret, G. (**1990**), in *Development of Sensory Systems in Mammals*, edited by J. R. Coleman (Wiley, New York,), pp. 289–315.
Ehret, G., and Romand, R. (**1981**). "Postnatal development of absolute auditory thresholds in kittens," J. Comp. Physiol. Psychol. **95**, 304–311.
Flynn, W. E., and Elliot, D. N. (**1965**). "Role of the pinna in hearing," J. Acoust. Soc. Am. **38**, 104–105.
Gilbert, S. G. (**1981**), *Pictorial Anatomy of the Cat* (University of Washington Press, Seattle, WA).
Hirsch, J. A., Chan, J. C., and Yin, T. C. T. (**1985**). "Responses of neurons in the cat's superior colliculus to acoustic stimuli. I. Monaural and binaural response properties," J. Neurophysiol. **53**, 726–745.
Huang, A. Y., and May, B. J. (**1996**). "Spectral cues for sound localization in cats: Effects of frequency domain on minimum audible angles in the median and horizontal planes," J. Acoust. Soc. Am. **100**, 2341–2348.
Irvine, D. R. F. (**1986**), *The Auditory Brainstem* (Springer-Verlag, Berlin).
Irvine, D. R. F. (**1987**). "Interaural intensity differences in the cat: Changes in sound pressure level at the two ears associated with azimuthal displacements in the frontal plane," Hear. Res. **26**, 267–286.
Jen, P. H., and Chen, D. M. (**1988**). "Directionality of sound pressure transformation at the pinna of echolocating bats," Hear. Res. **34**, 101–117.
Kinsler, L. E., Frey, A. R., Coppens, A. B., and Sanders, J. V. (**1982**), *Fundamentals of Acoustics* (Wiley, New, York).
Kitzes, L. M. (**1990**), in *Development of Sensory Systems in Mammals*, edited by J. R. Coleman (Wiley, New York), pp. 249–288.
Knudsen, E. I., Esterly, S. D., and Knudsen, P. F. (**1984a**). "Monaural occlusion alters sound localization during a sensitive period in the barn owl," J. Neurosci. **4**, 1001–1011.

Knudsen, E. I., Knudsen, P. F., and Esterly, S. D. (**1984b**). "A critical period for the recovery of sound localization accuracy following monaural occlusion in the barn owl," J. Neurosci. **4**, 1012–1020.
Koka, K., Read, H. L., and Tollin, D. J. (**2008**). "The acoustical cues to sound location in the rat: Measurements of directional transfer functions," J. Acoust. Soc. Am. **123**, 4297–4309.
Kuhn, G. F. (**1987**), in *Directional Hearing*, edited by W. A. Yost and G. Gourevitch (Springer-Verlag, New York), pp. 3–25.
Latimer, H. B. (**1931**). "The prenatal growth of the cat. II. The growth of the dimensions of the head and trunk," Anat. Rec. **50**, 311–332.
Latimer, H. B. (**1936**). "Weights and linear measurements of the adult cat," Am. J. Anat. **58**, 329–347.
Latimer, H. B. (**1967**). "Variability in body and organ weights in the newborn dog and cat compared with that in the adult," Anat. Rec. **157**, 449–456.
Leong, P., and Carlile, S. (**1998**). "Methods for spherical data analysis and visualization," J. Neurosci. Methods **80**, 191–200.
Litovsky, R. Y. (**1998**). "Physiological studies of the precedence effect in the inferior colliculus of the kitten," J. Acoust. Soc. Am. **103**, 3139–3152.
Maki, K., and Furukawa, S. (**2005**). "Reducing individual differences in the external-ear transfer functions of the Mongolian gerbil," J. Acoust. Soc. Am. **118**, 2392–2404.
Martin, R. L., and Webster, W. R. (**1989**). "Interaural sound pressure level differences associated with sound-source locations in the frontal hemisphere of the domestic cat," Hear. Res. **38**, 289–302.
Middlebrooks, J. C., and Green, D. M. (**1990**). "Directional dependence of interaural envelope delays," J. Acoust. Soc. Am. **87**, 2149–2162.
Middlebrooks, J. C. (**1999**). "Individual differences in external-ear transfer functions reduced by scaling in frequency," J. Acoust. Soc. Am. **106**, 1480–1492.
Middlebrooks, J. C., Makous, J. C., and Green, D. M. (**1989**). "Directional sensitivity of sound-pressure levels in the human ear canal," J. Acoust. Soc. Am. **86**, 89–108.
Middlebrooks, J. C., and Pettigrew, J. D. (**1981**). "Functional classes of neurons in primary auditory cortex of the cat distinguished by sensitivity to sound location," J. Neurosci. **1**, 107–120.
Moore, D. R., and Irvine, D. R. F. (**1979**). "A developmental study of the sound pressure transformation by the head of the cat," Acta Oto-Laryngol. **87**, 434–440.
Moore, D. R., and King, A. J. (**2004**), in "Development of the auditory system," *Springer Handbook of Auditory Research*, edited by T. N. Parks, E. W. Rubel, R. R. Fay, and A. N. Popper (Springer-Verlag, New York), pp. 96–172.
Moore, J. M., Tollin, D. J., and Yin, T. C. T. (**2008**). "Can measures of sound localization acuity be related to the precision of absolute localization estimates?" Hear. Res. **238**, 94–109.
Mrsic-Flogel, T. D., Schnupp, J. W. H., and King, A. J. (**2003**). "Acoustic factors govern developmental sharpening of spatial tuning in the auditory cortex," Nat. Neurosci. **6**, 981–988.
Musicant, A. D., Chan, J. C., and Hind, J. E. (**1990**). "Direction-dependent spectral properties of cat external ear: New data and cross-species comparisons," J. Acoust. Soc. Am. **87**, 757–781.
Norton, T. T. (**1974**). "Receptive-field properties of superior colliculus cells and development of visual behaviour in kittens," J. Neurophysiol. **37**, 674–690
Obrist, M. K., and Wenstrup, J. J. (**1998**). "Hearing and hunting in red bats (*Lasiurus Borealis, Vespertilionidae*): Audiogram and ear properties," J. Exp. Biol. **201**, 143–154.
Olmstead, C. E., and Villablanca, J. R. (**1980**). "Development of behavioral audition in the kitten," Physiol. Behav. **24**, 705–712.
Palmer, A. R., and King, A. J. (**1985**). "A monaural space map in the guinea-pig superior colliculus," Hear. Res. **17**, 267–280.
Parsons, C. H., Lanyon, R. G., Schnupp, J. W. H., and King, A. J. (**1999**). "Effects of altering spectral cues in infancy on horizontal and vertical sound localization by adult ferrets," J. Neurophysiol. **82**, 2294–2309.
Phillips, D. P., Calford, M. B., Pettigrew, J. D., Aitkin, L. M., and Semple, M. N. (**1982**). "Directionality of sound pressure transformation at the cat's pinna," Hear. Res. **8**, 13–28.
Populin, L. C., and Yin, T. C. T. (**1998**). "Behavioral studies of sound localization in the cat," J. Neurosci. **18**, 2147–2160.
Pujol, R., and Hilding, D. (**1973**). "Anatomy and physiology of the onset of auditory function," Acta Oto-Laryngol. **76**, 1–10.
Rice, J. J., May, B. J., Spirou, G. A., and Young, E. D. (**1992**). "Pinna-based spectral cues for sound localization in cat," Hear. Res. **58**, 132–152.
Rife, D. D., and Vanderkooy, J. (**1989**). "Transfer-function measurement with maximum-length sequences," J. Audio Eng. Soc. **37**, 419–444.
Rosowski, J. J., Carney, L. H., and Peake, W. T. (**1988**). "The radiation impedance of the external ear of the cat: Measurements and applications," J. Acoust. Soc. Am. **84**, 1695–1708.
Ruggero, M. A., and Temchin, A. N. (**2002**). "The roles of the external, middle, and inner ears in determining the bandwidth of hearing," Proc. Natl. Acad. Sci. U.S.A. **99**, 13206–13210.
Schnupp, J. W. H., Booth, J., and King, A. J. (**2003**). "Modeling individual differences in ferret external ear transfer functions," J. Acoust. Soc. Am. **113**, 2021–2030.
Schnupp, J. W. H., King, A. J., and Carlile, S. (**1998**). "Altered spectral localization cues disrupt the development of the auditory space map in the superior colliculus of the ferret," J. Neurophysiol. **79**, 1053–1069.
Shaw, E. A. G. (**1974**). "The external ear," in *Handbook of Sensory Physiology: Vol. V/1: Auditory System*, edited by W. D. Keidel and W. D. Neff (Springer, New York), pp. 455–490.
Timney, B. (**1988**), in *Advances in Neural and Behavioral Development*, edited by P. Shinkman (Ablex, Norwood, NJ), Vol. **3**, pp. 153–207.
Tollin, D. J., Populin, L. C., Moore, J. M., Ruhland, J. L., and Yin, T. C. T. (**2005**). "Sound-localizaiton performance in the cat: The effect of restraining the head," J. Neurophysiol. **93**, 1223–1234.
Tollin, D. J., and Yin, T. C. T. (**2003**). "Spectral cues explain illusory elevation effects with stereo sounds in cats," J. Neurophysiol. **90**, 525–530.
Villablanca, J. R., and Olmstead, C. E. (**1979**). "Neurological development of kittens," Dev. Psychobiol. **12**, 101–127.
Wallace, M. T., and Stein, B. E. (**1997**). "Development of multisensory neurons and multisensory integration in cat superior colliculus," J. Neurosci. **17**, 2429–2444.
Walsh, E. J., and McGee, J. (**1986**), in *Neurobiology of the Cochlea*, edited by R. A. Altschuler, R. P. Bobbin, and D. W. Hoffman (Raven, New York), pp. 247–269.
Wiener, F. M., Pfeiffer, R. R., and Backus, A. S. N. (**1966**). "On the sound pressure transformation by the head and auditory meatus of the cat," Acta Oto-Laryngol. **61**, 255–269.
Wotton, J. M., Haresign, T., and Simmons, J. A. (**1995**). "Spatially dependent acoustic cues generated by the external ear of the big brown bat, Eptesicus fuscus," J. Acoust. Soc. Am. **98**, 1423–1445.
Xu, L., and Middlebrooks, J. C. (**2000**). "Individual differences in external-ear transfer functions of cats," J. Acoust. Soc. Am. **107**, 1451–1459.
Young, E. D., Rice, J. J., and Tong, S. C. (**1996**). "Effects of pinna position on head-related transfer functions in the cat," J. Acoust. Soc. Am. **99**, 3064–3076.

Detecting incipient inner-ear damage from impulse noise with otoacoustic emissions

Lynne Marshall,[a] Judi A. Lapsley Miller, and Laurie M. Heller[b]
Naval Submarine Medical Research Laboratory, Groton, Connecticut 06349-5900

Keith S. Wolgemuth[c]
Naval Medical Center San Diego, San Diego, California 92134-5000

Linda M. Hughes
Naval Submarine Medical Research Laboratory, Groton, Connecticut 06349-5900

Shelley D. Smith
University of Nebraska Medical Center, Omaha, Nebraska 68198

Richard D. Kopke[d]
DOD Spatial Orientation Center, Naval Medical Center San Diego, San Diego, California 92134-5000

(Received 13 June 2008; revised 20 November 2008; accepted 21 November 2008)

Audiometric thresholds and otoacoustic emissions (OAEs) were measured in 285 U.S. Marine Corps recruits before and three weeks after exposure to impulse-noise sources from weapons' fire and simulated artillery, and in 32 non-noise-exposed controls. At pre-test, audiometric thresholds for all ears were ≤25 dB HL from 0.5 to 3 kHz and ≤30 dB HL at 4 kHz. Ears with low-level or absent OAEs at pre-test were more likely to be classified with significant threshold shifts (STSs) at post-test. A subgroup of 60 noise-exposed volunteers with complete data sets for both ears showed significant decreases in OAE amplitude but no change in audiometric thresholds. STSs and significant emission shifts (SESs) between 2 and 4 kHz in individual ears were identified using criteria based on the standard error of measurement from the control group. There was essentially no association between the occurrence of STS and SES. There were more SESs than STSs, and the group of SES ears had more STS ears than the group of no-SES ears. The increased sensitivity of OAEs in comparison to audiometric thresholds was shown in all analyses, and low-level OAEs indicate an increased risk of future hearing loss by as much as ninefold.
 [DOI: 10.1121/1.3050304]

PACS number(s): 43.64.Jb, 43.64.Wn [BLM] Pages: 995–1013

I. INTRODUCTION

Otoacoustic emissions (OAEs) are more sensitive than pure-tone audiometric thresholds in detecting the early stages of permanent noise-induced inner-ear damage in humans. Typical results for noise-exposed *groups* followed longitudinally show a decrease in OAE amplitudes, but no change in audiometric thresholds (Engdahl *et al.*, 1996; Murray *et al.*, 1998; Murray and LePage, 2002; Konopka *et al.*, 2005; Seixas *et al.*, 2005a, 2005b; Lapsley Miller *et al.*, 2006).[1] Most longitudinal studies do not last long enough to also see hearing loss in the noise-exposed group. A recent finding is that low-level or absent OAEs in noise-exposed *individual* ears may be a risk factor or predictor for hearing loss in their near future for continuous noise overlaid with impact noise (Lapsley Miller *et al.*, 2006). It is of interest to know for both theoretical and clinical reasons whether this finding generalizes to impulse noise.

Impulse noise is a common occupational and recreational hazard (Clark, 1991; Humes *et al.*, 2005). The waveform of the impulse as received at the ear is shaped by the individual pinna, ear canal, and middle ear, which may have an activated middle-ear reflex, sometimes even prior to the noise exposure in the case of an anticipatory reflex (e.g., Marshall *et al.*, 1975). To add to the complexity, higher-level sounds may result in less stapes motion and thus less damage than lower-level sounds (e.g., Price, 2007). In a work setting with considerable impulse-noise exposure (both self-generated and from other sources in the environment), as is the case for some military jobs, the impulse-noise exposure for any one individual can be difficult to quantify (unless one has the luxury of a microphone in the ear canal). We expect more variability in the noise exposure to the cochlea across these individuals than for individuals working in steady-state background noise, such as an engine room, where the noise exposure is more homogenous.

If the noise exposure reaching the cochlea is more variable across individuals and if a single exposure can cause inner-ear damage, we expect that the state of the inner ear

[a]Author to whom correspondence should be addressed. Electronic mail: lynne.marshall@med.navy.mil
[b]Present address: Department of Cognitive and Linguistic Sciences, Brown University, Providence, RI 02912.
[c]Present address: Department of Communicative Disorders, University of Redlands, Redlands, CA 92373.
[d]Present address: Hough Ear Institute, Oklahoma City, OK 73112.

prior to the noise exposure will not be as predictive of incipient risk for this group of people as for a group of people with more uniform noise exposure.

Noise exposure from live-fire training can produce hearing loss very quickly. Permanent threshold shifts (PTSs) have been reported in 10% or more of military personnel during weapons' training, including 300 rounds of M-16 live-fire training,[2] Army special forces undergoing routine weapons' training,[3] and Israeli army recruits firing an average of 420 M-16 rounds (Attias *et al.*, 1994), all in spite of wearing hearing protection. In the current study, Marine recruits undergoing basic training were chosen because some PTS within a short amount of time was expected, and the overall amount of noise exposure for each volunteer would be very similar.

The current study is complementary to the study reported in Lapsley Miller *et al.* (2006). The same experimental protocol was used, with each volunteer receiving both pure-tone audiometry and OAE tests before and after a significant multiday noise exposure in a military operational setting. The primary difference between the two studies is the type of noise exposure—with impulse-noise exposure (from weapons' fire) in the current study, in contrast to continuous noise overlaid with impact noise (from aircraft and machinery noise) in the previous study. The current study is also a subset of a larger interdisciplinary study investigating the auditory and genetic determinants of susceptibility to noise-induced hearing loss (NIHL).

II. METHOD

A. Participants

The study participants were 401 male volunteers who had just begun mandatory basic training as U.S. Marine Corps recruits.[4] No female volunteers were available, as this military installation provided training to male recruits only. Two experimental groups were formed from the volunteers who met the screening criteria and completed the study: the noise-exposed group ($N=285$; age at enrollment: range 17.4–28.1 years, median=19.2 years); and the control group (who were not exposed to noise between pre- and post-tests; $N=32$; age at enrollment: range 18.2–27.1 years, median =20.0 years).

Each group was tested twice with an identical protocol. Test times were determined by the recruit training schedule. The pre-test measurements occurred one to six weeks prior to the noise exposures (all volunteers had been noise-free for at least a day), and the post-test measurements occurred three weeks after the noise exposures. At pre-test, volunteers were screened for clear ear canals, audiometric thresholds of $\leq$25 dB HL from 0.5 to 3 kHz and $\leq$30 dB HL at 4 kHz, and peak immitance within the range of $\pm$50 daPa atmospheric pressure, with grossly normal amplitude, slope, and smoothness of the tympanogram. Volunteers who met these screening criteria proceeded to OAE testing. Volunteers who did not meet the screening criteria did not enroll in the study. At post-test, volunteers were checked for clear ear canals (cerumen was removed if present) and peak immitance within the range of $\pm$50 daPa atmospheric pressure, with grossly normal amplitude, slope, and smoothness of the tympanogram.

Between test sessions, the noise-exposed group received impulse-noise exposures (M-16 rifle, M-60 machine gun, and C-4 explosives) as prescribed by standard operating procedures.[5] A 5%–10% incidence of permanent hearing loss was expected, despite mandatory hearing protection. The volunteers in the control group were recruits who were in a "medical hold" status for minor nonauditory injuries, and a few U.S. Navy medical personnel. They were tested on two occasions within a 24–48 h period without any weapons' noise exposures before or between tests.

Data collection occurred from January to September 2000.

B. Noise exposures

During basic training, the noise-exposed group underwent approximately three and a half weeks of training that involved weapons' noise exposures at Marine Corps Base, Camp Pendleton, CA. All recruits spent six days on outdoor rifle ranges, where each individual fired 340 rounds with an M-16 rifle (~157 dB pSPL, US Army Center for Health Promotion and Preventive Medicine, 2008). Next, they made three 20–30 min runs through a combat obstacle course where they were exposed to M-60 machine-gun fire (~155 dB pSPL, US Army Center for Health Promotion and Preventive Medicine, 2008) and simulated artillery using C-4 explosives. Noise measurements performed with the Quest M-27 noise logging dosimeter revealed both the simulated artillery and C-4 explosions produced levels in excess of 146 dB pSPL (maximum limits of the M-27). Noise sources were located 5–20 ft from volunteers' ears depending on where individuals were located on the course at the time of an impulse-noise presentation. Finally, there were several days of simulated-combat exercises where each recruit fired an additional 50–75 M-16 rounds, as well as exposure to more M-60 machine-gun fire and simulated artillery using C-4 explosives. Because other recruits were simultaneously firing M-16 rounds on the rifle range and during combat exercises, each individual was exposed to more than the 390–415 rounds they fired from their own weapons. The exact number of M-16 rifle-fire exposures for each volunteer was not measured. The recruits fired from three positions: lying down in a prone position, in a sitting position, and standing up. The majority of the rounds fired were in the prone position (80%). The noise exposures were very similar for each volunteer for the rifle range and obstacle course, but there was a lot more variability in exposures through the simulated-combat exercises. The rest of the time, the recruits' activities were severely restricted, and they were not exposed to any significant levels of nonmilitary noise.

The recruits were provided with foam, disposable E·A·R Classic (Aearo Corporation) earplugs each day on the rifle range, obstacle course, and during combat exercises. These earplugs come in one size only and have a noise reduction rating of 29 dB (Berger, 2000), but only if properly and deeply fitted. The drill instructors informed large groups

of recruits how to use the earplugs. Because individual fitting was not done and only one size was provided, the actual field attenuation no doubt was much less than optimal.

C. Audiometric equipment, stimuli, and testing

Audiometric testing was performed using either a Maico MA-1000 PC audiometer or a Grason-Stadler G-117 portable audiometer. Pure-tone stimuli were delivered through Telephonics TDH-49 supra-aural earphones in MX41/AR cushions. Audiometers were calibrated (ANSI, 1996), and daily calibration and listening checks were performed each day of testing (Navy Occupational Health and Safety Program, 1999). Audiometric testing was performed one volunteer at a time in double-walled sound-attenuating chambers (ANSI, 1991). Earphone placement was checked by the examiner. At the pre-test audiogram, the recruits had been without much sleep for one to two nights. This necessitated the examiner being in the same room as the volunteer, testing him similarly to a pediatric patient (e.g., frequent animated verbal interaction and encouragement) to maintain alertness.

Audiometric thresholds were measured in both ears (the left ear was always tested first), using the standard U.S. Navy hearing-conservation program test protocol, which is an ascending, modified Hughson–Westlake procedure, with a 5 dB step size and frequencies tested in the order of 1, 0.5, 1, 2, 3, 4, and 6 kHz (Navy Occupational Health and Safety Program, 1999). All audiograms were collected manually by qualified technicians or audiologists.

Note that we did not use the results of the group hearing testing typically done for marine recruits as they enter basic training, because it was not reliable enough for our purposes (automated audiometry with up to eight recruits at a time in the booth).

D. Tympanometry equipment, stimuli, and testing

Middle-ear pressures were estimated from the peak of an immitance tympanogram with a 226 Hz tone using a Grason-Stadler GSI 33 version 2 analyzer at a sweep speed of 12.5 daPa/s to minimize hysteresis.

E. Otoacoustic emission equipment, stimuli, and testing

Two types of OAEs were measured: transient-evoked otoacoustic emissions (TEOAEs) and distortion-product otoacoustic emissions (DPOAEs). Both OAE types were measured with the ILO292 Echoport system (Otodynamics Ltd., England), using the DPOAE probe. To allow better placement and manipulation in the ear canal, an acoustic-immitance probe tip (which had been enlarged using a grinding tool) was inserted onto the DPOAE probe. The size of the probe tip was matched to the size of the ear canal, and was noted so that the same size could be used for the pre- and post-tests. Individual in-the-ear calibration was used for both TEOAE and DPOAE measurements. OAE testing was performed two volunteers at a time, with two testers also present, in a double-walled sound-attenuating chamber (ANSI, 1991).

An identical test battery was used as for Lapsley Miller *et al.* (2006). Before OAE testing (for both pre- and post-tests), an otoscopic examination was conducted with cerumen removal if necessary, and peak immitance was measured to ensure it was within ±50 daPa atmospheric pressure in both ears.

TEOAEs were evoked with a 74 dB pSPL click, presented in nonlinear mode, where responses to three clicks at one polarity and one click with opposite polarity and 9.5 dB higher were added together to reduce linear artifact from the stimulus (Bray, 1989). At pre-test, every attempt was made to get a flat stimulus spectrum during calibration by manipulating the depth and angle of the probe tip in the ear canal. At post-test, every attempt was made to get the same stimulus pattern during calibration as in the pre-test by referring to a screenshot printed out after the first test. TEOAEs were collected and averaged until 260 low-noise averages were obtained. The results were windowed (2.5 ms onset delay, 20.5 ms duration, with 2.56 ms rise/fall) and filtered (0.683–6.103 kHz bandpass filter), then analyzed into half-octave bands (0.7, 1, 1.4, 2, 2.8, 4, and 5.6 kHz).

In order of presentation, DPOAEs were measured with stimulus levels L_1/L_2=57/45, 59/50, 61/55, and 65/45 dB SPL (abbreviated herein to $DP_{57/45}$, $DP_{59/50}$, $DP_{61/55}$, and $DP_{65/45}$). For all stimulus levels, the f_2/f_1 ratio was 1.22, with f_2=1.8, 2.0, 2.2, 2.5, 2.8, 3.2, 3.6, 4.0, and 4.5 kHz.

F. Data definitions, cleaning, and reduction

The short testing time available for each volunteer meant that it was not always possible to obtain clean data. As in the previous study, OAE data were affected by electrical noise when running on line power (it was not possible to always run with batteries). Data points and/or test conditions contaminated with off-target stimulus levels, poor calibrations, high noise-levels, large differences in noise level between tests, or many unexplained outliers were removed from the data set in an objective fashion, using the same elimination rules across the entire dataset of all volunteers (see footnote 5, Lapsley Miller *et al.*, 2006).

A TEOAE was considered present if its amplitude was greater than the noise floor. A DPOAE was considered present if its amplitude was greater than the noise floor, which was redefined as the average noise floor from the three frequency bins above and below the $2f_1$-f_2 frequency bin plus two standard deviations. For some ears, a pre-test OAE was present, but the post-test OAE was absent. The noise-floor level accompanying an absent post-test OAE was substituted for the absent OAE providing the noise-floor level was lower than the *pre-test* OAE (see Lapsley Miller and Marshall, 2001, pp. 6 and 7; Lapsley Miller *et al.*, 2004, p. 311; and Lapsley Miller *et al.*, 2006, Sec. II E). Thus some OAE changes were potentially underestimated, but this was considered preferable to not using the data at all. The substitution was not done if the post-test noise-floor level was higher than the pre-test OAE, because a high noise-floor level could masquerade as an increase in OAE amplitude. For the susceptibility analyses, it was of interest to know if low or absent OAEs at pre-test increased the chance of STS

TABLE I. The number of ears and the total number of volunteers in each group that contributed to each analysis, listed by the section. The numbers varied at each test frequency, OAE level, and OAE type, because only valid data were used. The exception was for the ANOVAs where volunteers were required to have complete OAE data sets for both ears.

Section and analysis	Group	No. of ears (range)	Total No. of volunteers	Notes
II.A. Volunteers completing study	Noise	570	285	369 noise-exposed volunteers were enrolled, but only 285 completed the study. Not all noise-exposed volunteers were available for post-testing. All control volunteers completed the study.
	Control	64	32	
III.A. ANOVA	Noise	120	60	Ear was a factor in the ANOVAs.
III.B. Forming STS and SES criteria	Control	STS: 64	32	Ears were pooled (Tables II and III).
		SES: 36–54	32	
III.C. Identifying and describing STS and no-STS ears	Noise	STS: 42	36	STSs were detected in 15 left ears only; 15 right ears only; 6 bilateral.
		No-STS: 528	279	
III.D. Identifying and describing SES ears	Noise	SES: 42–49	32–43	The number of ears varied across the left/right ear, frequency and OAE type. There were many ears where SES status could not be determined. Also see Table VI.
		No-SES: 256–398	85–151	
III.E. Comparing STS and SES status	Noise	540	285	See Table V for the SES ears broken down by OAE type and STS status.
III.F. Susceptibility	Noise	STS: 17–21	21	Ear was a factor. The number of ears contributing to the analysis also varied across the OAE type, level, and frequency as all valid data were used.
		No-STS: 217–263	264	
IV.G. Susceptibility for worst ear	Noise	STS: 35–36	36	Ear with the lowest OAE amplitude was used as a predictor.
		No-STS: 234–249	249	
IV.E. Comparison to Lapsley Miller *et al.* (2006)	Noise	STS: 37–42	36	Ears were pooled.
		No-STS: 439–523	279	

at post-test. Absent pre-test OAEs were estimated where possible by substituting the noise floor for the absent OAE, providing the noise floor was sufficiently low, defined here as being in the tenth percentile of OAE amplitude (see Lapsley Miller *et al.*, 2006, Sec. II E).

To reduce the impact of unusable data, subsets of test frequencies and levels were used in the analyses: TEOAEs at 1, 1.4, 2, 2.8, and 4 kHz, or just the frequencies 2, 2.8, and 4 kHz for some analyses; and $DP_{65/45}$ and $DP_{59/50}$ at 2.5, 2.8, 3.2, 3.6, and 4.0 kHz, or just the frequencies 2.8, 3.2, and 4.0 kHz for some analyses. The TEOAE frequency bands at 0.7 and 5.6 kHz were excluded due to low amplitude resulting from the windowing and filtering used to extract the TEOAE. The DPOAE frequencies 1.8, 2.0, 2.2, and 4.5 kHz were excluded (a) due to electrical noise artifacts that (usually) elevated DPOAE amplitudes and/or noise-floor levels at 2.2 and 4.5 kHz, and (b) for noise-floor levels that were on average much higher than those at 2.5, 2.8, 3.2, 3.6, and 4.0 kHz. It was not possible to sensibly average across the remaining DPOAE frequencies or to compute DPOAE growth functions due to unusable data. As such, the DP levels $DP_{57/45}$ and $DP_{61/55}$ were not examined further.

We believe that we were measuring permanent changes in audiometric thresholds and OAEs because the post-tests were performed long enough after the noise exposure (three weeks) that any temporary threshold shifts (TTSs) or temporary emission shifts should have resolved. Nevertheless, because it was not possible to confirm the significant audiometric threshold and OAE shifts in individual ears with a follow-up test at a later time, we are careful here to refer to significant threshold shifts (STSs), rather than PTSs, and likewise for OAEs we refer to significant emission shifts (SESs), rather than permanent emissions shifts.

III. RESULTS

Table I provides an overview of the number of volunteers and ears in each group contributing to each analysis. Depending on the analyses, the noise-exposed group is further split into groups of ears with and without STSs (STS

and no-STS) and/or SESs (SES and no-SES). Volunteers may have had unilateral or bilateral significant shifts.

A. Changes in group OAE and audiometric thresholds after noise exposure

Separate repeated-measures analyses of variance (ANOVAs) were conducted on audiometric threshold, TEOAE, and DPOAE data for the subgroup of 60 volunteers (median age 19 years) from the noise-exposed group with complete data sets. A volunteer had a complete data set if, for both ears and for both pre- and post-tests, there was a set of audiometric thresholds (no missing data for any volunteer) and a set of measurable (or estimated) OAEs (TEOAEs at 1, 1.4 2, 2.8, and 4 kHz; and $DP_{65/45}$ and $DP_{59/50}$ at 2.5, 2.8, 3.2, 3.6, and 4.0 kHz; see Sec. II F). As described earlier, some absent post-test OAEs were estimated using the noise floor. Data from volunteers with incomplete data sets were not used for this analysis. Incomplete data sets were attributable to measurement errors, high noise, and/or absent OAEs at pre-test.[6] By selecting volunteers with complete data sets, a bias may have been introduced, because those volunteers with unusable data may have lower or absent OAEs from noise-induced damage—the lower OAEs being harder to detect from the noise floor. However, by using complete data sets, comparisons across OAE stimulus types, frequencies, and ears could be made more fairly.

A three-way repeated-measures ANOVA was conducted for audiometric thresholds (test: pre and post; ear: left and right; and frequency: 0.5, 1, 2, 3, 4, and 6 kHz). There was no significant change in audiometric thresholds (main effect) between pre- and post-tests ($F_{1,59}=0.03$, ns). There were, however, significant differences between ears ($F_{1,59}=4.4, p<0.05$) and across frequency ($F_{5,295}=14.8, p<0.05$). There was also a two-way interaction for test by frequency ($F_{5,295}=3.2, p<0.05$). Bonferroni *post hoc t*-test comparisons were used to establish whether any same-frequency pairs contributed to this interaction. The familywise significance level was $p<0.05$, so, for six comparisons, $p<0.008$ was used. None were significant.

A three-way repeated-measures ANOVA was conducted for TEOAE amplitude (test: pre and post; ear: left and right; and frequency: 1, 1.4, 2, 2.8, and 4 kHz). All three factors showed significant main effects. Particularly, there was a 0.94 dB decrease in TEOAE amplitude between pre- and post-testing ($F_{1,59}=14.4$, $p<0.05$). Ears also differed ($F_{1,59}=37.38, p<0.05$) as did frequency ($F_{4,236}=14.45, p<0.05$). There was one significant two-way interaction: test by frequency ($F_{4,236}=2.9, p<0.05$). Bonferroni *post hoc t*-test comparisons were used to establish whether any same-frequency pairs contributed to this interaction. The familywise significance level was $p<0.05$, so, for five comparisons, $p<0.01$ was used. The TEOAE amplitudes at the frequencies 1.4, 2, and 2.8 kHz contributed to the interaction with significant decreases between pre- and post-testing of 1.1, 1.3, and 1.0 dB, respectively.

A four-way repeated-measures ANOVA was conducted for DPOAE amplitude (test: pre and post; ear: left and right; level: stimulus levels of 65/45 and 59/50 dB SPL; and frequency: 2.5, 2.8, 3.2, 3.6, and 4.0 kHz). All four factors showed significant main effects. Particularly, there was a 0.84 dB decrement in DPOAE amplitude between pre- and post-testing ($F_{1,59}=8.6, p<0.05$). There were also main effects for ear ($F_{1,59}=4.9, p<0.05$), level ($F_{1,59}=140.3, p<0.05$), and frequency ($F_{4,236}=68.6, p<0.05$). There were three significant two-way interactions: test by level ($F_{1,59}=6.6, p<0.05$), ear by level ($F_{1,59}=4.1, p<0.05$), and level by frequency ($F_{4,236}=10.7, p<0.05$). Bonferroni *post hoc t*-test comparisons were used to establish which levels contributed to the test-by-level, two-way interaction. The familywise significance level was $p<0.05$, so, for two comparisons, $p<0.025$ was used. Neither was significant.

B. Significant threshold shift (STS) and significant emission shift (SES) criteria

Criteria for the detection of STS and SES in individual ears were developed using the same method as in Lapsley Miller *et al.* (2006),[7] which was based on the standard error of measurement (SE_{meas}) derived from the control-group data. Because the control group of 36 volunteers (64 ears) had not been exposed to noise between tests, the SE_{meas} represents the amount of variability attributable to other sources (i.e., fluctuations in the OAE level over time, differences in probe position or movement, etc.). Any OAE or audiometric threshold in a noise-exposed ear that exceeds these SES or STS criteria can be interpreted as being due to noise exposure (although there is always the possibility of a false positive).

Table II shows the STS criteria. STSs detected at 2, 3, or 4 kHz and the averaged shifts at 2 and 3 kHz, 3 and 4 kHz, and 2, 3, and 4 kHz were used to define the group of STS ears for subsequent analyses. Averaged shifts were included as they are commonly used by regulatory agencies to detect and define threshold shifts. As a crosscheck, no STSs were detected in any ear in the control group. Shifts at 0.5 and 1 kHz were not considered because on their own they are not diagnostic of NIHL, and we were mindful that each look increases the false-positive STS rate. We wanted to focus on those frequencies most likely to show NIHL. Shifts at 6 kHz, however, also were not considered because the STS criterion at 6 kHz was deemed less reliable for detecting noise-induced STS, based on the distribution of negative STSs in the noise-exposed group (there were more negative STS cases than positive).

Table III shows the SES criteria. The criteria for the TEOAEs ranged from approximately 4 to 6 dB, and tended to be smaller than for the DPOAE criteria. The criteria for $DP_{59/50}$ ranged from approximately 6 to 10 dB, and for $DP_{65/45}$ ranged from approximately 7 to 8 dB. The criteria tended to be smaller for the $DP_{65/45}$ compared with $DP_{59/50}$.

C. STSs detected in the noise-exposed group

A total of 36 out of 285 volunteers in the noise-exposed group (12.6%) were classified with a STS in at least one ear three weeks after the noise exposure (median age 19.1 years). When considering ears rather than volunteers, 42 out of 570 ears (7.4%) were classified with a STS.[8] There

TABLE II. STS criteria based on the standard error of measurement (SE_{meas}) from the control group (32 volunteers/64 ears) for individual audiometric-threshold frequencies and for averaged frequencies. Shown is the frequency, mean shift between post-testing and pre-testing, SE_{meas}, and the resulting STS criteria (see footnote 7). Note that although the STS criteria were calculated for all frequencies, only frequencies from 2 to 4 kHz and the averaged frequency bands were used to determine the STS status.

Frequency (kHz)	Average shift (dB)	SE_{meas} (dB)	STS (dB)
0.5	−1.4	4.0	20
1	−1.8	3.6	15
2	−2.0	3.4	15
3	−2.3	3.6	15
4	−2.0	3.9	15
6	−1.5	4.5	20
Mean 2 and 3	−2.2	3.1	10
Mean 3 and 4	−2.2	3.1	10
Mean 2, 3, and 4	−2.1	2.7	8.3

were 15 left STS ears, 15 right STS ears, and 6 bilateral STS ears. Table IV summarizes the STS and no-STS ears by left and right ears.

Figure 1 shows the average pre- and post-test audiograms for the STS ears (42 ears) compared with the no-STS ears (528 ears). The STS ears' average thresholds increased while the no-STS ears' average thresholds stayed the same. Although there were the same number of left STS ears and right STS ears, the left ear STSs on average were larger and broader-band than the right ear STSs. The largest average increases in threshold were 13.3 dB at 4 kHz for the right ears and 11.7 dB at 4 kHz for the left ears. The largest individual increases in threshold were 40 dB STSs at 4 kHz in both ears of one volunteer.

D. SESs detected in the noise-exposed group

SES status was determined for each ear of each volunteer in the noise-exposed group, separately for each OAE type. For comparison with the three frequencies used to assess STS, three frequencies/frequency bands were considered for each OAE type within the 2–4 kHz range. For TEOAEs, this was 2, 2.8, and 4 kHz half-octave bands. For DPOAEs this was 2.5, 3.2, and 4.0 kHz.[9] Three types of SES status were defined.

- *No-SES*. No OAE decrements at any of the three frequencies.
- *SES*. At least one significant decrease in OAE amplitude at any of the three frequencies. Other frequencies could have either no shifts or unusable data.
- *Unknown-SES*. At least one frequency band where SES status could not be determined[10] and no SES shifts at the other frequencies. In other words, it was unknown whether the ear should be a no-SES or a SES. No distinction is made here between OAEs below the noise-floor criterion and data loss due to measurement problems.

TABLE III. SES criteria based on the standard error of measurement (SE_{meas}) from the control group (32 volunteers/64 ears). Shown are the OAE type, f_2 frequency for DPOAEs or half-octave frequency band for TEOAEs, the number of ears contributing to the calculation, SE_{meas}, and the resulting SES criteria (see footnote 7). Note that although the SES criteria were calculated for all valid frequencies, only some frequencies were used to determine the SES status (2.5, 3.2, and 4 kHz for DPOAEs, and 2, 2.8, and 4 kHz for TEOAEs).

OAE type	Frequency (kHz)	Ears	SE_{meas} (dB)	SES (dB)
$DP_{59/50}$	2.5	40	2.0	6.0
	2.8	40	2.5	7.6
	3.2	48	3.3	9.9
	3.6	45	2.8	8.5
	4.0	51	2.9	8.7
$DP_{65/45}$	2.5	44	2.5	7.5
	2.8	48	2.5	7.5
	3.2	54	2.6	7.7
	3.6	54	2.3	7.0
	4.0	54	2.2	6.7
TEOAE	1.0	45	2.0	5.8
	1.4	48	2.0	6.0
	2.0	43	2.0	6.1
	2.8	42	1.7	5.2
	4.0	36	1.3	4.0

Summarizing from Table IV, which provides a breakdown of SES status by ear and OAE type, 42 out of 285 noise-exposed group volunteers (14.7%) showed a $DP_{59/50}$ SES, 32 volunteers (11.2%) showed a $DP_{65/45}$ SES, and 43 volunteers (15.1%) showed a TEOAE SES. This included two, ten, and six volunteers with bilateral SES, for each OAE type, respectively. When considering ears rather than volunteers, 44 out of 570 noise-exposed ears (7.7%) showed a $DP_{59/50}$ SES, 42 ears (7.4%) showed a $DP_{65/45}$ SES, and 49 ears (8.6%) showed a TEOAE SES. These percentages are similar to those seen for STS, but are underestimated due to the large number of ears with unknown-SES status. The true SES rate is likely to be much higher.[11]

To estimate the true SES rate (for individual frequencies), the percentages were recalculated (separately for each OAE type) for the group of volunteers where status was known for both ears (see Table IV). For these subgroups, 28 out of 158 volunteers (17.7%) showed a $DP_{59/50}$ SES, 29 out of 180 volunteers (16.1%) showed a $DP_{65/45}$ SES, and 28 out of 113 volunteers (24.8%) showed a TEOAE SES. When considering ears rather than volunteers, 30 out of 316 ears (9.5%) showed a $DP_{59/50}$ SES, 39 out of 360 ears (10.8%) showed a $DP_{65/45}$ SES, and 34 out of 226 ears (15.0%) showed a TEOAE SES. There was a tendency for more left

TABLE IV. Breakdown of the 285 volunteers in the noise-exposed group by STS status and SES status for the left/right ear and the measurement type. The first number is the count, and the number in parentheses is the overall percentage. The Unknown category represents those ears for which a SES determination could not be made, usually due to unusable data. See text for summaries of STS and SES rates for volunteers and ears.

Audiometric thresholds			Right ears		
			STS	No-STS	Unknown
	Left ears	STS	6 (2)	15 (5)	0 (0)
		No-STS	15 (5)	249 (87)	0 (0)
		Unknown	0 (0)	0 (0)	0 (0)
$DP_{59/50}$			Right ears		
			SES	No-SES	Unknown
	Left ears	SES	2 (1)	10 (4)	6 (2)
		No-SES	16 (6)	130 (46)	39 (14)
		Unknown	8 (3)	32 (11)	42 (15)
$DP_{65/45}$			Right ears		
			SES	No-SES	Unknown
	Left ears	SES	10 (4)	5 (2)	1 (0)
		No-SES	14 (5)	151 (53)	41 (14)
		Unknown	2 (1)	36 (13)	25 (9)
TEOAE			Right ears		
			SES	No-SES	Unknown
	Left ears	SES	6 (2)	15 (5)	7 (2)
		No-SES	7 (2)	85 (30)	26 (9)
		Unknown	8 (3)	38 (13)	93 (33)

ears to show DP SESs (~61% left ears and ~39% right ears) and for more right ears to show TEOAE SESs (38% left ears and 62% right ears).

Figures 2–4 show the average pre- and post-test OAE amplitudes for the SES ears compared with the no-SES ears, by left and right ears for each OAE type, without regard to STS status. Error bars are 95% confidence intervals. Note that SES status was determined from only three frequencies for each OAE type. To maximize the amount of data going into each point, there was no requirement for an ear to have

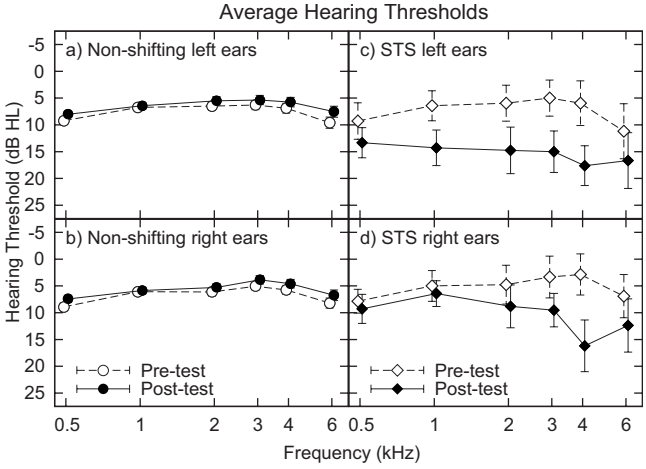


FIG. 1. Average pre-test and post-test audiometric thresholds for the noise-exposed group by STS status. Average pre-test thresholds for the 42 STS ears (21 left and 21 right ears) were essentially the same as for the 528 no-STS ears (286 left ears and 286 right ears). Post-test audiograms show that the average thresholds for the STS ears increased up to 13.3 dB (left ears, 4 kHz), while the no-STS ears stayed essentially the same. The error bars are 95% confidence intervals.

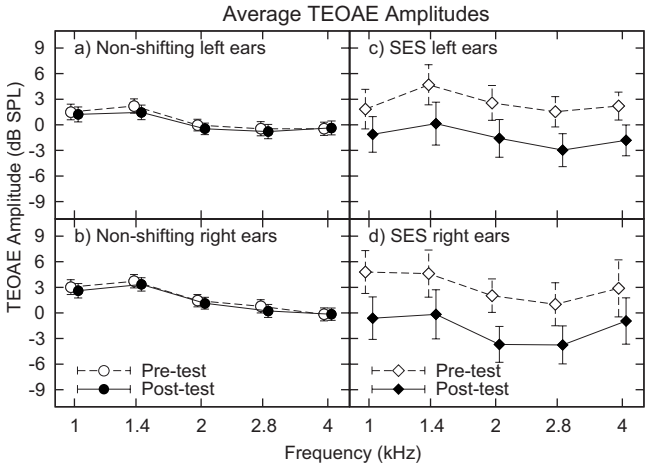


FIG. 2. Average pre-test and post-test TEOAE amplitudes for the noise-exposed group by significant TEOAE shift (SES) status. Average pre-test amplitudes for the SES ears (22–27 left ears and 14–21 right ears) were slightly higher than for the no-SES ears (112–118 left ears and 129–138 right ears). Average post-test TEOAE amplitudes decreased by approximately 4 dB from pre-test for the SES ears, while the post-test average for the no-SES ears stayed essentially the same. The error bars are 95% confidence intervals. The number of ears contributing to the average at each frequency varied because of some unusable data.

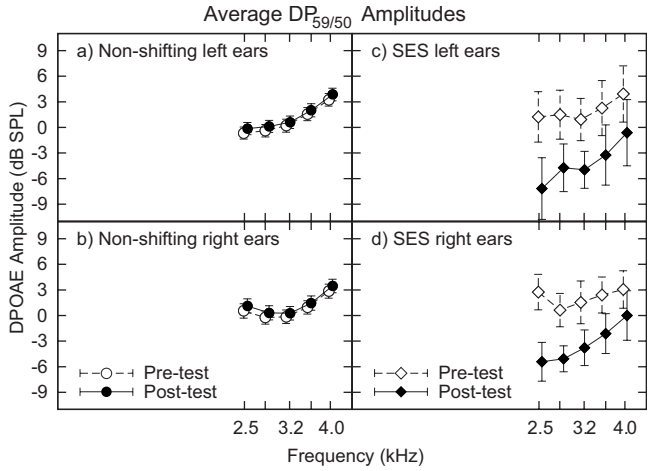


FIG. 3. Average pre-test and post-test $DP_{59/50}$ amplitudes for the noise-exposed group by significant $DP_{59/50}$ shift (SES) status. Average pre-test amplitudes for the SES ears (16–17 left ears and 25–26 right ears) were slightly higher than the average for the no-SES ears (180–185 left ears and 166–172 right ears). Average post-test $DP_{59/50}$ amplitudes decreased by approximately 6 dB from pre-test for the SES ears, while the post-test average for the no-SES ears stayed essentially the same. The error bars are 95% confidence intervals. The number of ears contributing to the average at each frequency varied because of some unusable data.

usable data at all displayed frequencies; however, an ear needed both a valid pre- and post-test measurement at a frequency to be included. There were more unusable data for the SES ears because these ears needed only one SES at one frequency to be considered a SES ear, whereas the no-SES ears were required to have no SESs at all three frequencies.

The SES ears showed on average an ~4 dB broadband TEOAE decrement and ~6 dB DPOAE decrements across 2.5–4 kHz in average OAE amplitude between pre- and post-testing, whereas the no-SES ears showed essentially no change. There was a tendency, especially for TEOAEs, for

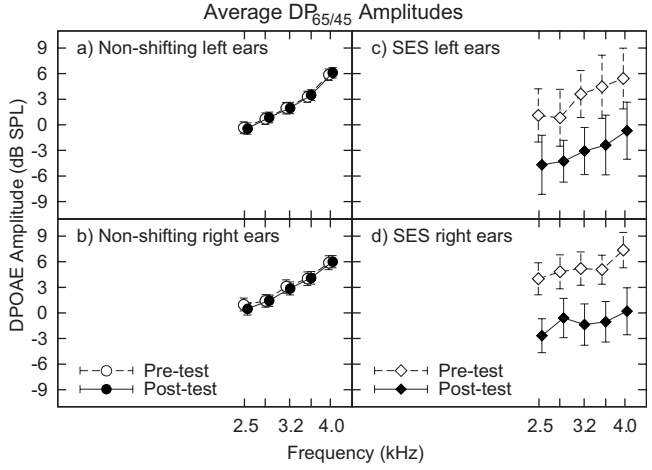


FIG. 4. Average pre-test and post-test $DP_{65/45}$ amplitudes for the noise-exposed group by significant $DP_{65/45}$ shift (SES) status. Average pre-test amplitudes for the SES ears (15–16 left ears and 24–26 right ears) were slightly higher than the average for the no-SES ears (201–206 left ears and 188–192 right ears). Average post-test $DP_{65/45}$ amplitudes decreased by approximately 6 dB from pre-test for the SES ears, while the post-test average for the no-SES ears stayed essentially the same. The error bars are 95% confidence intervals. The number of ears contributing to the average at each frequency varied because of some unusable data.

average pre-test amplitudes to be higher for the SES ears compared with the no-SES ears, but this may be due solely to the small N. In general, these graphs indicate that the method used to determine the SES status was appropriate.

E. Comparison of STS and SES

Table V shows the resulting 2×3 matrices for the STS and no-STS ear versus the SES, no-SES, and unknown-SES ears. The amount of data is small in some cells, and is also unevenly balanced, so it is important to not overinterpret the findings. The count of the SES ears is likely to be an underestimate. First, many potential SES ears are in the unknown-SES category because the SES is masked by noisy measurements. Second, ears with low-level or absent OAEs at pre-test cannot show a SES at post-test; these ears are examined in more detail in Sec. III F.

The nonparametric phi coefficient (Siegel, 1956) was used as a measure of association for the 2×2 matrices to determine whether STSs and SESs tended to occur together in the same ear (the unknown-SES category was not included). The phi coefficient can be interpreted similarly to a correlation coefficient and can be used for small data sets. Coefficients below 0.35 are considered to indicate no more than trivial associations (Fleiss *et al.*, 2003). There was essentially no association between the STS status and the SES status for TEOAEs (left ears, phi=0.25; right ears, phi =0.22), $DP_{65/45}$ (left ears, phi=0.11; right ears, phi=0.10), or $DP_{59/50}$ (left ears, phi=0.18; right ears, phi=0.02).

To further assess whether STSs and SESs were associated, conditional probabilities were considered for the ears in Table V. As shown in Table VI, in general, the probability (P) of a SES in an ear was higher than the probability of a STS. Further, $P(\text{STS}|\text{SES})$, which is the conditional probability of the STS in the subgroup of the SES ears, was higher than the $P(\text{STS})$, which is the STS base rate, indicating that STS ears were overrepresented among the SES ears. Going the other way, $P(\text{SES}|\text{STS})$ which is the conditional probability of the SES in the subgroup of the STS ears, was higher than $P(\text{SES})$, which is the SES base rate, indicating that SES ears were overrepresented among the STS ears. In the STS ears, TEOAEs tended to show SES more than did DPOAEs. Because of the small numbers in some of the cells, any further analysis would be inappropriate.

F. OAE predictors of susceptibility to NIHL

Pre-test OAE amplitudes were used as predictors of the STS status in the noise-exposed ears.[12] There were two groups of interest: the 42 STS ears and the 528 no-STS ears. Due to potential differences in the NIHL susceptibility in the left and right ears, they were kept separated in the analyses. As described earlier, where possible, pre-test OAE amplitudes below the noise floor were estimated with the noise-floor level, providing the noise floor was sufficiently low. Between 17 and 21 STS ears and 217–263 no-STS ears contributed to the analysis for each frequency, OAE type, and ear.

The positive predictive value (PPV) (Zhou *et al.*, 2002) is the conditional probability of an ear from the noise-

TABLE V. STS vs SES matrices. The first number is the count, and the number in parentheses is the overall percentage. For the left and right ears separately, ears were grouped by whether they were classified as STS and/or SES ears. The unknown-SES category is for those ears where there were unusable data; these ears were not used in the analysis of the matrices.

		SES status					
		Left ear			Right ear		
OAE type	STS status	No-SES	SES	Unknown-SES	No-SES	SES	Unknown-SES
$DP_{59/50}$	No-STS	174 (61)	14 (5)	76 (27)	161 (56)	24 (8)	79 (28)
	STS	11 (4)	4 (1)	6 (2)	11 (4)	2 (1)	8 (3)
$DP_{65/45}$	No-STS	191 (67)	13 (5)	60 (21)	178 (62)	22 (8)	64 (22)
	STS	15 (5)	3 (1)	3 (1)	14 (5)	4 (1)	3 (1)
TEOAE	No-STS	111 (39)	21 (7)	132 (46)	130 (46)	16 (6)	118 (41)
	STS	7 (2)	7 (2)	7 (2)	8 (3)	5 (2)	8 (3)

exposed group *with* STS after basic training, *given* a test result of a low-level OAE. A low-level OAE is defined as an OAE amplitude that is less than a cutoff value. By varying the cutoff value over the entire range of OAE amplitudes, the entire PPV function may be generated. Once the entire PPV function is known, an optimal cutoff point may be chosen to define a "low-level" OAE (which may vary depending on the purpose, and the outcomes associated with the diagnosis). The PPV is also known as the *a posteriori* conditional probability: $P(\text{STS}|\text{OAE} \leq \text{cutoff})$.

Figure 5 shows the PPV as a function of OAE amplitude for each ear, OAE type, and frequency. Without knowledge of the OAE level in a given ear, there was a probability of around 0.07–0.08 that the ear would be classified with STS. With knowledge of the OAE level, the probability an ear would be classified with STS rose to a maximum of 0.67, indicating an eight- to ninefold increased risk for STS. TEOAEs at 2.8 and 4 kHz tended to be better predictors for the left ears, with OAE amplitudes below approximately −5 dB SPL indicating an increased risk for STS. DPOAEs at 4 kHz tended to be better predictors for the right ears, with OAE amplitudes below approximately −5 to −10 dB SPL indicating an increased risk for STS. To summarize the risk, Table VII provides the maximum increased risk (maximum PPV/base-rate over the OAE amplitude) across all OAE types by frequency.

To relate these figures to the percentage of volunteers at increased risk, the PPV functions are replotted in Fig. 6 for each ear and each OAE type at 4 kHz after transforming OAE amplitudes into percentiles. For the left ear (but not the right ear), TEOAEs at 4 kHz show an increased risk for STS in the bottom quartile, whereas for the left ear (but not the right ear), DPOAEs at 4 kHz show an increased risk for STS in the bottom decile.

G. Susceptibility to NIHL for volunteers rather than ears using "worst ear" as a predictor

In a clinical situation, the focus is more on the risk of STS for an individual person, rather than individual ears. One way to use the information from both ears is to take the results from the worst ear (the ear with the lowest OAE amplitude) and use that as the predictor for the STS risk. Figure 7 shows the results of such an analysis for the two best DP frequencies and the three best TEOAE frequencies. For each noise-exposed volunteer and for each OAE type

TABLE VI. STS ears are over-represented in the group of SES ears, compared with the probability (P) of a STS in general. Likewise SES ears are over-represented in the group of STS ears, compared with the probability of a SES in general. This finding holds over the left and right ears and for all three OAE types, except for $DP_{59/50}$ in the right ears, where representation was proportional, and where OAEs had the highest variability. The pooled category represents the results pooled over the ear and the OAE type. The true SES rate is underestimated because there are likely to be some unidentified SES ears in the large group of unknown-SES ears, where SES status could not be determined at all three frequencies. (For the underlying cell counts, see Table V.)

OAE type	Ear	P(STS)	P(STS|SES)	STS represented in SES group	P(SES)	P(SES|STS)	SES represented in STS group	P(unknown-SES)
$DP_{59/50}$	Left	0.07	0.22	Over	0.09	0.27	Over	0.29
	Right	0.07	0.08	Proportionally	0.13	0.15	Proportionally	0.31
$DP_{65/45}$	Left	0.08	0.19	Over	0.07	0.17	Over	0.22
	Right	0.08	0.15	Over	0.12	0.22	Over	0.24
TEOAE	Left	0.10	0.25	Over	0.19	0.50	Over	0.49
	Right	0.08	0.24	Over	0.13	0.38	Over	0.44
Pooled		0.08	0.19	Over	0.12	0.27	Over	0.33

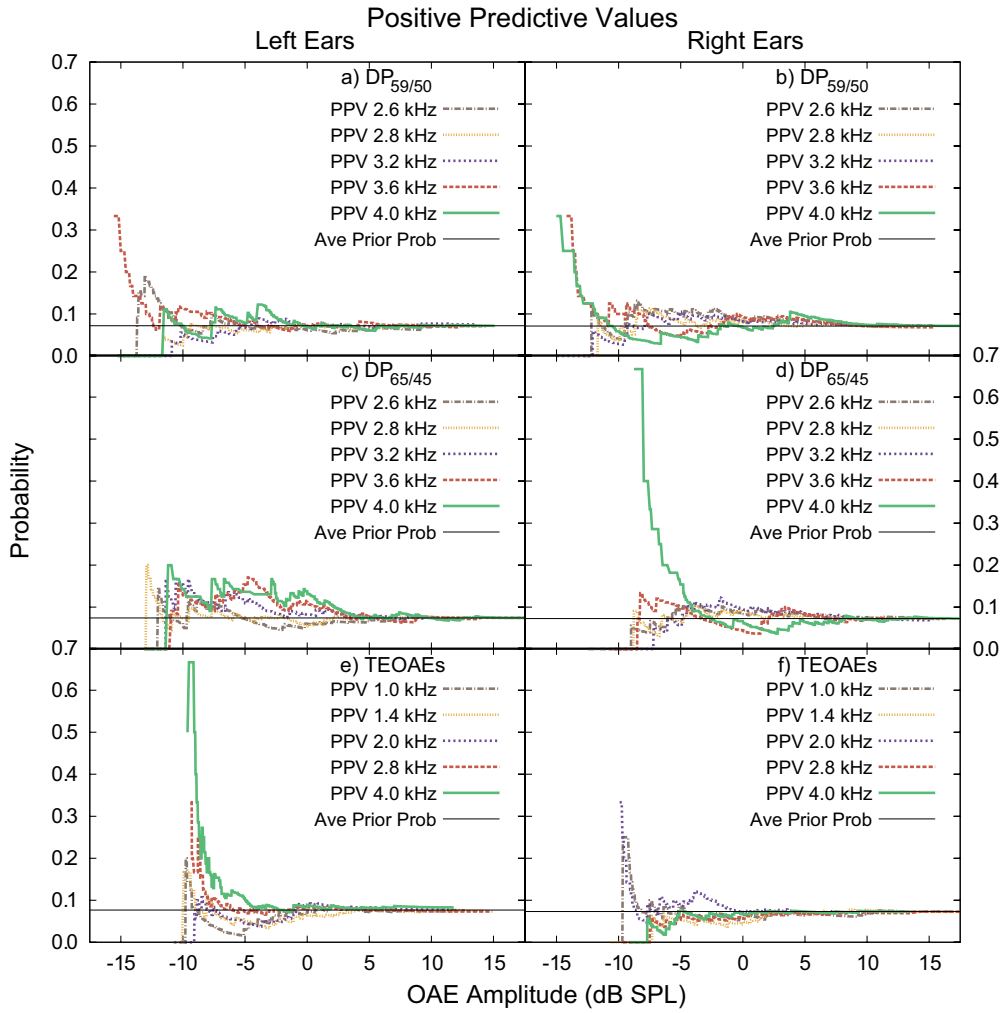


FIG. 5. (Color online) PPV, for the left and right ears separately, as a function of the PPV criterion, which is OAE amplitude (in dB SPL). PPV is the probability that an ear was classified with a STS given an OAE amplitude less than the criterion. As the OAE amplitude decreased, PPV tended to increase for the higher-frequency bands, but not for all OAE types and frequencies. [(a) and (b)] $DP_{59/50}$,[(c) and (d)] $DP_{65/45}$, and [(e) and (f)] TEOAEs. The thin solid horizontal line represents the prior probability of a STS averaged over the displayed frequencies.

and frequency, the lowest OAE amplitude of the two ears was chosen, or if there were valid data for only one ear then that ear constituted the worst ear. If there were no valid data for either ear, the volunteer was not included in the analysis at that test point. Note that unlike the analyses so far, the contralateral ears of the 30 volunteers with unilateral STS were included with the STS ears rather than with the no-STS ears (in the cases where that ear had the lowest OAE amplitude). When choosing the worst ear, TEOAEs, especially at 4 kHz, were the best predictor of incipient NIHL.

IV. DISCUSSION

A. OAEs are more sensitive than audiometric thresholds to noise exposure

The repeated-measures ANOVA indicated that OAEs were more sensitive to noise-induced changes to the inner ear than were audiometric thresholds. Both DPOAEs and TEOAEs showed significant decreases in OAE levels after the noise exposure, but there was no change in audiometric thresholds for the subgroup of 60 noise-exposed volunteers with complete data sets. The 2×2 matrices for SES versus

TABLE VII. Maximum increased risk for STS (PPV/base-rate) for each OAE type and frequency, by ear. Each number represents how many times more likely a STS is given a low pre-test OAE result relative to the base rate.

OAE type	Frequency (kHz)	Maximum increased risk	
		Left ears	Right ears
$DP_{59/50}$	2.5	2.6	1.8
	2.8	1.1	1.6
	3.2	1.2	1.5
	3.6	4.7	4.9
	4.0	1.7	4.7
$DP_{65/45}$	2.5	1.9	1.4
	2.8	2.7	1.4
	3.2	2.3	1.7
	3.6	2.3	1.9
	4.0	2.7	9.2
TEOAE	1.0	2.6	3.4
	1.4	2.2	1.1
	2.0	1.4	4.4
	2.8	4.6	1.0
	4.0	8.1	1.0

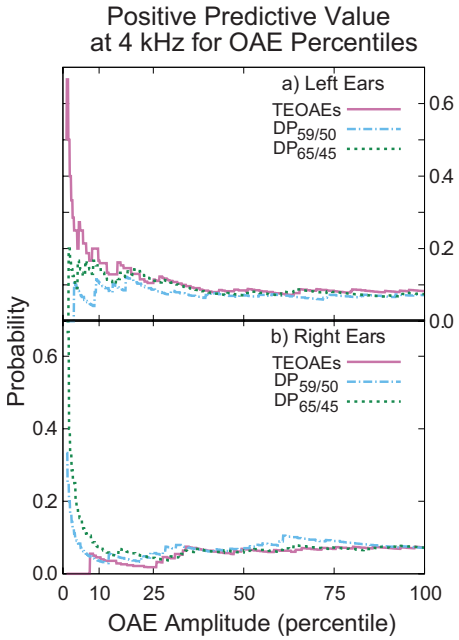


FIG. 6. (Color online) PPV at 4 kHz (from Fig. 5). for the left and right ears separately, for each OAE type replotted as a function of OAE amplitude in percentiles.

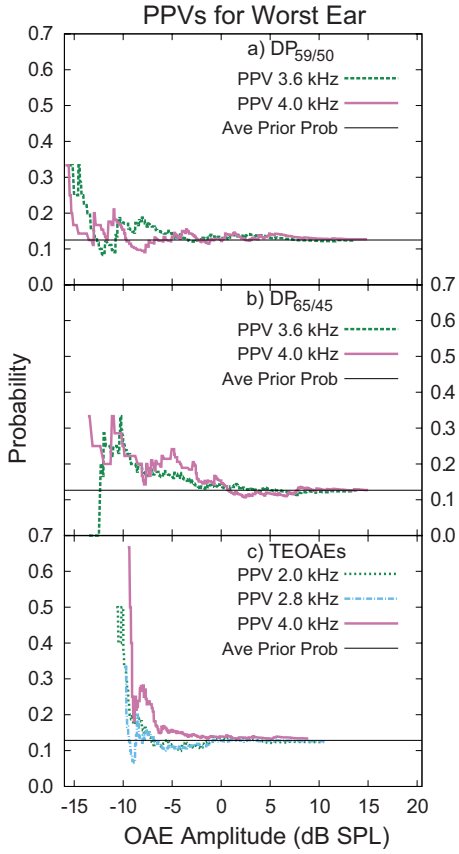


FIG. 7. (Color online) PPVs as a function of PPV criterion, which in this case is OAE amplitude for the worst ear, which is the way it would be implemented in occupational audiology programs. For each volunteer, the ear with the lowest OAE amplitude was used as the predictor. (a) $DP_{59/50}$ at 3.6 and 4 kHz, (b) $DP_{65/45}$ at 3.6 and 4 kHz and (c) TEOAEs at 2, 2.8, and 4 kHz. The thin solid horizontal line represents the prior probability of a STS averaged over the displayed frequencies.

STS status indicated a higher SES rate compared with the STS rate in the noise-exposed ears, and there was a tendency for the STS ears to also have SESs and for SES ears to also have STSs. TEOAE SES status showed more consistency with the STS status than with the DPOAEs, despite the larger amount of unusable data with the TEOAEs. These findings are consistent with Lapsley Miller *et al.* (2006), where the same experimental protocol was used, and with many other longitudinal studies that also showed changes in OAEs without accompanying changes in audiometric thresholds (Engdahl *et al.*, 1996; Murray *et al.*, 1998; Murray and LePage, 2002; Konopka *et al.*, 2005; Seixas *et al.*, 2005a; Duvdevany and Furst, 2006).

These results could be due to on-frequency inner-ear damage in the 2–4 kHz range that causes subclinical changes insufficient to affect audiometric thresholds but to which OAEs are sensitive. This is consistent with observations in animals that damage to outer hair cells (OHCs) can be extensive with no concomitant change in audiometric thresholds (Hamernik *et al.*, 1989; Hamernik *et al.*, 1996), and consistent with the theory that there is OHC redundancy (LePage *et al.*, 1993), where it is thought that there are many more OHCs than what is required for normal hearing. The OHC loss therefore shows up in OAE measurements before audiometric threshold measurements because OAE measurements more directly measure OHC activity. Alternatively, the results could be due to unmeasured higher-frequency inner-ear damage (that may or may not affect high-frequency hearing thresholds) that affects OAEs measured at lower frequencies, but not audiometric thresholds at those lower frequencies. This higher-frequency damage might influence the transmission of a lower-frequency OAE out to the middle ear (Lonsbury-Martin and Martin, 2007). Furthermore, with some OAE stimulus configurations (containing high-frequency energy), high-frequency damage could lessen the distortion-component OAE from the high-frequency place that creates a lower-frequency stimulus-frequency OAE (SFOAE),[13] which can interact with OAEs generated at that lower frequency. It was not possible in the current study to measure high-frequency hearing thresholds or high-frequency OAEs, making it difficult to disentangle the two theories; however, the results of others offer some clues.

For DPOAEs in humans, diminished OAE amplitudes without an accompanying hearing loss in the same frequency region have been associated with a hearing loss at higher frequencies (Arnold *et al.*, 1999; Dorn *et al.*, 1999). Arnold *et al.* (1999) used 50 subjects, the majority of whom were males, with normal hearing (<20 dB HL from 0.25 to 8 kHz) and ages from 17 to 37 years. They reported minimal noise exposure, but people, particularly males, living in modern civilizations do tend to accumulate damage from noise exposure. The multivariate analyses of Dorn *et al.* (1999) included hundreds of subjects, age 1–96 years, and hearing levels from −5 to 120 dB HL at 0.75–8 kHz. In contrast, Schmuziger *et al.* (2005) minimized the effects of previous noise exposure by using younger subjects (age 16–19 years), and fewer males (38%) in the group (all with normal hearing, as well as reports of minimal previous noise exposure). The high-frequency (8–16 kHz) thresholds for this group were only minimally related to lower-frequency

DPOAEs. In rodents, Withnell and Lodde (2006) did not see lower-frequency DPOAE amplitude losses when higher-frequency regions were damaged by noise. These results suggest that the more likely explanation for the influence of high-frequency thresholds on much lower-frequency DPOAEs is subclinical damage at the lower frequency.

For TEOAEs in humans, decreased OAE amplitudes without an accompanying audiometric-threshold decrement in the same-frequency region also were associated with a hearing loss at higher frequencies (Avan *et al.*, 1997; Konopka *et al.*, 2005). The subjects of Avan *et al.* (1997) (nearly half of whom were males) had normal hearing up to 4 kHz, and were older—ages 24–54 years. The group from Konopka *et al.* (2005) consisted of 92 young, male, noise-exposed soldiers. In a young population with less noise exposure, and with normal hearing up to 8 kHz, high-frequency (8–16 kHz) thresholds were not related to lower-frequency TEOAEs (Schmuziger *et al.*, 2005).

Yates and Withnell (1999), using a novel measurement technique that allowed the measurement of high-frequency TEOAES, observed that TEOAEs evoked by a high-pass click included frequencies lower than those in the stimulus in guinea pigs. The generation of new frequencies that were not present in the stimulus implies that the OAEs were generated from a distortion mechanism; these OAEs also act as a stimulus that elicits reflection-component SFOAEs at lower frequencies. After noise exposure, which damaged the high-frequency region in guinea pigs, lowered TEOAE amplitudes were found not only at the higher frequency where the eighth-nerve compound-action-potential (CAP) thresholds were lowered, but also at lower frequencies where the CAP thresholds were not lowered (Withnell *et al.*, 2000). In humans, however, not only is the distortion mechanism relatively smaller than it is in guinea pigs (Shera and Guinan, 1999), but the method used for clinical TEOAE measurements windows out the first few milliseconds of the TEOAE to reduce stimulus artifact (Bray and Kemp, 1987). This leaves only the TEOAE reflection component (Knight and Kemp, 1999; Kalluri and Shera, 2007; Sisto *et al.*, 2007; Withnell *et al.*, 2008). Furthermore, behavioral thresholds at ultrahigh frequencies would not influence most TEOAE measurements with humans because the TEOAE stimulus usually does not have much energy above 5 kHz. With a TEOAE stimulus that extends up to 5 kHz, the TEOAE and SFOAE spectra in individual ears are nearly identical, at least up to 2.4 kHz, implying that with TEOAEs, the lower-frequency SFOAE that gets generated due to the higher-frequency distortion component does not have much effect on the lower-frequency SFOAE that is solely generated from that place (e.g., Kalluri and Shera, 2007). The results from these TEOAE studies also suggest that on-frequency subclinical damage is the predominant reason why OAE amplitudes can decrease when hearing levels remain unchanged in humans.

Studies that have shown TEOAE decrements in individual ears to be broader than DPOAE decrements may be indicative of TEOAEs being more sensitive to subclinical damage than DPOAEs (e.g., Lapsley Miller *et al.*, 2004; Lapsley Miller *et al.*, 2006), consistent with Shera's (2004) view that the reflection component should be more sensitive to noise damage. TEOAEs as we typically measure them are primarily reflection mechanism, and DPOAEs are a mix of the two mechanisms. Therefore, it is not surprising that DPOAEs and TEOAEs often are equally sensitive for groups, but TEOAEs tend to edge out DPOAEs in sensitivity for individual ears.[14]

There are some studies that do not show OAEs as being more sensitive than audiometric thresholds. Lapsley Miller *et al.* (2006) found significant changes in group audiometric thresholds along with changes in OAEs, but there was little consistency between changes in thresholds and OAEs in individual ears. Duvdevany and Furst (2007) measured hearing and TEOAEs in the same individuals annually three times—neither hearing nor TEOAEs changed in the second year, but both did in the third year. Their TEOAE stimulus was 84 dB pSPL, which would not be maximally sensitive to noise damage.[15]

Aging and sex differences can be discounted in the current study because all the participants were young men and the study duration of 13 weeks was too short for aging to have any measurable impact. Nor is audiometric resolution an explanation for the greater sensitivity of OAEs to noise-induced changes in the inner ear. The standard clinical protocol, which produces a resolution of 5 dB, may hinder the detection of small changes in audiometric thresholds, even in the group average. However, if the only reason for the difference between OAEs and audiometric thresholds is resolution, all the STS ears that were identified should also show SESs, but this was not the case. Even within the subset of ears with both SESs and STSs, there was not much consistency across frequency and OAE type (results not reported here in detail).

The typical finding for the STS ears was either an accompanying SES (not necessarily across all OAE types) or low-level or absent OAEs. For a couple of ears where there was STS but no SES and normal OAEs, the STS was small and possibly a false positive (it was not possible to do a confirmation audiogram).

B. STS and SES criteria

The SE_{meas} values underlying the STS criteria were smaller here than in Lapsley Miller *et al.* (2006), perhaps because the current study used double-walled sound-attenuating chambers, whereas the earlier study used single-walled chambers in a noisier shipboard environment. The current STS criteria were identical to those developed in Lapsley Miller *et al.* (2004), where the testing environment was similar.

If possible, it is important to derive STS (and SES) criteria from a control group tested in the same environment so that there is some certainty that the shifts are significantly different from test-retest variability. For instance, in the current study, using all the derived STS criteria (Table II), STS was detected in 36 noise-exposed volunteers (42 ears) or 12.6% of volunteers (7.4% of ears). If the strict clinical criterion that is commonly used by regulatory agencies was used instead, which is an average shift at 2, 3, and 4 kHz of

at least 10 dB (Mining Safety and Health Administration, 1999; Department of Defense, 2004; Federal Railroad Administration, 2006; Occupational Safety & Health Administration, 2007), STS would have been detected in only 17 volunteers (18 ears) or 5.9% of volunteers (3.2% of ears). The opposite can also happen, where the criteria chosen are too lax. For instance, if someone arbitrarily chose a STS criterion of a 10 dB shift at any single frequency (without any reference to a control group), in our current study STS would have been detected in 87 volunteers (109 ears) or 30% of volunteers (19.1% of ears). Most of these shifts are false-positive STSs because the criterion was less than the test-retest variability.

The detection of a STS does not necessarily mean the ear had a hearing loss as there is a chance that the STS was a false positive. An upper limit for our false-positive rate (for STS at three single frequencies and three averaged frequencies; assuming independence of frequencies) is approximately 11% (for both positive and negative STSs). The probability of a false positive across six frequencies is one minus the probability of no false positives at any frequency. Because our STS criteria were based on a 98% confidence interval (see footnote 7), this is $1-0.98^6$. The actual false-positive rate is likely to be lower because (a) we looked only at positive STS, so the false-positive rate would be only 5.5% at most, (b) the frequencies are correlated, and (c) the 5 dB resolution of the audiogram meant that we rounded up the raw criteria based on multiples of the SE_{meas} to the next largest available step so the probability of a false positive at each frequency was lower. This is borne out when applying the STS criteria to the control group where there were no STSs detected. In the larger noise-exposed group, however, we would expect some of the STSs to be false positives, and indeed there are some shifts that are not compelling from a clinical viewpoint (e.g., a shift only at 2 kHz). A stricter STS criterion, however, would have meant missing more true STSs. It was unfortunate that the recruits' schedule did not allow time for immediate retesting of STSs, which would have decreased the false-positive rate.

NIOSH (National Institute for Occupational Safety and Health, 1998) suggested a different STS criterion—a 15 dB shift at any tested frequency (0.5, 1, 2, 3, 4, 6, or 8 kHz), with an immediate retest being optional. ASHA (American Speech-Language-Hearing Association, 1994) also suggested a significant change criterion (for monitoring ototoxic hearing loss) greater than 15 dB at any one frequency or greater than or equal to 10 dB at two or more adjacent frequencies. The current study, as well as previous ones (e.g., Marshall and Hanna, 1989; Lapsley Miller *et al.*, 2004) found that the SE_{meas} at a single frequency varies as a function of frequency, with lower and higher frequencies having a larger SE_{meas}. Shaw (1966) demonstrated that supra-aural earphone-placement variations have the largest effect at these frequencies. For our data, the NIOSH criterion was the same as ours at 1–4 kHz (National Institute for Occupational Safety and Health, 1998), but too lax at 0.5 and 6 kHz. The ASHA criterion was too strict for 1–4 kHz, but applicable above and below that, as well as for the 10 dB two-frequency average criterion.

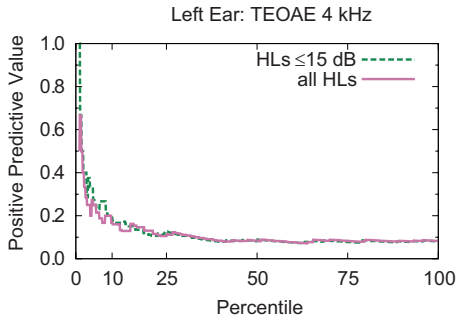


FIG. 8. (Color online) PPV as a function of OAE amplitude in percentiles for TEOAEs at 4 kHz for the left ears from Fig. 6 (solid line) compared with the same data after excluding ears with audiometric thresholds >15 dB HL (dashed line).

The NIOSH suggestion of retesting immediately following a STS is a good one. For example, in the current study, we estimated an upper bound for the false-positive STS rate to be 5.5%. If a STS is retested, then we would expect the STS rate to diminish to 0.5% (probability of a STS over six frequencies multiplied by the probability of a STS at one frequency, 0.055×0.01; assuming only positive STSs are of interest). Basing the STS criterion on the known test-retest reliability of a test situation is necessary for control of false-positive STSs. Our STSs are a better estimate of true STS than may be the case when arbitrary values are chosen.

The SE_{meas} values underlying the TEOAE SES criteria were similar to Lapsley Miller *et al.* (2004), but were slightly larger than in Lapsley Miller *et al.* (2006), where the equipment was run on battery power more often, which tended to produce a lower noise floor. The SE_{meas} values underlying the DPOAE SES criteria were also slightly larger than in Lapsley Miller *et al.* (2006), but could not be directly compared to Lapsley Miller *et al.* (2004) because here they were based on measurements at individual frequencies, rather than averaged within half-octave bands. In general, the SE_{meas} values were mostly comparable to those reported elsewhere (Franklin *et al.*, 1992; Beattie and Bleech, 2000; Beattie, 2003; Seixas *et al.*, 2005b; Wagner *et al.*, 2008).

C. Susceptibility to NIHL from impulse noise

In the analyses discussed so far, all ears used had pre-test OAE amplitudes that were measurable. Many of the ears that ended up in the unknown-SES category had OAE signal-to-noise ratios that did not meet the criteria for presence. The second thrust of the analyses showed that low-level pre-test OAE amplitudes were predictive of subsequent STS status for some OAE types, frequencies, and left and right ears (Fig. 5). The increased risk of a STS for those ears with low pre-test OAE levels cannot be explained by pre-test audiometric-threshold differences between the STS ears and the no-STS ears as there was essentially no difference between pre-test audiometric thresholds for these groups, as shown in Fig. 1. To further illustrate this point, the analysis underlying Fig. 6 for TEOAEs at 4 kHz, for left ears, was rerun after excluding all ears with audiometric thresholds >15 dB HL at 4 kHz (see Fig. 8). Seventeen no-STS ears and zero STS ears were excluded using this new criterion.

The exclusion of these ears only enhances the finding that pre-test OAE levels can be predictive of subsequent STS. This is supported by an animal study by Perez *et al.* (2004) that showed that ears with existing hearing loss were less likely to get further hearing loss. This implies that ears with low-level OAEs and hearing loss would be less likely to get further hearing loss. It is those ears with low-level OAEs and normal hearing that are at risk.

The OHC redundancy theory described earlier (LePage *et al.*, 1993) is consistent with these findings. Earlier subclinical damage to some of the OHCs would show up as low-level or absent OAEs, but not necessarily as hearing loss. Further noise exposures damaging further OHCs would then be more likely to lead to hearing loss compared to the ears with more intact OHCs.

Animal studies also provide some clues as to why low-level OAEs could be associated with an increase in the likelihood of future PTS. Chinchilla studies have shown that small amounts of OHC loss have a more significant effect (reduction) on DPOAE amplitude levels than on measures of threshold sensitivity, suggesting that OAEs may also indicate an early onset of cochlear damage in humans (Davis *et al.*, 2005). These results also suggest that OAEs be considered, within the context of hearing-conservation practices, as a complement to existing hearing-threshold tests in detecting OHC loss resulting from noise exposure. The results indicate that on the basis of threshold information alone, without information about the OAEs, one might underestimate the sensory cell loss. This conclusion is supported by the results of others, which show up to 30% OHC loss in subjects with less than 10 dB of PTS (Hamernik *et al.*, 1989; Hamernik and Qiu, 2000; Davis *et al.*, 2004). Bohne *et al.* (1987) also showed that 20%–30% OHC loss in the low frequencies was often not accompanied by corresponding behaviorally measured threshold shifts in the chinchilla. They also explained that a relatively large reduction (12–15 dB) in DPOAEs in the presence of smaller OHC losses at some frequencies may be accounted for not only by the OHC loss but also by morphological changes (e.g., cilia defects or intracellular changes) that can affect the function of cells that are present and for which the cochleogram provides no information.

There were differences between the ears, with no one frequency consistently being a good predictor of the STS status across ears. TEOAEs appeared to be a better predictor of the STS status for the left ear, and DPOAEs appeared to be a better predictor of the STS status for the right ear. Except for TEOAEs at 4 kHz in the left ear, STS risk did not increase greatly until the OAE amplitude moved into the bottom decile.

It is unclear why DPOAEs would be a better predictor of a STS in the right ears and TEOAEs a better predictor of a STS in the left ears. The most likely explanation is that the small number of STS ears contributing to each analysis gives some spurious results. There are other possible contributing factors. The recruits did not get in-depth training on proper insertion of hearing-protection devices; indeed, they sometimes reported that the earplug fell out during the live-fire exercise. Duvdevany and Furst (2007) showed large increases in PTS during a time period when hearing-protection devices apparently were not worn much. The variability in noise exposure across volunteers could be large compared to the number of Marine recruits that were in this study.

Although we did not keep track of handedness, we would expect approximately 95% of the group to be right handed.[16] Two studies indicated that the left ear is more susceptible to NIHL than is the right ear, irrespective of handedness (Job *et al.*, 1998; Nageris *et al.*, 2007). In the current study, there was an equal number of STSs in the left ear and the right ear (21 ears for each, including 6 ears with bilateral STS). This indicates that there was considerable noise in the environment (more than just from firing one's own gun), as well as the poor quality of the hearing protection.

Figure 1 indicates that on average the left ear STSs were broader than the right ear STSs, suggesting that the left ear suffered more extensive damage from the impulse-noise exposures than the right ear. The implication is that low-level TEOAEs could be a better predictor of broadband STS and that low-level DPOAEs might be a better predictor of narrowband STS. If this is true, it may be due to higher-order physiological asymmetries (e.g., efferent innervations) that somehow treat tonal stimuli differently from click stimuli depending on which ear the stimuli are presented to (e.g., Sininger and Cone-Wesson, 2004), but any mechanism is speculative at best. Furthermore, the DPOAE results may be greatly affected by the measurements and data analysis being seen at single widely-spaced frequencies, making it difficult to determine if a low-level DPOAE is just due to the test frequency coinciding with a null in the DPOAE microstructure in what is otherwise a strong DPOAEgram.

Even though we suspect that the apparent differences are due to the relatively small amount of data, the results are suggestive enough that future studies should continue to look at ear differences. If the ear differences seen in this study are also found in future studies, it will be important to parse out whether the differences are due to external factors (e.g., noise exposures), innate factors (e.g., efferent innervation), existing preclinical damage, or methodological idiosyncrasies.

D. Susceptibility to NIHL from impulse noise compared with continuous noise

It is of interest to compare the results of the susceptibility analysis to that in Lapsley Miller *et al.* (2006), as it is the only other known analysis that considers if OAE amplitude is a predictor of NIHL. Likelihood ratios were used to make this comparison, rather than PPVs, because the PPV is dependent on the prior probability of STS/PTS, which differed across the two studies. The likelihood ratio is a ratio of two probabilities: the probability of a particular test result among patients with a condition to the probability of that particular test result among patients without the condition (Zhou *et al.*, 2002). In the current context, the likelihood ratio indicates the relative probability that a pre-test OAE amplitude was below a given percentile in the group of ears that subsequently were classified with STS, relative to the same result in the group of ears that did not.

To ensure a fair comparison with Lapsley Miller *et al.* (2006), ears for the current study were combined, but only

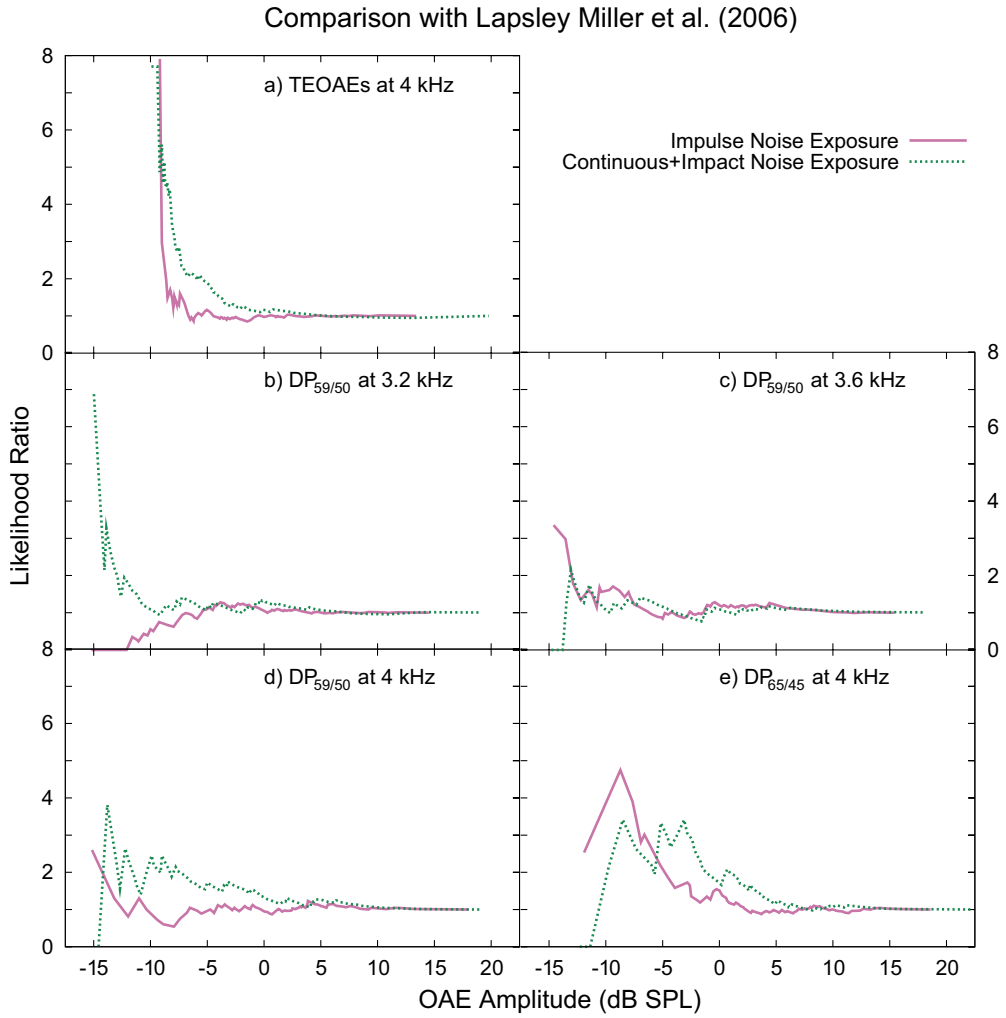


FIG. 9. (Color online) Comparisons of likelihood ratio as a function of OAE amplitude (in dB SPL), indicating susceptibility to noise-induced hearing loss, between the current study (Marine recruits exposed to impulse noise, solid line) and Lapsley Miller *et al.* (2006) (deployed aircraft carrier sailors exposed to continuous noise overlaid with impact noise, dashed line) for the OAE test frequencies where there were no large differences in the amplitude distributions between the ears: (a) TEOAEs at 4 kHz (half-octave band), (b) $DP_{59/50}$ at 3.2 kHz, (c) $DP_{59/50}$ at 3.6 kHz, (d) $DP_{59/50}$ at 4 kHz, and (e) $DP_{65/45}$ at 4 kHz.

for those test frequencies which showed little to no difference in OAE amplitude percentiles between ears.[17] These included TEOAEs at 4 kHz, $DP_{65/45}$ at 4 kHz, and $DP_{59/50}$ at 3.2, 3.6, and 4 kHz. Figure 9 shows likelihood ratio as a function of OAE amplitude. For the current study, likelihood ratios decreased when ears were combined, due to the asymmetry of results between ears. Compared with Lapsley Miller *et al.* (2006), likelihood ratios for the DPOAEs were comparable at 4 kHz. At lower frequencies, the likelihood ratios in the earlier study were higher than for the current study. The biggest difference was for TEOAEs at 4 kHz, with the likelihood ratios in the earlier study showing increased risk for ears with OAE amplitudes below the 50th percentile (approximately 0 dB SPL), whereas for the current study, risk did not increase until about the 5th percentile (approximately −8.5 dB SPL). The maximum likelihood ratio, however, was essentially the same at around 8.

Overall, the general trend across both studies is for low-level OAEs to be predictive of subsequent PTS and/or STS. In both studies, OAEs—particularly TEOAEs—in the 4 kHz region were the best predictors. Larger studies with many more PTS/STS ears are essential to better establish this relationship. We expect this predictive power to be greater in situations with continuous noise than in situations with impulse noise due to greater variability of the sound power of the sound source reaching the inner ear especially in environments where much gunfire is in the general environment. This supposition is supported by the comparison shown in Fig. 9 where NIHL risk for impulse noise showed up for much lower TEOAE amplitudes, compared with continuous noise. Further, if hearing protectors are not worn (or if they fit poorly) for impulse-noise exposures, we expect the predictive power to diminish even more because there may be sufficient damage over a relatively short amount of exposure time to cause hearing loss irrespective of whether or not there were previously missing OHCs.

When considering both the current and the earlier study, in general, TEOAEs were better predictors than DPOAEs (however, the ear asymmetry shown in the current study indicates that DPOAEs cannot be discounted). There are also a number of mostly practical pros and cons for choosing one OAE type over another. If only a few DPOAE frequencies are tested (which makes the test faster—an important consideration in field studies with humans), it is possible that the

test point will fall into a null of the DPOAE microstructure. To obtain results that are independent of the microstructure, many points per octave need to be measured and averaged (Kemp, 2007), and there are further issues when it comes to combining across frequencies if there are unusable data, particularly when comparing measurements where data may be unusable at different frequencies in different measurements.

Furthermore, the DPOAE measurement is a mix of both reflection-source and distortion-source OAEs (Shera and Guinan, 1999), whereas TEOAEs as measured in humans are essentially reflection source (Withnell *et al.*, 2008). Techniques to separate out the two sources for DPOAEs exist (e.g., Long and Talmadge, 1997; Heitmann *et al.*, 1998; Talmadge *et al.*, 1999; Konrad-Martin *et al.*, 2001; Dhar and Shaffer, 2004; Shaffer and Dhar, 2006; Long *et al.*, 2008), but either are not yet implemented and tested for clinical applications or else have inherent limitations for clinical application (Dhar and Shaffer, 2004). Although in our studies there were more unusable TEOAE data compared with DPOAE data, modern instruments have lower noise floors and faster data collection allowing for more averaging, so we anticipate unusable data to be less of a problem in the future.

E. Concluding remarks

OAEs are predictive of incipient NIHL. It is unknown whether prior noise exposures or innate factors explain why some normal-hearing ears had low-level or absent OAEs. Most recruits indicated that they had prior noise exposures typical of a modern lifestyle, including weapons' fire, amplified music, and machinery noise. Regardless, having a test that indicates ears susceptible to noise-induced hearing loss is a boon for hearing conservation and audiology in general. The current study extends the earlier findings to include impulse noise.

If identifying those individuals and groups most at risk for hearing loss from noise exposure in their near future is possible by detecting the early stages of inner-ear changes, then steps can be taken to prevent or mitigate further damage. While the auditory medial-olivocochlear-bundle reflex (MOCR) may be another way to assess future risk (Maison and Liberman, 2000; Backus and Guinan, 2007), there is at present no test in humans that sufficiently differentiates large and small AERs within the test time available for clinical testing. Furthermore, such a test requires OAEs with reasonable amplitude, thereby precluding the use of such a test in many noise-exposed individuals who do not have strong OAEs. In the future, a very powerful predictive OAE test battery might consist of both OAE level and MOCR strength.

ACKNOWLEDGMENTS

Thanks to Linda Westhusin, Michael McFadden, Denise Cline, Jackie Adler, Joy Houston, and Brian Ferris for their assistance with data collection. A special thanks to the staff and recruits of the Marine Corps Recruit Depot (MCRD) San Diego and Charles Jackson of the Naval Medical Center San Diego Occupational Audiology Department. Thanks to Tom Taggart for his input into the overall experimental design, help with logistics, and feedback on preliminary analyses. Thanks to Chris Shera for helpful discussions on the theoretical aspects. Thanks to the two anonymous reviewers whose considered opinions substantively improved the manuscript. This research was supported primarily by grants from the Office of Naval Research. The views expressed in this article are those of the authors and do not reflect the official policy or position of the Department of the Navy, the Department of Defense, or the United States Government.

[1]In laboratory studies on humans, only TTSs can be studied, and there is typically a close relationship between changes in OAEs and changes in audiometric thresholds (Marshall and Heller, 1998; Marshall *et al.*, 2001). However, TTSs and PTSs are physiologically different (Saunders *et al.*, 1985; Slepecky, 1986; Nordmann *et al.*, 2000), so the results from TTS experiments cannot be expected to generalize to PTS. Furthermore, laboratory experiments examining PTS in animals may not generalize to humans (see summary in Lapsley Miller *et al.*, 2004, p. 308). Therefore, to understand PTS in humans, there is no substitute for actually measuring PTS in humans. These human PTS experiments invariably have to be conducted in field settings where there is usually neither the time nor facilities to make measurements comparable in quality to those made in the laboratory. Nevertheless, the stated results have been found repeatedly across a range of studies.

[2]Data from K. S. Wolgemuth from a 1998 study on NIHL from Marine infantry training at Camp Pendleton.

[3]Data from N. Vausse from a 1994 study on NIHL from Army training at Fort Bragg.

[4]Informed consent briefings, conducted by one of the study principal investigators, took place immediately prior to the Marine Corps recruits beginning their medical evaluation on day 2 of basic training. There were approximately 80 recruits in each briefing and Informed Consent forms were passed out prior to beginning the briefing. The recruits were given ample opportunity to ask any questions about the study, and participation was voluntary. The voluntary aspect was made very clear to them given this was a military basic training facility where most activity is mandatory. Approximately 10% of the recruits declined to participate in the study. The volunteers in the control group were also briefed, reviewed the informed consent form, and were asked to participate. The informed consent form was approved under Naval Medical Center, San Diego-approved research protocol No. S-99-085.

[5]It was not possible to control for potential pharmacological influences across subjects, which may have included over-the-counter and prescription medications such as Erythromycin, Motrin, cold medications, and aspirin.

[6]It was not possible to do ANOVA on the control group, because only four volunteers had complete data sets.

[7]As described in Lapsley Miller *et al.* (2006, footnote 6), the SE_{meas} can be used to specify the magnitude of a statistically significant change within an individual (Ghiselli, 1964), and is defined as $SE_{meas}=\sqrt{\frac{1}{2}(s_1^2+s_2^2)(1-r)}$ where s_1^2 and s_2^2 are the pre- and post-test variances, and r is the correlation between pre- and post-tests. Because the focus here is on the difference between pre- and post-tests, ΔSE_{meas} is defined as $\sqrt{2}SE_{meas}$ (Beattie, 2003; Beattie *et al.*, 2003). Multiplying ΔSE_{meas} by an appropriate multiplier then gives the desired confidence interval. Here a multiplier of 2.12 is used, which gives a 98% confidence interval.

[8]13 ears were classified with a STS in just a two- or three-frequency averaged band, and 23 ears were classified with STSs at both individual frequencies and across averaged frequency bands. Six right ears were classified with a STS at only one individual frequency (with no shifts in the contralateral ear). No left ear was classified with a STS at just one frequency.

[9]The other two frequencies in this range (2.8 and 3.6 kHz) were not used as it would increase the false-positive SES rate when detecting SESs for DPOAEs; it was decided that having three frequencies/frequency bands for each OAE type and audiometric threshold would be a fairer balance. Shifts in averaged frequency bands were not considered because too many ears had unusable data at one or more frequencies.

[10]SES status could not be determined if (a) the OAE was below the noise floor on the pre-test; (b) the OAE was below the noise floor on the post-test, and the post-test noise floor was high so that the OAE level could not

be estimated; or (c) data loss on the pre-test or post-test.

[11]In addition, if we had used additional criteria based on averages across frequencies, as we did with STS, the SES rate would be expected to increase, especially as the SES criteria for averaged bands are likely to be smaller thereby allowing smaller wide-band shifts to be detected (Lapsley Miller *et al.*, 2004).

[12]It is not possible to use pre-test audiometric thresholds as predictors here because the volunteers were prescreened for hearing levels. As shown in Fig. 1, pre-test audiometric thresholds were essentially the same for the STS ears compared with the no-STS ears.

[13]SFOAEs are OAEs generated with the same frequency as the evoking tonal stimulus (e.g., Shera and Guinan, 1999).

[14]Our implementation of DPOAEs may have also been a factor. We were trying to capitalize on growth functions (testing various stimulus levels at specific frequencies) when, in hindsight, using more frequencies might have been a better bet. By testing with a sparse frequency spacing it is possible that for some ears, some of the test points fell into a null in the DPOAE microstructure (Shaffer *et al.*, 2003). If the DPOAEs had been tested at a sufficient number of frequencies, then we could average them to get the total energy in a frequency band, which would be equivalent in that way to TEOAEs. Of course, it would be a much slower test than the TEOAE test, at least with current instrumentation. In the future, new methods to measure DPOAEs may enable easier comparisons with TEOAEs, and DPOAEs might be less dependent on their implementation (e.g., Long *et al.*, 2008).

[15]TEOAEs from lower-level stimuli are more sensitive to cochlear changes due to quinine (Karlsson *et al.*, 1991). Data of ours (to be published) from a study using similar methodology to Marshall and Heller (1998) indicate that TEOAEs evoked with a lower stimulus level show greater sensitivity to noise-induced TTS.

[16]For left-handed shooters, an adaptor was used on the rifle range so the recruit would not get hit in the face by a very hot brass shell casing.

[17]There were some differences between the two studies: In Lapsley Miller *et al.* (2006), confirmed PTS ears were compared with no-PTS ears, and ears with unconfirmed STS were not included. Ears were combined in analyses, because there were not enough PTS ears (13 left ears and 5 right ears, including 3 bilateral) to analyze ears separately. The noise exposures tended to be continuous noise overlaid with impact noise. For the current study, unconfirmed STS ears were compared with no-STS ears. There were differences between ears in OAE amplitude at many frequencies, so ears were analyzed separately. The noise exposures were impulse noise. OAE amplitude criteria were created from binning into percentile categories, although the results are plotted as a function of OAE amplitude, not percentile. Further, the study was part of a larger study investigating genetic factors of hearing loss, with only a subset of volunteers receiving OAE testing. A blood sample was taken at study enrollment and used to identify those volunteers with the Connexion 26 (35delG) GJB2 polymorphism. Those volunteers with this polymorphism were asked back for OAE testing if they had not already been tested. Thus, the sample had a higher percentage of volunteers with 35delG (13 out of 285; 4.6%) than the general population. Initial analyses did not indicate that the 35delG group was in any way different to the other volunteers, so they were included in all analyses. Only two 35delG volunteers were classified with a STS, which was not sufficient to establish if 35delG was a predictor of STS. Recently, evidence has come to hand showing no relationship between 35delG and NIHL (Van Eyken *et al.*, 2007).

American Speech-Language-Hearing Association (**1994**). "Audiologic management of individuals receiving cochleotoxic drug therapy [guidelines]," ASHA **34**, 11–19.

ANSI (**1991**). *Maximum Permissible Ambient Noise Levels for Audiometric Test Rooms (ANSI S3.1)* (American National Standards Institute, New York).

ANSI (**1996**). *Specifications for Audiometers (ANSI S3.6-1996, R1973)* (American National Standards Institute, New York).

Arnold, D. J., Lonsbury-Martin, B. L., and Martin, G. K. (**1999**). "High-frequency hearing influences lower-frequency distortion-product otoacoustic emissions," Arch. Otolaryngol. Head Neck Surg. **125**, 215–222.

Attias, J., Weisz, G., Almog, S., Shahar, A., Wiener, M., Joachims, Z., Netzer, A., Ising, H., Rebentisch, E., and Guenther, T. (**1994**). "Oral magnesium intake reduces permanent hearing loss induced by noise exposure," Am. J. Otolaryngol. **15**, 26–32.

Avan, P., Elbez, M., and Bonfils, P. (**1997**). "Click-evoked otoacoustic emissions and the influence of high-frequency hearing losses in humans," J. Acoust. Soc. Am. **101**, 2771–2777.

Backus, B. C., and Guinan, J. J. (**2007**). "Measurement of the distribution of medial olivocochlear acoustic reflex strengths across normal-hearing individuals via otoacoustic emissions," J. Assoc. Res. Otolaryngol. **8**, 484–496.

Beattie, R. C. (**2003**). "Distortion product otoacoustic emissions: Comparison of sequential versus simultaneous presentation of primary-tone pairs," J. Am. Acad. Audiol **14**, 471–484.

Beattie, R. C., and Bleech, J. (**2000**). "Effects of sample size on the reliability of noise floor and DPOAE," Br. J. Audiol. **34**, 305–309.

Beattie, R. C., Kenworthy, O. T., and Luna, C. A. (**2003**). "Immediate and short-term reliability of distortion-product otoacoustic emissions," Int. J. Audiol. **42**, 348–354.

Berger, E. H. (**2000**). "Hearing protection devices," in *The Noise Manual*, 5th ed., edited by E. H. Berger, L. H. Royster, J. D. Royster, D. P. Driscoll, and M. Layne (American Industrial Hygiene Association (AIHA) Press, Fairfax, VA).

Bohne, B. A., Yohman, L., and Gruner, M. M. (**1987**). "Cochlear damage following interrupted exposure to high-frequency noise," Hear. Res. **29**, 251–264.

Bray, P., and Kemp, D. T. (**1987**). "An advanced cochlear echo technique suitable for infant screening," Br. J. Audiol. **21**, 191–204.

Bray, P. J. (**1989**). "Click evoked otoacoustic emissions and the development of a clinical otoacoustic hearing test instrument," Ph.D. thesis, Institute of Laryngology and Otology, University College and Middlesex School of Medicine, London.

Clark, W. W. (**1991**). "Noise exposure from leisure activities: A review," J. Acoust. Soc. Am. **90**, 175–181.

Davis, B., Qiu, W., and Hamernik, R. P. (**2004**). "The use of distortion product otoacoustic emissions in the estimation of hearing and sensory cell loss in noise-damaged cochleas," Hear. Res. **187**, 12–24.

Davis, B., Qiu, W., and Hamernik, R. P. (**2005**). "Sensitivity of distortion product otoacoustic emissions in noise-exposed chinchillas," J. Am. Acad. Audiol **16**, 69–78.

Department of Defense (**2004**). Department of Defense Instruction 6055.12: DOD Hearing Conservation Program (HCP) (USD/AT&L).

Dhar, S., and Shaffer, L. A. (**2004**). "Effects of a suppressor tone on distortion product otoacoustic emissions fine structure: Why a universal suppressor level is not a practical solution to obtaining single-generator DP-grams," Ear Hear. **25**, 573–585.

Dorn, P. A., Piskorski, P., Gorga, M. P., Neely, S. T., and Keefe, D. H. (**1999**). "Predicting audiometric status from distortion product otoacoustic emissions using multivariate analyses," Ear Hear. **20**, 149–163.

Duvdevany, A., and Furst, M. (**2006**). "Immediate and long-term effect of rifle blast noise on transient-evoked otoacoustic emissions," J. Basic Clin. Physiol. Pharmacol. **17**, 173–185.

Duvdevany, A., and Furst, M. (**2007**). "The effect of longitudinal noise exposure on behavioral audiograms and transient-evoked otoacoustic emissions," Int. J. Audiol. **46**, 119–127.

Engdahl, B., Woxen, O., Arnesen, A. R., and Mair, I. W. (**1996**). "Transient evoked otoacoustic emissions as screening for hearing losses at the school for military training," Scand. Audiol. **25**, 71–78.

Federal Railroad Administration (**2006**). "Occupational noise exposure for railroad operating employees; final rule (49 CRF Parts 227 and 229)," Fed. Regist. **71**, 63066–63168.

Fleiss, J. L., Levin, B. A., and Paik, M. C. (**2003**). *Statistical Methods for Rates and Proportions* (Wiley, Hoboken, NJ).

Franklin, D. J., McCoy, M. J., Martin, G. K., and Lonsbury-Martin, B. L. (**1992**). "Test/retest reliability of distortion-product and transiently evoked otoacoustic emissions," Ear Hear. **13**, 417–429.

Ghiselli, E. E. (**1964**). *Theory of Psychological Measurement* (McGraw-Hill, New York).

Hamernik, R. P., Ahroon, W. A., and Lei, S. F. (**1996**) "The cubic distortion product otoacoustic emissions from the normal and noise-damaged chinchilla cochlea," J. Acoust. Soc. Am. **100**, 1003–1012.

Hamernik, R. P., Patterson, J. H., Turrentine, G. A., and Ahroon, W. A. (**1989**). "The quantitative relation between sensory cell loss and hearing thresholds," Hear. Res. **38**, 199–211.

Hamernik, R. P., and Qiu, W. (**2000**). "Correlations among evoked potential thresholds, distortion product otoacoustic emissions and hair cell loss following various noise exposures in the chinchilla," Hear. Res. **150**, 245–257.

Heitmann, J., Waldmann, B., Schnitzler, H., 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.
Humes, L. E., Joellenbeck, L. M., and Durch, J. S., eds. (**2005**). *Noise and Military Service: Implications for Hearing Loss and Tinnitus* (Institute of Medicine: Medical Follow-Up Agency, Washington, DC).
Job, A., Grateau, P., and Picard, J. (**1998**). "Intrinsic differences in hearing performances between ears revealed by the asymmetrical shooting posture in the army," Hear. Res. **122**, 119–124.
Kalluri, R., and Shera, C. A. (**2007**). "Near equivalence of human click-evoked and stimulus-frequency otoacoustic emissions," J. Acoust. Soc. Am. **121**, 2097–2110.
Karlsson, K. K., Berninger, E., and Alvan, G. (**1991**). "The effect of quinine on psychoacoustic tuning curves, stapedius reflexes and evoked otoacoustic emissions in healthy volunteers," Scand. Audiol. **20**, 83–90.
Kemp, D. (**2007**). "The basics, the science, and the future potential of otoacoustic emissions," in *Otoacoustic Emissions: Clinical Applications*, 3rd ed., edited by M. S. Robinette and T. J. Glattke (Thieme, New York), pp. 7–42.
Knight, R. D., and Kemp, D. T. (**1999**). "Relationships between DPOAE and TEOAE amplitude and phase characteristics," J. Acoust. Soc. Am. **106**, 1420–1435.
Konopka, W., Pawlaczyk-Luszczynska, M., Sliwinska-Kowalska, M., Grzanka, A., and Zalewski, P. (**2005**). "Effects of impulse noise on transiently evoked otoacoustic emission in soldiers," Int. J. Audiol. **44**, 3–7.
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.
Lapsley Miller, J. A., and Marshall, L. (**2001**). "Monitoring the effects of noise with otoacoustic emissions," Semin. Hear. **22**, 393–403.
Lapsley Miller, J. A., Marshall, L., and Heller, L. M. (**2004**). "A longitudinal study of changes in evoked otoacoustic emissions and pure-tone thresholds as measured in a hearing conservation program," Int. J. Audiol. **43**, 307–322.
Lapsley Miller, J. A., Marshall, L., Heller, L. M., and Hughes, L. M. (**2006**). "Low-level otoacoustic emissions may predict susceptibility to noise-induced hearing loss," J. Acoust. Soc. Am. **120**, 280–296.
LePage, E. L., Murray, N. M., Tran, K., and Harrap, M. J. (**1993**). "The ear as an acoustical generator: Otoacoustic emissions and their diagnostic potential," Acoust. Aust. **21**, 86–90.
Long, G., Talmadge, C., Prieve, B., and Lahtinen, L. (**2008**). "Extraction of DPOAE generator and reflection components in the time domain in adults and infants," Assoc. Res. Otolaryngol. Abstr. **31**, 61.
Long, G. R., and Talmadge, C. L. (**1997**). "Spontaneous otoacoustic emission frequency is modulated by heartbeat," J. Acoust. Soc. Am. **102**, 2831–2848.
Lonsbury-Martin, B., and Martin, G. K. (**2007**). "Distortion product otoacoustic emissions in populations with normal hearing sensitivity," in *Otoacoustic Emissions: Clinical Applications*, 3rd ed., edited by M. S. Robinette and T. J. Glattke (Thieme, New York), pp. 107–130.
Maison, S. F., and Liberman, M. C. (**2000**). "Predicting vulnerability to acoustic injury with a noninvasive assay of olivocochlear reflex strength," J. Neurosci. **20**, 4701–4707.
Marshall, L., Brandt, J. F., and Marston, L. E. (**1975**). "Anticipatory middle-ear reflex activity from noisy toys," J. Speech Hear Disord. **40**, 320–326.
Marshall, L., and Hanna, T. E. (**1989**). "Evaluation of stopping rules for audiological ascending test procedures using computer simulations," J. Speech Hear. Res. **32**, 265–273.
Marshall, L., and Heller, L. M. (**1998**). "Transient-evoked otoacoustic emissions as a measure of noise-induced threshold shift," J. Speech Lang. Hear. Res. **41**, 1319–1334.
Marshall, L., Lapsley Miller, J. A., and Heller, L. M. (**2001**). "Distortion-product otoacoustic emissions as a screening tool for noise-induced hearing loss," Noise Health **3**, 43–60.
Mining Safety and Health Administration (**1999**). "Health standards for occupational noise exposure; final rule (30 CFR Parts 56 and 57)," Fed. Regist. **64**, 49548–49634.
Murray, N. M., and LePage, E. L. (**2002**). "A nine-year longitudinal study of the hearing of orchestral musicians," paper presented at the International Auditory Congress, Melbourne, Australia, March.
Murray, N. M., LePage, E. L., and Mikl, N. (**1998**). "Inner ear damage in an opera theatre orchestra as detected by otoacoustic emissions, pure tone audiometry and sound levels," Aust. J. Audiol. **20**, 67–78.
Nageris, B. I., Raveh, E., Zilberberg, M., and Attias, J. (**2007**). "Asymmetry in noise-induced hearing loss: Relevance of acoustic reflex and left or right handedness," Otol. Neurotol. **28**, 434–437.
National Institute for Occupational Safety and Health (**1998**). *Criteria for a Recommended Standard: Occupational Noise Exposure (Revised Criteria 1998). No. 98–126* (U.S. Department of Health and Human Services (NIOSH), Cincinnati, OH).
Navy Occupational Health and Safety Program (**1999**). *OPNAVINST 5100.23E: Hearing Conservation and Noise Abatement* (Chief of Naval Operations, Washington, DC).
Nordmann, A. S., Bohne, B. A., and Harding, G. W. (**2000**). "Histopathological differences between temporary and permanent threshold shift," Hear. Res. **139**, 13–30.
Occupational Safety & Health Administration (**2007**). "Occupational noise exposure standard (29 CFR 1910.95)," *Code of Federal Regulations* (U.S. Department of Labor, Washington, DC).
Perez, R., Freeman, S., and Sohmer, H. (**2004**). "Effect of an initial noise induced hearing loss on subsequent noise induced hearing loss," Hear. Res. **192**, 101–106.
Price, G. R. (**2007**). "Validation of the auditory hazard assessment algorithm for the human with impulse noise data," J. Acoust. Soc. Am. **122**, 2786–2802.
Saunders, J. C., Dear, S. P., and Schneider, M. E. (**1985**). "The anatomical consequences of acoustic injury: A review and tutorial," J. Acoust. Soc. Am. **78**, 833–860.
Schmuziger, N., Probst, R., and Smurzynski, J. (**2005**). "Otoacoustic emissions and extended high-frequency hearing sensitivity in young adults," Int. J. Audiol. **44**, 24–30.
Seixas, N. S., Goldman, B., Sheppard, L., Neitzel, R., Norton, S. J., and Kujawa, S. G. (**2005a**). "Prospective noise induced changes to hearing among construction industry apprentices," Occup. Environ. Med. **62**, 309–317.
Seixas, N. S., Neitzel, R., Brower, S., Goldman, B., Somers, S., Sheppard, L., Kujawa, S. G., and Norton, S. (**2005b**). "Noise-related changes in hearing: A prospective study among construction workers," paper presented at the 30th Annual NHCA National Hearing Conservation Conference, Tucson, AZ, 26 February.
Shaffer, L. A., and Dhar, S. (**2006**). "DPOAE component estimates and their relationship to hearing thresholds," J. Am. Acad. Audiol **17**, 279–292.
Shaffer, L. A., Withnell, R. H., Dhar, S., Lilly, D. J., Goodman, S. S., and Harmon, K. M. (**2003**). "Sources and mechanisms of DPOAE generation: Implications for the prediction of auditory sensitivity," Ear Hear. **24**, 367–379.
Shaw, E. A. (**1966**). "Earcanal pressure generated by circumaural and supraaural earphones," J. Acoust. Soc. Am. **39**, 471–479.
Shera, C. A. (**2004**). "Mechanisms of mammalian otoacoustic emission and their implications for the clinical utility of otoacoustic emissions," Ear Hear. **25**, 86–97.
Shera, C. A., and Guinan, J. J. (**1999**). "Evoked otoacoustic emissions arise by two fundamentally different mechanisms: A taxonomy for mammalian OAEs," J. Acoust. Soc. Am. **105**, 782–798.
Siegel, S. (**1956**). *Nonparametric Statistics for the Behavioral Sciences* (McGraw-Hill, New York).
Sininger, Y. S., and Cone-Wesson, B. (**2004**). "Asymmetric cochlear processing mimics hemispheric specialization," Science **305**, 1581.
Sisto, R., Moleti, A., and Shera, C. A. (**2007**). "Cochlear reflectivity in transmission-line models and otoacoustic emission characteristic time delays," J. Acoust. Soc. Am. **122**, 3554–3561.
Slepecky, N. (**1986**). "Overview of mechanical damage to the inner ear: Noise as a tool to probe cochlear function," Hear. Res. **22**, 307–321.
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.
US Army Center for Health Promotion and Preventive Medicine (**2008**). "Noise levels of common army equipment," retrieved from http://usachppm.apgea.army.mil/hcp/noiselevels.aspx on 1 May 2008 (Aberdeen Proving Ground, MD).
Van Eyken, E., Van Laer, L., Fransen, E., Topsakal, V., Hendrickx, J. J., Demeester, K., Van de Heyning, P., Maki-Torkko, E., Hannula, S., Sorri, M., Jensen, M., Parving, A., Bille, M., Baur, M., Pfister, M., Bonaconsa, A., Mazzoli, M., Orzan, E., Espeso, A., Stephens, D., Verbruggen, K., Huyghe, J., Dhooge, I., Huygen, P., Kremer, H., Cremers, C., Kunst, S., Manninen, M., Pyykko, I., Rajkowska, E., Pawelczyk, M., Sliwinska-Kowalska, M., Steffens, M., Wienker, T., and Van Camp, G. (**2007**). "The

contribution of GJB2 (Connexin 26) 35delG to age-related hearing impairment and noise-induced hearing loss," Otol. Neurotol. **28**, 970–975.
Wagner, W., Heppelmann, G., Vonthein, R., and Zenner, H. P. (**2008**). "Test-retest repeatability of distortion product otoacoustic emissions," Ear Hear. **29**, 378–391.
Withnell, R. H., Hazlewood, C., and Knowlton, A. (**2008**). "Reconciling the origin of the transient evoked otoacoustic emission in humans," J. Acoust. Soc. Am. **123**, 212–221.
Withnell, R. H., and Lodde, J. (**2006**). "In search of basal distortion product generators," J. Acoust. Soc. Am. **120**, 2116–2123.
Withnell, R. H., Yates, G. K., and Kirk, D. L. (**2000**). "Changes to low-frequency components of the TEOAE following acoustic trauma to the base of the cochlea," Hear. Res. **139**, 1–12.
Yates, G. K., and Withnell, R. H. (**1999**). "The role of intermodulation distortion in transient-evoked otoacoustic emissions," Hear. Res. **136**, 49–64.
Zhou, X.-H., Obuchowski, N. A., and McClish, D. K. (**2002**). *Statistical Methods in Diagnostic Medicine* (Wiley-Interscience, New York).

High-frequency click-evoked otoacoustic emissions and behavioral thresholds in humans[a)]

Shawn S. Goodman,[b)] Denis F. Fitzpatrick, John C. Ellison, Walt Jesteadt, and Douglas H. Keefe
Boys Town National Research Hospital, Omaha, Nebraska 68131

(Received 25 June 2008; revised 2 December 2008; accepted 4 December 2008)

Relationships between click-evoked otoacoustic emissions (CEOAEs) and behavioral thresholds have not been explored above 5 kHz due to limitations in CEOAE measurement procedures. New techniques were used to measure behavioral thresholds and CEOAEs up to 16 kHz. A long cylindrical tube of 8 mm diameter, serving as a reflectionless termination, was used to calibrate audiometric stimuli and design a wideband CEOAE stimulus. A second click was presented 15 dB above a probe click level that varied over a 44 dB range, and a nonlinear residual procedure extracted a CEOAE from these click responses. In some subjects (age 14–29 years) with normal hearing up to 8 kHz, CEOAE spectral energy and latency were measured up to 16 kHz. Audiometric thresholds were measured using an adaptive yes-no procedure. Comparison of CEOAE and behavioral thresholds suggested a clinical potential of using CEOAEs to screen for high-frequency hearing loss. CEOAE latencies determined from the peak of averaged, filtered temporal envelopes decreased to 1 ms with increasing frequency up to 16 kHz. Individual CEOAE envelopes included both compressively growing longer-delay components consistent with a coherent-reflection source and linearly or expansively growing shorter-delay components consistent with a distortion source. Envelope delays of both components were approximately invariant with level.
 [DOI: 10.1121/1.3056566]

PACS number(s): 43.64.Jb, 43.66.Yw, 43.64.Kc [BLM] Pages: 1014–1032

I. INTRODUCTION

In response to a short-duration sound presented in the ear canal, a click-evoked otoacoustic emission (CEOAE) is generated within the cochlea and transmitted back through the middle ear into the ear canal, where it is detected using a miniature microphone (Kemp, 1978). The fact that a normal-functioning cochlea produces greater CEOAE signal energy than an impaired cochlea has led to the use of CEOAE testing to identify ears with a sensorineural hearing loss. Common techniques of measuring CEOAEs reviewed below share the property that the upper frequency of the CEOAE spectral response is limited to approximately 5 kHz. This also serves as the upper frequency of hearing for which a hearing loss can be identified using a CEOAE test. This report presents results showing a new form of CEOAE test, which can be used to measure CEOAE spectral energy and CEOAE latency up to 16 kHz in subjects with normal high-frequency hearing. The problem of measuring high-frequency CEOAEs is intimately related to the problem of assessing behavioral thresholds at high frequencies. A new incident-pressure technique to measure high-frequency audiometric thresholds is described that avoids effects of standing waves within the coupler used to calibrate the threshold sound pressure level (SPL).

This section reviews calibration issues related to the assessment of high-frequency hearing, followed by an overview of the procedure used to measure high-frequency CEOAEs. Methodological issues related to the measurement of behavioral thresholds and high-frequency calibration of the probe used to measure CEOAEs are next discussed. The experimental results of measurements of high-frequency CEOAEs are presented, which demonstrate the ability to measure level and latency responses in some ears up to 16 kHz. Potential applications of such a CEOAE test are described to highlight its possible utilization in audiological screening and diagnostic tests and as a probe of cochlear mechanics at high frequencies in human ears.

A. Assessment of high-frequency hearing

While normal-hearing humans can detect sound energy at frequencies up to 20 kHz, clinical hearing assessment, both behavioral and physiological, typically examines hearing only up to 8 kHz. One reason is that the majority of speech signals are present at frequencies of 8 kHz and below. Therefore, while the ability to hear at higher frequencies contributes to listening to music or other nonspeech auditory signals, sensitivity at the higher frequencies is usually of lesser concern. Another reason may be that responses at frequencies above 8 kHz are more difficult to measure accurately due to the short wavelengths involved and the increased difficulties of audiometry at high frequencies.

Nevertheless, there are situations where audiometric threshold measurements at frequencies above 8 kHz may be useful. Ototoxic damage to the cochlea typically occurs at

[a)]Portions of this work were presented at the 29th Annual MidWinter Research Meeting of the Association for Research in Otolaryngology, Baltimore, MD, February 2006.

[b)]Present address: Department of Speech Pathology and Audiology, University of Iowa, Iowa City, IA 52242. Electronic mail: shawn-goodman@uiowa.edu

the basal end of the cochlea and proceeds apically (Brummett, 1980; Konishi *et al.*, 1983; Komune *et al.*, 1981; Nakai *et al.*, 1982; Schweitzer *et al.*, 1984). Tests for detecting ototoxic hearing loss are therefore most sensitive at high frequencies, whether tested behaviorally (Dreschler *et al.*, 1989; Fausti *et al.*, 1999, 2003; Ress *et al.*, 1999; Tange *et al.*, 1985; van der Hulst *et al.* 1988) or using otoacoustic emissions (OAEs) (Mulheran and Degg, 1997; Ress *et al.*, 1999; Stavroulaki *et al.*, 2001; Stavroulaki *et al.*, 2002). The measurement of behavioral thresholds and OAEs above 8 kHz may also be useful for detection of and/or monitoring noise induced hearing loss (Kuronen *et al.*, 2003), presbycusis (Lee *et al.*, 2005; Matthews *et al.*, 1997), high-frequency loss associated with otitis media (Margolis *et al.*, 2000; Hunter *et al.*, 1996), and possibly loss associated with other etiologies. The ability to noninvasively assess physiological correlates to hearing over the full bandwidth of hearing would also help improve our understanding of cochlear and middle-ear mechanics in human ears at high frequencies.

One of the main difficulties with high-frequency threshold testing is calibration error arising from acoustic standing waves. Measurements SPL, whether made in the ear canal or a mechanical coupler, have pressure minima at the microphone relative to the eardrum or terminating wall of the coupler. These minima are present near frequencies having quarter wavelengths equal to the distance between the microphone and the termination (to the extent that the acoustic volume velocity is approximately zero at the termination). At yet higher frequencies, alternating maxima and minima of SPL are present in the ear canal. The difference in SPL between the microphone and eardrum can be as large as 20 dB (Stinson *et al.*, 1982; Siegel, 1994). Such differences may result in underestimation of the pressure at the eardrum, so that calibrated output levels are too high.

Various methods have been proposed to overcome these difficulties at high frequencies. Threshold-measurement systems have been calibrated up to 16–20 kHz based on measured responses in a flat-plate coupler (Fausti *et al.*, 1979). High-frequency thresholds calibrated in this manner were elevated in subjects with a history of noise exposure compared to thresholds in young adults (Fausti *et al.*, 1981). To estimate SPL at the eardrum in individual ears, Stevens *et al.* (1987) used a sound delivery system to the ear canal through a lossy cylindrical tube of 60 cm length between the sound source and ear canal. Their calibration involved fitting a model to the ear-canal spectrum to remove its zeros, indicative of standing-wave minima in the ear canal, and thus produce a spectrum that varied slowly with frequency. Because evanescent modes in the ear canal produce the largest effects at the frequencies at which the single propagating mode has a minimum, this calibration may have been influenced by evanescent modes. Stevens *et al.* (1987) also showed a frequency-response spectrum of the sound delivery system coupled to a long reflectionless cylindrical tube, and this spectrum was without of any standing-wave effects up to 20 kHz. This system was used to measure thresholds up to 20 kHz (Green *et al.*, 1987) in terms of both the voltage level applied to the source transducer as well as an extrapolation of SPL to high frequencies based on the fitting model. Green *et al.* (1987) concluded that a substantial number of subjects would have thresholds determined at the higher frequencies that might be in error by 10 dB or more.

Stelmachowicz *et al.* (1988) reported measurements using the audiometer of Stevens *et al.* (1987) and found increased variability of thresholds using the extrapolated SPL approach than the input voltage-level approach. Thresholds measured using a system calibrated with supra-aural headphones on a flat-plate coupler were more reliable above 11 kHz than those calibrated with the audiometer of Stevens *et al.* (1987) (Stelmachowicz *et al.*, 1989b). These studies indicate the complexity of estimating the SPL at the eardrum at high frequencies. With the simpler goal of calibrating in a flat-plate coupler and not attempting to estimate SPL at the eardrum, normative threshold data in the 8–20 kHz range were more elevated in older subjects relative to thresholds for the youngest age group (10–19 years) (Stelmachowicz *et al.*, 1989a).

ANSI S3.6 (2004) (Specification for Audiometers) describes coupler calibration of headphones for testing up to 16 kHz. However, the standard describes only the use of circumaural headphones, which are not typically used in OAE testing; it contains no recommendation of a calibration technique to use with insert earphones, which are routinely used for audiometric and OAE measurements. One possible solution is to simply calibrate at frequencies below 2–3 kHz, determine the voltage level applied to the sound source that is required to generate a given SPL in an ear or coupler, and then use that same voltage to generate sound at higher frequencies. This constant-voltage method assumes that the acoustical output spectra of the transducers are relatively flat within the frequency range of interest. In practice, the output of many transducers rolls off at frequencies above 5–8 kHz.

High-frequency calibration errors may be reduced if an ear-canal simulator is used whose length matches the length of the ear canal of the subject being tested (Gilman and Dirks, 1986). Because of large variability in ear-canal lengths (Stinson and Lawton, 1989; Kruger and Rubin, 1987), subjects' ear-canal lengths would need to be measured individually using, for instance, an operating microscope (Zemplenyi *et al.*, 1985). A further problem is that the tympanic membrane lies at an oblique axis to that of the ear canal, so that the ear-canal length is not uniquely specified. Another calibration method is to attach a probe tube to the microphone and place it very close to the eardrum. The small radius and long length of such a probe tube attenuate high frequencies, which may limit its usefulness for recording low-level OAEs. Precise placement of the probe tube is required for this method. Chan and Geisler (1990) described an acoustic method, utilizing the presence of standing waves in the ear canal to locate the probe tip near the eardrum. Dreisbach and Siegel (2001) used an endoscope to aid in placing a probe tube very close to the eardrum. Depth of insertion was adjusted so that pressure minima in probe-tube responses were shifted above 20 kHz. While these methods have been successfully used in research settings, their practical use in the clinic may be limited because of increased test time.

B. Procedure to measure CEOAEs

CEOAEs are low-level sounds produced by the healthy cochlea in response to a brief acoustic stimulus (see Probst *et al.*, 1991 for a review). CEOAEs are widely used in newborn screening protocols and are also used to test young children and difficult-to-test patients. By virtue of their short duration, CEOAE stimuli are broadband and therefore can be used to measure a response over a wide range of frequencies. In principle, a very broad range could be tested, from 0 Hz up to $\frac{1}{2}$ the sample rate of digitized signals, but measurement limitations have constrained CEOAE studies to frequencies up to approximately 4 kHz (Probst *et al.*, 1991).

It should be noted that distortion-product OAEs have been measured in human ears up to 16 kHz (Dreisbach and Siegel, 2001; Dreisbach and Siegel, 2005) with adequate repeatability (Dreisbach *et al.*, 2006). Stimulus-frequency OAEs (SFOAEs) have also been measured up to 14 kHz (Dreisbach *et al.*, 1998). The ability to measure CEOAEs to high frequencies would complement the existing large literature relating to CEOAE measurements at lower frequencies, as well as studies based on measurements of high-frequency Distortion Product Otoacoustic Emissions (DPOAEs) and SFOAEs. CEOAEs have the advantage that a broad bandwidth of cochlear response is assessed in a single time-averaged response.

In practice, however, the CEOAE bandwidth may be limited by several factors. An electrical impulse is often used as the stimulus input to the sound source (earphone) producing the click used to evoke a CEOAE response. However, the earphones, which transduce the electrical signal into a sound wave, may not have a sufficiently flat magnitude transfer function extending into the high frequencies. As a result, the click energy at high frequencies may be less than that at lower frequencies, which contribute to the difficulty of detecting high-frequency CEOAEs.

Extraction of high-frequency CEOAEs from the stimulus and noise floor poses special challenges. CEOAEs recorded in human ears are typically on the order of 30 dB smaller than the stimuli that elicit them and may overlap the stimulus in both frequency and time. Low- and midfrequency (<5 kHz) CEOAEs can be extracted by taking advantage of the time it takes for the signal to travel from the eardrum to its characteristic places along the basilar membrane and for the emissions to travel back along the reverse pathway (Kalluri and Shera, 2007; Shera *et al.*, 2007). Because of the tonotopic organization of the cochlea, low-frequency signals travel further along the cochlea and are delayed relative to higher-frequency signals. Based on an assumed source equivalence of CEOAEs and SFOAEs, the expected delays for the 1, 2, 4, 8, and 16 kHz components of CEOAEs are approximately 11, 7.1, 4.6, 3.0, and 1.9 ms (Shera *et al.*, 2002). For one commonly used CEOAE system (Otodynamics ILO 88), the electrical pulse duration delivered to the loudspeaker is on the order of 80 μs, resulting in a biphasic response that decays over 2–3 ms due to the impulse response duration of the earphone and multiple internal reflections between probe and eardrum (Kemp *et al.*, 1990; Glattke and Robinette, 2007). From these durations it can be seen that at frequencies $\geqslant$8 kHz, the emission begins to overlap the acoustic stimulus in time. To avoid stimulus artifact in the CEOAE response, many analysis programs zero the first 2.5 ms of the response waveform and window the onset of the remaining response (Kemp *et al.*, 1990). Based on the expected delay times described above, this process can be expected to eliminate the CEOAE response above 6–8 kHz.

When the stimulus and CEOAE overlap in both time and frequency, they may still be separated by exploiting the nonlinear (i.e., compressive) amplitude growth of the emission. This method has been used in several forms (Kemp and Chum, 1980, Zwicker and Schloth, 1984; Brass and Kemp, 1991; Keefe, 1998; Keefe and Ling, 1998). This technique consists of presenting the stimulus at a relatively low level and recording the resulting ear canal pressure. A second stimulus differing in level, frequency, or both is then presented one or more times, and the ear-canal pressure is again recorded. By appropriate scaling and vector subtraction techniques, the linear portions of the stimulus and emission are canceled. The remaining nonlinear component is thought to consist mostly of the emission. Section IV describes the CEOAE extraction procedure, in which the initial part of the response was retained, thereby preserving the high-frequency content of the emission.

II. ADAPTIVE MAXIMUM-LIKELIHOOD PROCEDURE TO MEASURE THRESHOLD

An adaptive maximum-likelihood (ML) procedure that is based on a set of single-interval responses to a yes-no task was used to measure thresholds across the frequency range of hearing. Adaptive ML threshold estimation methods were developed by Green (1993) and generalized by Gu and Green (1994) to include catch trials to decrease the false-alarm rate. ML procedures provide a bias-free automatic technique to assess threshold that is more efficient than other commonly used bias-free threshold-measurement procedures.

The modified adaptive ML procedure used in the present research is described in more detail elsewhere (Keefe *et al.*, 2009). The modification is that the stimulus level in each of the first four trials was selected by randomly choosing without replacement one of these subranges for the trial and then randomly choosing a stimulus level within the subrange. An initial threshold estimate was calculated using the ML estimate based on these initial four yes-no responses. In accord with the Green (1993) procedure, the stimulus level of the fifth and subsequent trials was chosen based on this current threshold estimate. This modified procedure substantially reduced the sensitivity of the usual ML procedure to errors when testing human subjects.

Preliminary data were acquired with a number of trials N as large as 30, and $N=15$ trials were selected as a sufficiently large number of trials. A false-alarm rate was also estimated. The 15 trials per run included three catch trials in which no stimulus was presented. The order of these three catch trials was uniformly randomized within the last 11 trials of a run. Once N was set at a total of 15 trials, preliminary data were acquired using a greater and lesser number of catch trials before settling on three trials as adequate, corre-

sponding to a catch-trial rate of 20%. The responses to catch trials were included in the ML estimate of threshold and false-alarm rate by assuming that the stimulus level of the catch trial was the minimum stimulus level in the dynamic range.

Finally, if the subject answered *No* three times in succession, with the stimulus level set at its maximum value, the run was halted before $N=15$ trials were performed. This condition was not assessed until after the initial four trials. This condition gave sufficiently high confidence that the subject's threshold was above the maximum stimulus level used by the system.

III. STIMULUS DESIGN AND CALIBRATION AT HIGH FREQUENCIES

The design of an electrical stimulus to serve as the input to an earphone producing a short-duration acoustic stimulus (a "click") is complicated by the need for a wide stimulus bandwidth. Part of the complexity resides in the bandwidth limitations of existing earphones used in OAE probes, and another part has to do with how the sound output from the probe is calibrated at high frequencies.

Historically, insert earphones used in clinical hearing tests and hearing-research measurements in adult human subjects have been calibrated either in a 2 cm^3 coupler or an artificial ear. The 2 cm^3 coupler is a reference coupler for acoustic immittance measurements at low frequencies (i.e., at 226 Hz). When the task is to calibrate probe levels above 8 kHz, the 2 cm^3 coupler and the artificial ear have limitations. A significant problem in 2 cm^3 couplers conforming to standards is that the coupler is used with a reference microphone with a 1 in. diameter. This diameter is large compared to the diameter of the ear canal and large also compared to the acoustic wavelength at 16 kHz. The artificial ear was not designed for use at these high frequencies because the human impedance data on which its design is based do not encompass such high frequencies. An alternative and simpler coupler with acoustic properties appropriate for calibration across the audio frequency range is that of a long rigid-walled cylindrical tube, into which the source and microphone transducers are inserted in a leak-free manner.

Suppose that the sound source outputs a transient of a given duration D. This transient travels down the tube away from the source, is reflected at its far end, and travels back toward the source end to the measurement microphone with a round-trip travel time T_{rt}. If the measurement time is longer than D and less than T_{rt}, then the microphone response includes only the incident signal. There are no standing waves, so that the corresponding spectrum of the measured sound pressure has no maxima or minima associated with standing waves. This property has been used in studies measuring the ear-canal acoustic reflectance (Keefe, 1997; Keefe and Simmons, 2003). Such an anechoic, or reflectionless, coupler is well suited for probe calibration at high frequencies by eliminating the variability within the coupler due to standing waves.

A cylindrical brass tube of length 91.4 cm served as the anechoic termination; the tube was closed at its far end to prevent contamination from room noise. The tube had a circular cross-sectional diameter (8.02 mm), similar to the diameter of an adult human ear canal. A probe assembly, composed of a pair of insert earphones (ER-2, Etymotic) and microphone (ER-10B+, Etymotic), was inserted into one end of the tube. This probe assembly was also used in the CEOAE measurements. The round-trip travel time in this tube was approximately 5.3 ms. When windowed to include slightly less than the first 5.3 ms of the response, the microphone recorded the pressure response associated only with the outgoing, or incident, acoustic wave.

Stimuli were digitally generated and recorded at a sample rate $f_s=44\,100$ Hz (sample period $T=1/f_s$) using a computer, Digital Audio Labs 24 bit sound card (CardDeluxe), and custom software. The ER-10B+ microphone frequency response provided by the manufacturer was flat to within +1.8 dB up to 8 kHz with a nominal sensitivity level of −26 dB (re 1 V/Pa). Above 8 kHz, there was one maximum and one minimum in sensitivity relative to this nominal sensitivity.[1] The manufacturer's calibration of frequency-dependent sensitivity was used in this study for all spectral results. The microphone preamplifier was used with +20 dB gain, which boosted the effective sensitivity level to −6 dB (re 1 V/Pa). No phase calibration was provided for this microphone so that the pressure phase response above 8 kHz was approximated by the phase response of the voltage signal recorded by the analog-to-digital converter (ADC) of the sound card. Thus, the waveform plots reported in Fig. 3 of an output waveform from the microphone preamplifier and of a nonlinear residual CEOAE waveform, are voltage waveforms that have been scaled to pressure waveforms by the nominal sensitivity of −6 dB. Otherwise, all results reported herein, including CEOAE envelope delays, were independent of microphone phase sensitivity.

ADC voltage waveforms were time averaged and analyzed spectrally after applying the frequency-dependent microphone sensitivity. Although a departure from previous work on CEOAEs, the sound level of the CEOAE was calculated using the sound-exposure spectrum level (SEL spectrum), which is a standard means of calculating the sound level of a transient (ANSI S1.1, 1994). The Appendix describes the implementation of the SEL spectrum and its relation to the SPL spectrum. The SEL spectrum was calculated using the windowed time-averaged pressure waveform sequence $p[n]$ at sample n, which was zero padded out to $N=1024$ samples. Using this buffer length N, the discrete Fourier transform (DFT) of $p[n]$ was $P[k]$. The band SEL spectrum L_{Eb} in the kth frequency bin was calculated as follows, which is based on an averaging time NT equal to the DFT buffer length,

$$L_{Eb} = 10 \log_{10}\left(\frac{\frac{2}{N^2}|P[k]|^2}{P_{\text{ref}}^2}\right). \quad (1)$$

The click was designed using a method modified from Agullo *et al.* (1995). The microphone output voltage waveform was measured at discrete-time samples n in response to an electrical delta-function input from the digital-to-analog converter (DAC). This was the voltage impulse response

$h[n]$ of the measurement system. A finite impulse response (FIR) filter was designed using the Kaiser window technique to create a short-duration impulse response $g[n]$ for a signal with a passband from 0.5 to 16 kHz and stop bands at 0.043 and 17 kHz. An inverse filtering technique was used to find the electrical input $x[n]$ such that the voltage output response $g[n]$ of the microphone was a multiple of this filter shape, that is, to find $x[n]$ such that $g[n]=h[n]^*x[n]$, where * represents convolution. A MATLAB implementation of a conjugate gradient method (Hansen, 2001) was used to find a stable solution for the electrical input. The resulting acoustic click spectrum recorded in the anechoic tube was approximately flat from 0.5 out to 16 kHz. The stimulus waveform so devised did not include the frequency variation of the recording microphone (ER-10B+) so that the actual pressure spectrum was inversely proportional to the maximum and minimum in microphone sensitivity described in footnote 1. These variations in microphone sensitivity level were within +2 dB at all third octave frequencies up to 12.7 kHz, but the sensitivity level was reduced by 5.2 dB at 16 kHz compared to the microphone preamplifier output voltage spectrum recorded by the ADC.

A possible source of measurement error is the close proximity of the microphone and earphone tubes in the probe assembly, which can lead to evanescent-mode effects in some situations (Burkhard and Sachs, 1977; Rabinowitz, 1981; Huang *et al.*, 1998, 2000). The manufacturer provides the ability to modify the ER-10B+ probe by extending the sound tube further beyond the microphone tube in order to reduce evanescent effects. An extension of 5 mm is consistent with Burkhard and Sachs (1977) and is recommended by the manufacturer (Etymotic), but the 1/4 wavelength of a 16 kHz tone is only 5.4 mm. Thus, a 5 mm extension is longer than what would be desirable for high-frequency measurements in the ear canal and what would complicate the interpretation of pressures measured near the tympanic membrane. Features of the Etymotic ER-10B+ microphone frequency response have been described along with acoustic studies of its 5 mm extension tube (Siegel, 2007).

Probe responses were measured using lesser extensions of 4 and 2 mm, which are generally consistent with recommendations in Keefe and Benade (1980), as well as using a zero-length extension (0 mm) to examine contributions of nonplanar and evanescent modes. The click stimulus was recorded using the ER-10B+ probe assembly with no extension and with extensions (ER10B-3/ER7-14C, Etymotic) of 0, 2, and 4 mm. In addition to moving the earphone port further away from the microphone (in the 2 and 4 mm cases), the extension also reduced the cross-sectional area of the aperture. Adding an extension replaced the usual 255 of 1.35 mm inner-diameter tubing with a slightly shorter segment of inner diameter 1.35 mm coupled to a pair of short segments with inner diameters of 0.86 and 0.5 mm.

Spectral results are shown for recordings in the anechoic tube (Fig. 1, top) and a 2 cm^3 coupler (Fig. 1, bottom). The 2 cm^3 coupler used in all measurements was an HA-1 coupler as specified in ANSI 3.7 (1995). Each spectrum was calculated based on the waveform truncated to eliminate all reflected signals from the end of the tube and zero padded out to 1024 samples. The top panel of Fig. 1 shows that for the spectrum measured in the anechoic tube, the main effect of any extension was to attenuate the SEL spectrum compared to the no-extension condition. This was due to greater viscothermal dissipation within the reduced area of the plastic tube within the capillary. Aside from overall attenuation, the SEL spectrum varied little with extension length in the anechoic tube.

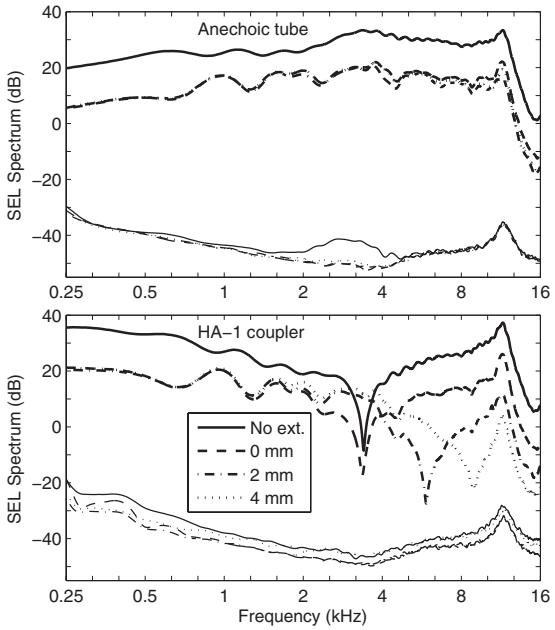


FIG. 1. Comparison of click spectra measured in the anechoic tube (top panel) and in a 2 cm^3 (HA-1) coupler (bottom panel). In each panel the thicker lines toward the top show the signal SEL spectrum, and the thinner lines near the bottom show the noise SEL spectrum. The noise SEL is the standard error of the mean and is extremely low due to the large number of averages (N=4050). The legend indicates the length of the ER-10B+ extension tube associated with each spectra, or No ext. for the no-extension condition, which is the subsequent default configuration of the ER-10B+ used in the study.

The bottom panel of Fig. 1 shows that for the spectrum measured in the 2 cm^3 coupler, there were level differences as large as 50 dB between the conditions. The SEL spectrum for the no-extension condition was higher than the SEL spectrum for all extension conditions at low frequencies up to 3 kHz, while the SEL spectral differences for the three extension conditions were within a few dB up to 2 kHz and much larger above 2 kHz. This large variability shows that the 2 cm^3 coupler should not be used with the ER-2/ER-10B+ probe assembly to assess stimulus level above 2 kHz. The diameter (18–21 mm) of the cylindrical cavity comprising the main volume of the 2 cm^3 coupler [HA-1 as described in ANSI S3.7 (1995)] is large compared to its length (5.4–7.3 mm), which enhances the acoustical effect of the evanescent modes. In addition to an overall attenuation of the SEL spectrum, large differences appear as the sound tube is extended, especially in the location of notch frequencies between 3 and 8 kHz. A comparison of the top and bottom panels in Fig. 1 shows that the spectra measured in the anechoic tube were much smoother across frequenies than those measured in the 2 cm^3 coupler. The anechoic tube removed standing-wave effects in calibration, making it more

accurate in calibrating at higher frequencies. When used in real-ear tests, the sound field in the ear canal includes this calibrated incident-pressure signal and the reflected pressure signal from the tympanic membrane. The complex sum of these pressures in the ear canal forms standing waves in the ear-canal pressure that vary with the measurement location of the microphone and the source reflectance of the probe. The present method provides a measure of the incident pressure delivered to the ear, as further described below.

Based on these responses, the ER-10B+ was used in subsequent CEOAE and audiometric measurements using the no-extension configuration. This provided a SEL spectrum that was approximately 13 dB larger than in any of the configurations with extensions. This increased sound level was useful in measuring CEOAE responses with limited signal-to-noise ratios (SNRs). Another benefit was that there was no need to adjust the plastic tubing to a particular extension distance, which would have been a source of variability between subjects.

The main difficulty with higher-order evanescent mode interactions in ear-canal or tube measurements arises from the fact that the total input impedance measured at the surface of the probe is the sum of the input impedance associated with the plane-wave acoustic excitation in the ear canal and the input impedance associated with the evanescent modes, which acts as an inertance (Keefe and Benade, 1980). That is, the plane-wave and evanescent-mode impedances act in series. A reactance that is inertance dominated increases linearly with frequency, so that the effects of evanescent modes grow in importance at high frequencies. Evanescent-mode effects become important when the inertive reactance is on the order of the magnitude of the plane-wave impedance. This occurs first near the minima of the input reactance, which occur at frequencies near pressure minima as predicted by models of plane acoustic waves in cylindrical ducts. These are the frequencies of the sharp minima in the 2 cm^3 coupler responses (see bottom panel of Fig. 1). At frequencies away from impedance minima, the plane-wave impedance dominates and the evanescent-mode effects are negligible. This means that there will be isolated narrow ranges of frequency in practical ear-canal measurements at high frequencies in which evanescent-mode effects can play a role. Because the middle ear is much more efficient at absorbing sound energy than are the hard walls of a coupler, the pressure minima are not as deep so that the effects of evanescent modes are smaller than in coupler measurements.

Restricting attention to the no-extension configuration of the ER-10B+ probe (including ER-2 earphones), the SEL spectrum in the 2 cm^3 coupler was measured both by the probe microphone and by the reference microphone, i.e., the 1 in. Bruel and Kjaer microphone (type 4144) used in the HA-1 coupler specified in ANSI S3.7 (1995). The reference microphone was used with a sound level meter (Bruel and Kjaer type 2231). Responses were measured using both the click stimulus used in the CEOAE procedure and using tone bursts that were similar to those used in the ML threshold procedure, but of longer duration to provide time to read the visual display of the sound level meter. All measurements using the visual display of the sound level meter were an average of three measurements. The individual calibration provided by Bruel and Kjaer of the sensitivity of the reference microphone was used in all spectral analyses in this study, including frequency-dependent variations in sensitivity important at higher frequencies up to 16 kHz.

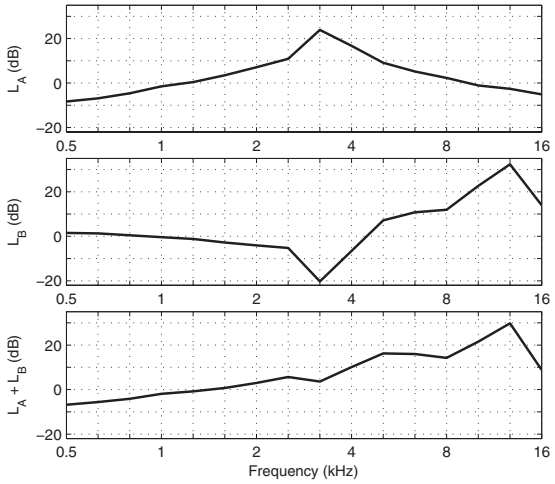


FIG. 2. 1/3-octave averaged transfer-function levels are plotted as follows. Top: L_A for the ER-10B+ in the anechoic tube re its level in the 2 cm^3 coupler. Middle: L_B for the ER-10B+ in the 2 cm^3 coupler re the level of a 1 in. Bruel and Kjaer microphone in the 2 cm^3 coupler. Bottom: L_A+L_B. Frequency-dependent sensitivities have been applied to each microphone response in each panel.

Three relevant transfer functions are defined and plotted in Fig. 2. The top panel shows the transfer-function level L_A between the probe-microphone response in the anechoic tube relative to the probe-microphone response in the 2 cm^3 coupler. This transfer function shows the effect of acoustic termination on probe response. L_A was measured as the difference in the third-octave averaged sound level of the click responses. The click stimulus was the appropriate choice because it has a sufficiently short duration that the tube acts as an anechoic termination. This would not be the case for the tone-burst stimulus. L_A in the top panel of Fig. 2 was negative below 1.2 kHz, which means that the sound level measured by the ER-10B+ was less in the anechoic tube than in the compliance-dominated 2 cm^3 coupler at low frequencies. The maximum in L_A near 3.2 kHz was created by a minimum in the sound level in the 2 cm^3 coupler (bottom panel of Fig. 1, no-extension condition).

The middle panel shows the transfer-function level L_B measured in the 2 cm^3 coupler by the probe microphone relative to that measured by the reference microphone. The tone-burst stimulus was the appropriate choice because earphones are calibrated using the SPL measured by the reference microphone according to ANSI S3.6 (2004). The probe-microphone response was recorded in the 2 cm^3 coupler by the computer measurement system and converted to SPL using the peak-to-peak pressure waveform difference during the steady-state portion of each frequency-specific tone burst. L_B was measured as the difference in these SPLs to represent the variability in SPL within the 2 cm^3 coupler between the two measurement locations. L_B in the middle panel of Fig. 2 approached 0 dB at the lowest measurement frequencies to within measurement precision. The absolute value of L_B as-

sessed level differences within the 2 cm^3 coupler and did not exceed 5.3 dB at any frequency up to 2.5 kHz. However, strong spatial effects were present within the coupler above 2.5 kHz. A zero crossing in L_B occurred near 4.5 kHz, and L_B was as large as 33 dB at 12.7 kHz. Acoustic calibrations of probe microphones using the HA-1 coupler and its reference microphone strongly depend on spatial variations in sound field above 2.5 kHz.

The sum of these transfer-function levels, L_A+L_B, describes the relationship between sound level measured by the ER-10B+ microphone in the anechoic tube relative to the sound level measured in the HA-1 coupler by the 1 in. Bruel and Kjaer microphone (bottom panel, Fig. 2). The L_A+L_B increased with increasing frequency up to 30 dB at 12.7 kHz and decreased to 9 dB at 16 kHz. Because these are transfer functions, they can be used to adjust either SEL or SPL spectra.

These transfer-function measurements do not directly calibrate the probe microphone with respect to a reference microphone at high frequencies, inasmuch as a reference microphone was not incorporated into the anechoic coupler. One such approach is described in Siegel (2007). The present system with its anechoic probe termination provides an incident-pressure stimulus for use in calibrating behavioral thresholds. A power-based system for calibrating thresholds would require a measurement of the power absorbed by the middle ear (e.g., Keefe *et al.*, (1993), but this method would require wideband aural acoustic admittance or related transfer-function measurements and would still not quantify any power internally lost within the middle ear that is not absorbed by the cochlea. The incident-pressure calibration is simpler in that it requires only sound level measurements of the incident signal, yet it provides a high-frequency calibration that is not influenced by standing waves in the ear canal. An incident-pressure calibration in a long tube is also applicable to time-gated tonal or noise signals, as long as their duration is less than the round-trip travel time within the tube.

IV. METHODS

A. Subjects

Responses included in the analyses were obtained from 49 ears (24 left and 25 right ears) of 29 normal-hearing subjects (25 females and 4 males), in whom the mean age ± 1 standard deviation (SD) was 20.5 ± 3.4 years over an age range of 14–29 years. All subjects had pure-tone air-conduction thresholds <15 dB HL at octave frequencies from 0.5–8 and 226 Hz tympanograms within normal limits. During testing, subjects were seated comfortably inside a sound-attenuated booth. The experimental protocol was approved by the Institutional Review Board at Boys Town National Research Hospital, and written informed consent was obtained from all participants.

B. Behavioral threshold-measurement procedures

Behavioral hearing thresholds were measured using the adaptive ML procedure in a yes-no task at octave frequencies from 0.5 to 4 kHz and at 11 third-octave frequencies from 4 to 16 kHz. Each tone-burst stimulus had a total duration of 250 ms, which included 25 ms onset and offset ramps using cosine-squared envelopes. The threshold procedure was automated, and subjects provided responses using a custom-built response box that included text feedback via a visual display. The feedback indicated whether a yes or no response had been recorded on the previous trial and alerted the subject to the upcoming trial. The system waited after each stimulus presentation (gated tone or silence) until the subject depressed the Yes or No button.

The ML threshold was determined first at 0.5 kHz and then at other frequencies in ascending order. One reason for testing in this order is that all subjects were presented with stimuli in their suprathreshold range at the beginning of the test and finished with the frequencies above 8 kHz for which a tone might be inaudible. Any effect of test order was outside the scope of this study.

Two runs of the adaptive yes-no threshold procedure were performed, each run composed of 15 trials with four initial trials in which levels were set as described above in a manner that was independent of the subject responses. Three catch trials were included in the last 11 trials of each run. If the SD in the threshold between the two runs did not exceed 3 dB, the mean threshold was saved as the threshold estimate, and the test moved to the next frequency. Otherwise, a new run was performed, and the mean and SD of the threshold were again calculated. The mean threshold was stored whenever the 3 dB criterion was attained. If five runs were performed at this frequency without reaching the criterion, data collection was halted at this frequency, the mean frequency was saved as the threshold estimate across the pair of successive runs with the lowest SD, and the test proceeded to the next frequency.

Initial data collection at 0.5 kHz was considered a training run. The automated threshold test was paused when the criterion at 0.5 kHz was attained or after five runs. The operator then provided verbal feedback to the subject that the criterion had been attained, and, if so, data collection continued under the subject's control at the next test frequency. If the criterion was not obtained after five runs, the operator so informed the subject, repeated the instructions on the use of the response box, and asked if tones were audible. Then, the threshold was again measured at 0.5 kHz (after discarding the old data), and this feedback step was repeated until the subject was either trained to criterion performance or excluded from the study.

The ML thresholds measured using the ER-2 earphones joined to the ER-10B+ with no coupling extension tube (see "No ext." curve in top panel of Fig. 1) were compared to clinical thresholds (GSI-33 audiometer) measured at octave frequencies up to 8 kHz. The clinical thresholds were each calibrated in HL according to ANSI S3.6 (2004) using the 1 in. microphone in the 2 cm^3 coupler. Thus, behavioral threshold and CEOAEs were measured with the same probe. Clinical thresholds at each frequency took approximately 0.5 min to acquire, and ML thresholds for two runs took approximately 1.5 min.

C. CEOAE measurement paradigm

CEOAEs were collected using a nonlinear residual method (Keefe, 1998) based on three responses (sample duration of 1124 samples or 25.5 ms), each elicited using a different stimulus. The first stimulus (s_1) was presented through a first earphone, followed by a presentation of the second stimulus (s_2) though a second earphone. Then both stimuli were presented simultaneously ($s_{1,2}$), each through its own earphone. The ear-canal sound pressure p_1 was measured in response to stimulus s_1, p_2 to stimulus s_2, and $p_{1,2}$ to stimulus $s_{1,2}$. The nonlinear residual, p_d, was extracted by calculating $p_d = p_1 + p_2 - p_{1,2}$. By this process, the linear-system responses to the stimuli were canceled along with any isochannel system distortion, leaving only the nonlinear residual, which was interpreted as a biological response, i.e., as the OAE, in the absence of any system intermodulation distortion. Such system distortion was assessed from coupler recordings in an artificial ear (Bruel and Kjaer type 4157) that approximated the impedance of an average adult human ear (IEC 711 standard). This distortion level never exceeded the measured noise level.

The click stimulus described above was presented as s_1, the "probe click." The same click presented at a level of 15 dB above s_1 was used as s_2, the "second click." Based on SF OAE studies (Dreisbach *et al.*, 1998; Shera and Guinan, Jr., 1999; Schairer *et al.*, 2003), a second signal with level 15 dB above the probe level is sufficient to substantially fully recover the SFOAE. Kalluri and Shera (2007) considered compression and suppression methods of extracting SFOAEs, and this CEOAE method likely involves both mechanisms, as well as a distortion mechanism (Withnell *et al.*, 2008) further described below. Measurements in human ears (Konrad-Martin and Keefe, 2005; Kalluri and Shera, 2007) provide evidence over limited ranges of frequencies and moderate levels that SFOAEs and CEOAEs are generated by the same underlying cochlear mechanism, as predicted by the coherent reflection emission theory (Shera and Guinan, Jr., 1999). This suggests that a CEOAE response may be interpreted over these limited ranges as a superposition of SFOAE responses at frequencies within the passband of the CEOAE (this interpretation is revisited in Sec. VI). This source equivalence between SFOAE and CEOAE responses in human ears suggested the use of the simultaneous presentation of clicks differing in level by 15 dB. Previous CEOAE measurements based on the nonlinear residual technique of the present study used equal levels for the click (and chirp) stimuli s_1 and s_2 (Keefe and Ling, 1998), which resulted in slightly lower SNR levels than in the present study because of a less complete extraction of the CEOAE.

The overall duration of each recording buffer was 1124×3 samples, or 76.5 ms. This is slightly less than the typical buffer duration (80 ms) used in clinical CEOAE tests, which is comprised of three elementary buffers of a repeated click and a fourth elementary buffer of a click at three times the amplitude and opposite polarity (Kemp *et al.*, 1990). The levels of the click stimuli were calculated in peak equivalent SPL (peSPL), which equals the SPL of a continuous tone having the same peak-to-peak pressure amplitude as the maximum peak-to-peak pressure amplitude of the click waveform.[2] Probe clicks were presented in 6 dB steps at levels from 43 to 73 dB peSPL. Levels were always presented in order from highest to lowest level. For each level, $M = 4050$ independent buffers were collected (duration of 5.2 min). The CEOAE signal and noise sound levels were calculated using the coherent and incoherent averaging procedures described in the Appendix of Schairer *et al.* (2003) and expressed as SEL spectra (see Appendix)

The CEOAE test time for each subject was approximately 1 h. During CEOAE testing within the booth, subjects typically viewed a television program using a closed-caption DVD (i.e., without audible sound), which helped them remain awake and alert. To reduce variability associated with probe placement, the probe was not removed from the ear during the test session unless it inadvertently needed refitting.

D. CEOAE *post hoc* analysis

Data were analyzed using custom MATLAB software. Artifact rejection was used to detect and reject data buffers contaminated by intermittent noise that were outliers in the sense of Hoaglin *et al.* (1983). The data were first filtered using a high-pass FIR filter (0.354 kHz cutoff, 0.5 kHz passband lower frequency, 5.7 ms group delay) to remove low-frequency noise below the analysis band, and the total energy in each buffer was calculated. Individual responses with energy >2.25 times the interquartile range (IQR) of the buffer set were discarded. This method of *post hoc* artifact rejection resulted in an average of $N = 3690$ buffers retained for each recording; e.g., 8.7% of buffers were rejected on average across subjects with a SD of 4%.

In order to elicit the best possible high-frequency emissions, relatively high click levels were used. The combination of higher click levels and a nonlinear residual paradigm results in the possibility of middle-ear muscle (MEM) reflex activation, which could be mistaken for OAEs. If present, the MEM reflex would tend to affect the nonlinear residual response at lower frequencies. Further, many subjects had synchronous spontaneous otoacoustic emissions (SSOAEs) present below 2 kHz. It was therefore decided to limit the CEOAE spectral analysis to frequencies ≥ 2 kHz. This decision was consistent with the fact that CEOAE spectra below 2 kHz are well understood. However, as described later, CEOAE latency measurements were extended downward to 1 kHz.

A time-domain window encompassing the higher frequencies in the CEOAE waveform was devised for the CEOAE spectral analyses as follows: The window onset was positioned at the centroid of the click stimulus; this onset occurred at $t = 0$ ms in the click stimulus response plotted in Fig. 3 (top panel). Based on data from Shera *et al.* (2002), the expected group delay for a 2 kHz SFOAE is 7 ms, with an upper 95% confidence interval of 8.3 ms. The duration of the window was therefore chosen to be 8.3 ms. The effects of this window onset and duration are evident in the CEOAE waveform plotted as a black line in the bottom panel of Fig. 3. The window was gated on and off using half cosine-

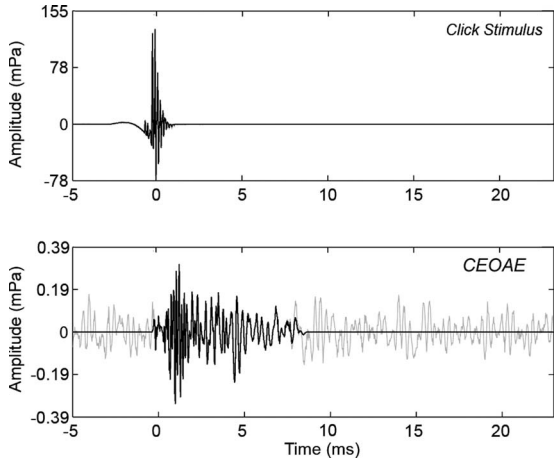


FIG. 3. Waveforms of time-averaged click stimulus (top) and CEOAE response (bottom). Zero on the time axis is set to the centroid of the click. The full CEOAE response is shown in gray, which is the waveform of the nonlinear residual. The windowed portion of the CEOAE extends from 0 to 8.3 ms and is expected to contain energy from 2 to 16 kHz (shown in black). The windowed portion was used in subsequent analyses in this paper.

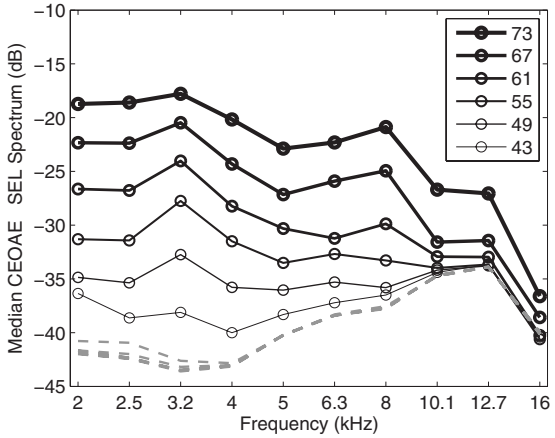


FIG. 4. Median CEOAE SEL spectra (1/3 octave average) for each of six click stimulus levels are plotted with increasing line thickness for increasing stimulus level. The peSPL of each click stimulus is specified in the legend. The median noise SEL spectra (1/3 octave average) are also plotted (dashed lines). The line thickness is the same for the noise levels because they had no level dependence.

squared windows of duration equal to three periods of the expected CEOAE frequency. The dominant frequency at the onset was expected to be on the order of 16 kHz (period of 0.0625 ms), so the window was gated on over a duration of 0.187 ms. The dominant frequency at the offset was expected to be on the order of 2 kHz (period of 0.5 ms), so its window was gated off over a duration of 1.5 ms. This windowing had the added benefit of reducing the noise floor by eliminating the noise in the rest of the CEOAE response (Whitehead *et al.*, 1995a), which occurred primarily at frequencies below 2 kHz. The original CEOAE waveform calculated as the nonlinear residual is plotted as the gray line in the lower panel in Fig. 3. The windowed CEOAE waveform (black line in the lower panel in Fig. 3) was further analyzed using the DFT. In particular, the windowed waveform has its initial sample at the centroid of the click stimulus response ($t=0$ ms) and was zero padded out to a duration of 1024 samples.

CEOAE delays were examined in the time domain using overlapping third-octave bandpass filters, with center and edge frequencies computed according to ANSI S1.11 (2004). A Kaiser-based window method was used to design the filters. Filter order decreased as center frequency and bandwidth increased. The filters spanned four octaves (1–16 kHz), and the filter orders ranged from 500 to 32, which corresponded to filter group delays of 11.3 to 0.7 ms, respectively. Each subject's mean CEOAE waveform was bandpass filtered with each of the 13 third-octave filters. The group delays of the filters were subtracted from the filtered waveforms to ensure correct temporal alignment. A Hilbert transform was computed on each filtered waveform, and the temporal envelopes were computed as the magnitude of the transform.

V. RESULTS

A. CEOAE spectra

Figure 4 shows the median across subjects of the CEOAE SEL spectra for the six highest click levels that were analyzed (solid lines with circle symbols), with the corresponding median noise SEL at these click levels (dashed lines). The CEOAE spectra show an orderly increase in level with stimulus level, while the noise SEL was independent of click level. The CEOAE levels shown in Fig. 4 were above the noise floor for third-octave frequencies in the range of 2–8 kHz for all the stimulus click levels shown. A decreased CEOAE level at higher stimulus levels was evident above 8 kHz. At frequencies of 10–16 kHz, only the highest click levels (67 and 73 dB peSPL) resulted in CEOAE levels more than 4 dB above the noise SPL. Subsequent analysis therefore concentrates on CEOAEs elicited at the highest click stimulus level, which resulted in the largest SNRs.

Figure 5 shows the variability across subjects of CEOAE signal and noise SEL spectra recorded using the highest click level (73 dB peSPL) as a box and whiskers

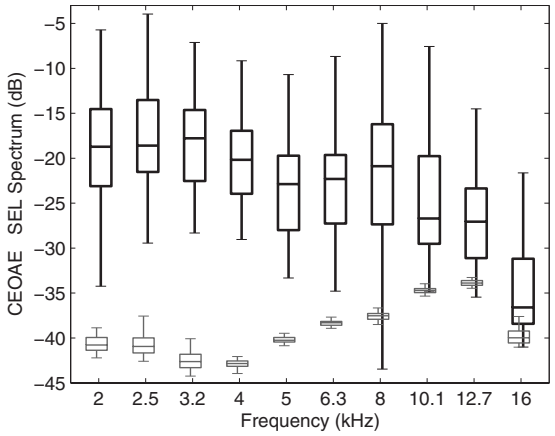


FIG. 5. Box and whiskers plots of the CEOAE SEL signal and noise spectra elicited at the highest click level (73 dB peSPL) are shown for $N=49$ ear responses. The black plot with thicker lines shows the CEOAE SEL spectrum, and the gray plot shows the noise SEL spectrum. The boxes show the lower quartile, median, and upper quartile values. The whiskers extend from each end of the boxes to show the extent of the rest of the data, with the maximum length being the lesser of the full range of SEL spectrum and 1.5 times the IQR. Outliers are not plotted, and signal and noise plots are slightly displaced horizontally to clarify the main features of the box plots.

plot. The "box" represents the IQR of the distributions of SEL. The IQRs for CEOAE signal SEL spectra were similar for frequencies of 2–16 kHz. The IQRs for noise levels were also similar across this frequency range. However, it was necessary to filter out an intermittent narrow-band noise spike in the microphone output near 15 kHz, which was present in some coupler recordings and some ear recordings, but not in others. This was a measurement system artifact and not a noise source of biological origin. The noise levels in Fig. 5 in the 16 kHz third octave represent the output after this exclusion of noise-contaminated bins (it should also be noted that data were analyzed only in the lower half of the third octave at 16 kHz).

Measurements made in the artificial ear and in human subjects suggest that a 3 dB SNR is an appropriate value for detecting the presence of an emission. CEOAES were present using this criterion between 2 and 6.3 kHz for all ears tested. At center frequencies of 8, 10.1, 12.7, and 16 kHz, CEOAEs were present in 92%, 78%, 66%, and 52% of ears, respectively.

B. CEOAE latency

The CEOAE latency of each third-octave filtered CEOAE output was calculated as the time corresponding to a maximum in the group-averaged temporal envelope of the CEOAE output waveform with respect to the time of the maximum temporal envelope of the click stimulus. When the energy in the temporal envelopes was averaged across subjects at each time step, results at the highest stimulus level (73 dB peSPL) showed the expected decrease in CEOAE latency with increase in filter center frequency (Fig. 6). The CEOAE latency is represented by an asterisk above the envelope response of each third-octave frequency. For frequencies higher than those previously reported for CEOAEs, the latency decreased from 2.0 down to 0.98 ms as frequency increased from 5 up to 16 kHz.

An envelope peak centered at $t=0$ was evident in the filter outputs at center frequencies at and above 8 kHz (Fig. 6). This energy was due to stimulus artifact in the click presented at $t=0$ (i.e., see Fig. 3). In terms of interpretation of CEOAE recordings, the peaks aligned at $t=0$ were ignored as unrelated to cochlear mechanics. The envelope peaks at later times in these high-frequency filter outputs provided a direct measure of CEOAE latency.

The CEOAE latencies derived from the mean envelope peaks at a stimulus level of 73 dB peSPL are replotted in Fig. 7 in units of the number of periods at each filter center frequency. The vertical error bars, which indicate the latencies at the half-power bandwidths of the peaks, provide a measure of the variability of latencies of the averaged group data. SFOAE latency data from Shera *et al.* (2002) (dashed lines show mean ±1 confidence interval) and Schairer *et al.* (2006) (thick solid line) are overlaid for comparison. Latencies in the present study were consistent with the predictions of Shera *et al.* (2002) and the measurements of Schairer *et al.* (2006) for frequencies between 1 and 2 kHz. Between 2 and 4 kHz, CEOAE latencies were slightly shorter than predicted by Shera *et al.* (2002) and slightly longer than measured by

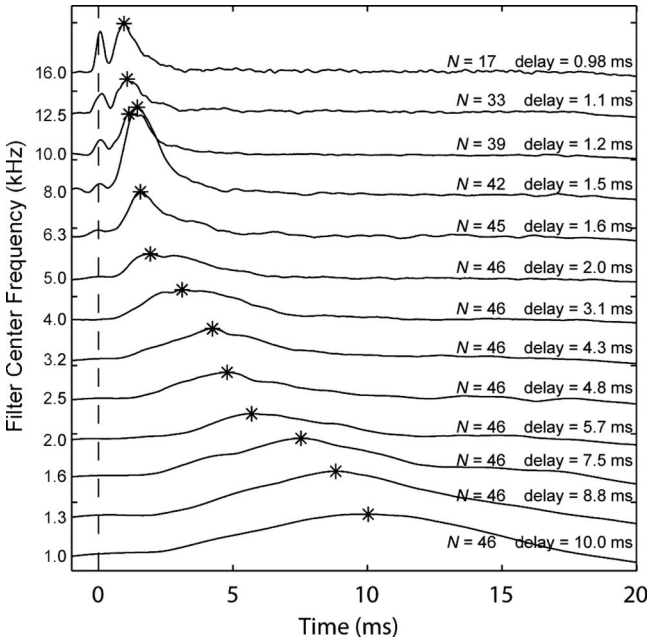


FIG. 6. Mean 1/3-octave filtered CEOAE waveform envelopes as a function of time. Data are from the highest stimulus level used (73 dB peSPL). Tracings are offset on the *y*-axis for clarity and arranged by filter center frequency in decreasing order. Only recordings with a SNR >3 dB were included in the averages. The number of ears (*N*) included in the mean is indicated for each tracing. The vertical dashed line indicates the stimulus onset. The peak of each envelope is indicated by an asterisk, and the corresponding delay (in ms) relative to stimulus onset is also displayed.

Schairer *et al.* (2006), but all latencies agreed to within measurement variability (the variability in Schairer *et al.*(2006) is reported in their article but not plotted in Fig. 7). At frequencies >5 kHz, the present latency data were shorter than the results of Shera *et al.* (2002). The cause of these differences is unknown, but possible explanations include differences in emission type (CEOAE versus SFOAE), differences

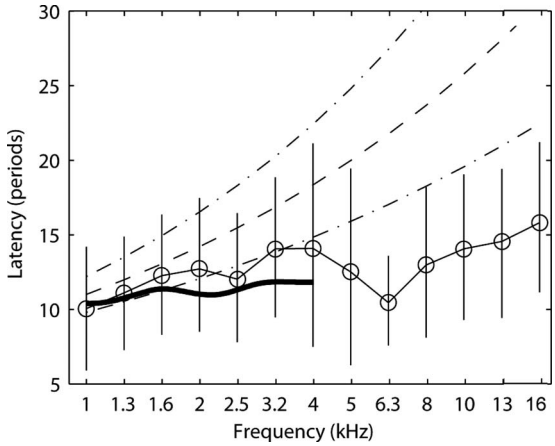


FIG. 7. Group means of CEOAE latency (thin solid lines and open circle symbols) are plotted nondimensionally in units of the number of periods at the filter center frequency. Data are for CEOAEs recorded at the highest stimulus level (73 dB peSPL). Vertical error bars indicate the latencies at the half-power bandwidths of the envelope peaks. The broken lines show the mean and 95% confidence intervals of SFOAE latencies reported by Shera *et al.* (2002) for stimulus levels of 40 dB SPL. The thick solid line shows the mean latencies of SFOAE latencies reported by Schairer *et al.* (2006) for stimulus levels of 40 dB SPL [variability not plotted for clarity but are shown in Schairer *et al.* (2006)].

in effective stimulus levels, differences in methodologies, and/or the presence of multiple components contributing to the measured latency. The role of level-dependent multiple components in CEOAEs is discussed below.

When averaged group latency data in the present study were compared across stimulus levels, a trend of decreasing CEOAE latency with increasing level was seen. This effect has been reported previously (e.g., Prieve *et al.*, 1996; Tognola *et al.*, 1997). However, most previous studies with human subjects have removed the first 2.5 ms of the ear-canal CEOAE recording prior to analysis in order to avoid stimulus artifact (Kemp *et al.*, 1990). The techniques used in the present study allowed the earlier portions of the CEOAE to be included in the analysis. These group results are not shown here because of the following properties observed in responses in individual ears.

Examination of individual subjects' data contributing to the group results in Figs. 6 and 7 suggested that there are two initial components to the CEOAEs: a longer-delay component that tends to dominate at lower stimulus levels and a shorter-delay component that dominates at higher levels. Examples of such individual-subject data are shown in Fig. 8. On average, the longer and shorter delays differed by a factor of approximately 1.6. A comparison of the envelope amplitudes of the earlier and later sources of individual-subject data suggested differences in growth as a function of stimulus level. Longer-delay components showed compressive growth, while shorter-delay components showed a more nearly linear growth. This is evident in the individual-ear responses plotted in Fig. 8.

In order to examine group data, growth was quantified as follows: Plots similar to the panels in Fig. 8 were produced, and the main delay components (peaks) were visually identified. Each component consisted of the responses to four stimulus levels. In some subjects, four peaks were identified at a particular stimulus level. In other subjects, the responses to lower stimulus levels were below the noise floor. Wherever pairs of peaks at adjacent stimulus levels were identified in an ear, the peak amplitude at the current stimulus amplitude was divided by the peak amplitude at the next lower stimulus amplitude; this lower stimulus amplitude was one-half the amplitude of the current stimulus (i.e., −6 dB in relative level). The mean of these ratios was taken as the average growth of the CEOAE residual for that component. These average growth values were normalized so that a value of 1 indicated linear growth, <1 indicated compressive growth, and >1 indicated expansive growth. A value of 0.5 represented full saturation, i.e., no change in emission level as stimulus increased. A value of 2 indicated growth at twice the rate of linear growth, i.e., a fourfold increase in amplitude for each doubling of stimulus amplitude.

Figure 9 shows scatter plots of growth values plotted as a function of component delay using data from the four filter center frequencies (2, 4, 8, and 10 kHz) used in these group analyses. Data are also grouped by component number, with open circles, asterisks, and open triangles showing growth values from the first, second, and third components, respectively. The growth rate of the CEOAE residuals was more compressive as component delays increased. CEOAE residuals were sometimes fully saturated (a growth value of 0.5), but they never showed negative growth. Because the growth decayed approximately exponentially with component delay, the data were fitted with an exponential function of the form

$$f(x) = \alpha \exp(-\beta x) + 0.5, \tag{2}$$

where x is peak delay in milliseconds. Values of α and β were determined for each frequency using an iterative nonlinear least squares fitting procedure with bisquares weighting (MATLAB CURVE FITTING TOOLBOX 1.2.1).

Note that the CEOAE temporal envelopes described above are envelopes of the nonlinear residual extracted from the measured waveforms (as described in Sec. IV C). Thus, an apparent linear growth in the envelope amplitude in Figs. 8 and 9 corresponds to a quadratic growth in the total

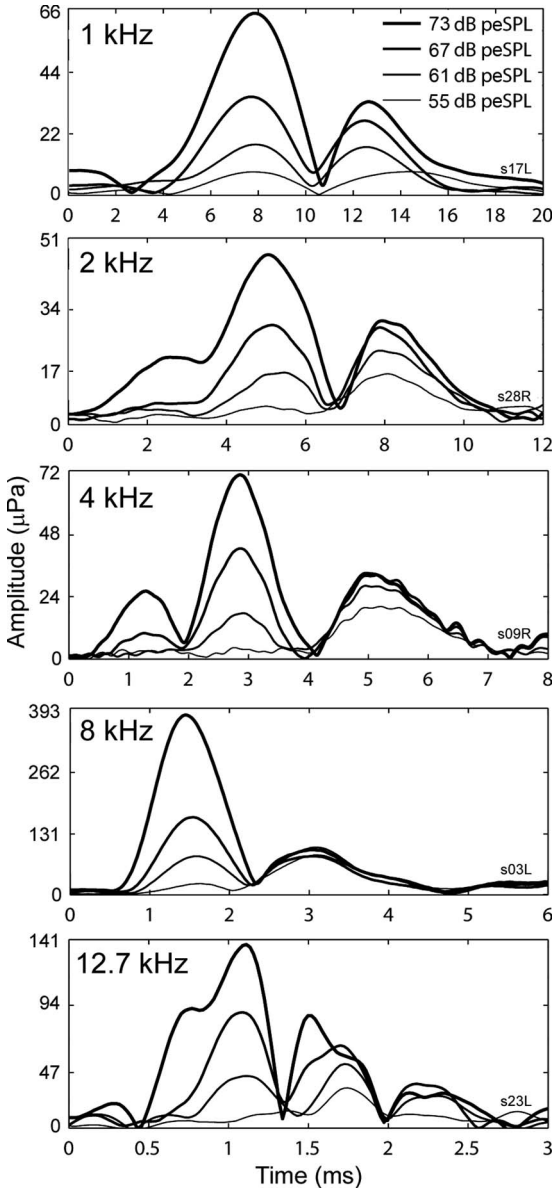


FIG. 8. 1/3-octave filtered CEOAE waveform envelopes as a function of time and stimulus level. Panels show increasing filter center frequency (kHz). The tracings in each panel are from a different subject. The plotting axes in each panel encompass different ranges to allow the examination of salient details. No synchronized spontaneous OAEs were present in the data shown.

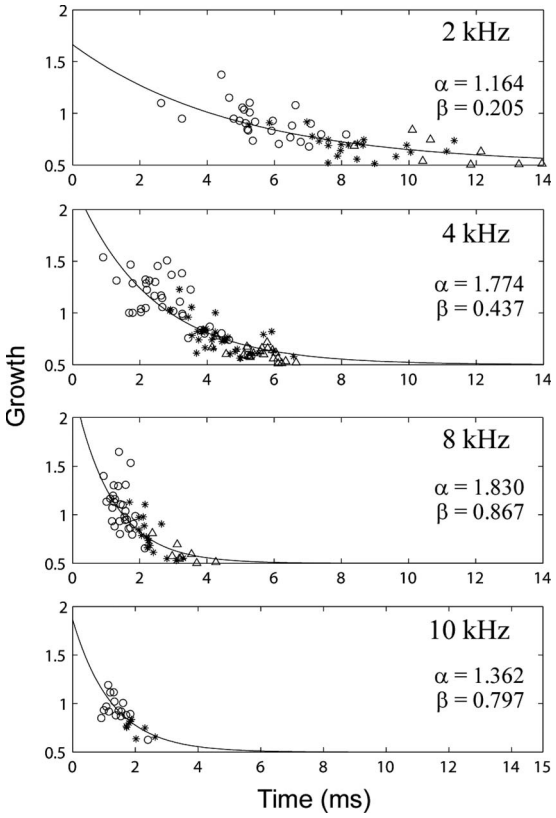


FIG. 9. Growth in peaks of emission as a function of peak delay. From top to bottom, panels show increasing filter center frequency. Open circles, asterisks, and open triangles show growth values from the first, second, and third components, respectively. α and β are the parameters of the exponential fit (solid line).

CEOAE. The relevant property of the shorter-delay components is that their growth was faster than compressive growth. The growth patterns of the later components are consistent with a coherent reflection mechanism on the basilar membrane (Zweig and Shera, 1995; Prieve *et al.*, 1996; Kalluri and Shera, 2007), as would be the result of a compressive-growth basilar-membrane input-output function. The more linear or expansive-growth patterns of the earlier components in the CEOAE residual are consistent with a nonlinear distortion source (Withnell *et al.*, 2008). Components with the longest latencies may also have included effects of one or more internal reflections within the cochlea between its base and tonotopic place.

An important observation regarding both short- and long-delay components is that although their relative dominance shifted as a function of stimulus level, there was essentially no change in the delay of the envelope of either component as a function of level. This pattern is evident in Fig. 8. Group data were examined using a strategy similar to that described above. The main delay components were identified, each consisting of responses to four stimulus levels (55, 61, 67, and 73 dB peSPL). Wherever two or more peaks were identified, the maximum amplitude of each peak was recorded. A linear regression line was fitted to the points, with its slope indicating the rate of change in delay per doubling of stimulus amplitude.

Because earlier and later components had different growth rates, it was hypothesized that they might show different changes in delay as a function of stimulus level. Accordingly, individual-ear data were visually grouped by component number, with the earliest component designated as component 1. Some subjects had only one component identified, while others had two or three. A few subjects had four components, but fourth components were not considered in this analysis. The data comprising component 1, as a group, showed expansive or linear growth. The data comprising component 2 were linear or compressive, and the data comprising component 3 were mostly compressive in growth. These group patterns were similar to the individual patterns in Fig. 8. Data were statistically analyzed for three components at each of 2, 4, and 8 kHz and at 10 kHz for a total of 11 components tested. T-tests were performed on each component to test the null hypothesis that component delay was independent of stimulus level. A false discovery rate adjustment (Benjamini and Hochberg, 1995) was made to control for the proportion of falsely rejected hypotheses when conducting multiple significance tests. Three of the 11 groups were significantly different from zero at the 5% level: Component 1 at 2 and 8 kHz and component 2 at 4 kHz each had decreased latency with increasing stimulus level. The significant effects were small compared to the component latency: the component shifts were −0.058, −0.038, and −0.026 ms per 6 dB relative increase in stimulus SPL. In other words, the change in delay seen between a stimulus of 55 dB peSPL and a stimulus of 73 dB peSPL was on the order of 0.16 ms.

Summarizing the results of group data analysis, it was found that earlier-occurring components tended to grow expansively, consistent with a distortion mechanism, and showed a small but significant decrease in delay as the stimulus level increased. Later-occurring components tended to grow compressively, consistent with a coherent reflection mechanism, and showed no change in delay as a function of stimulus level.

C. Maximum-likelihood thresholds

Air-conduction thresholds obtained using the ML procedure were compared with air-conduction thresholds obtained using standard clinical procedures at frequencies up to 8 kHz at which both procedures were used. Figure 10 shows scatter plots of the clinical versus the ML audiometric thresholds for octave frequencies between 0.5 and 8 kHz, with brackets indicating a range of +1 SD. For these results, the HL reference for the probe (as well as the clinical system) was based on a calibration in the HA-1 coupler according to ANSI S3.6 (2004). The mean clinical and ML thresholds measured using the 250 ms tone bursts were within ±1 SD at 0.5, 1, and 2 kHz, but the mean ML thresholds were slightly lower at 4 and 8 kHz.

ML thresholds were also analyzed at frequencies up to 16 kHz based on the reference SPL in the anechoic tube, as determined using measurements in the HA-1 coupler and the transfer function L_A+L_B (see Fig. 2). Box and whisker plots of the ML threshold SPL are shown in Fig. 11. The median of this threshold SPL is defined at each frequency as the

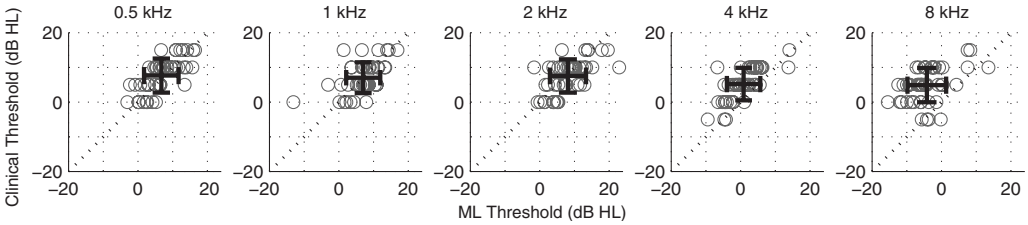


FIG. 10. Scatter plots of clinical vs the maximum-likelihood audiometric thresholds (in HL) for five frequencies. Individual ear thresholds are shown as open gray circles. The mean ±1 SD of the thresholds is plotted as bars in black. Median thresholds are shown as black filled circles.

reference equivalent threshold sound pressure level (RETSPL) based on the incident-pressure procedure.

While subjects were included only if their clinical thresholds were <15 dB HL up to 8 kHz, there was no clinical reference available at higher frequencies. Consequently, a more elevated and wider range of hearing was expected and observed; e.g., the results in Fig. 11 show a 77 dB range in thresholds measured at 16 kHz. At frequencies >8 kHz, the median SPL threshold increased with frequency, in qualitative agreement with previous high-frequency audiometric studies (Green *et al.*, 1987; Stelmachowicz *et al.*, 1989a, 1986). The ML threshold technique had similar efficiency at all frequencies, and thresholds at high frequencies were no more complicated to measure than at low frequencies. The RETSPL calibrated according to the ER-10B+ microphone in the HA-1 coupler would be unduly influenced by the high-frequency standing waves evident in the transfer function L_B between this microphone and the reference microphone (see middle panel of Fig. 2). These standing waves produce a total variation of over 50 dB in L_B across frequency.

The significance of these results is that the determination of the incident-pressure RETSPL in a group of young normal-hearing listeners, as measured by measurements in the anechoic tube using the ER-10B+ microphone, makes possible behavioral audiometry and OAE measurements using the same probe to frequencies as high as 16 kHz.

The CEOAE SEL spectrum decreased with increasing

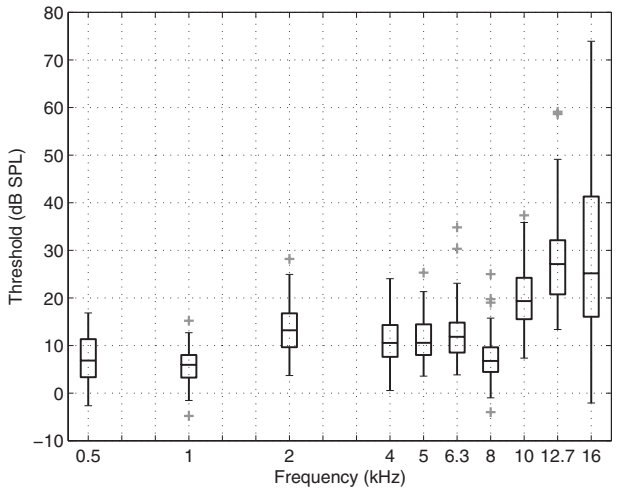


FIG. 11. ML thresholds expressed as the incident SPL at the threshold (i.e., the SPL on each tone burst that would be measured in the anechoic tube), as referenced to the calibration of the ER10-*B*+ and ER-2 earphones in the no-extension probe condition. These thresholds are shown as a box and whiskers plot, with each outlier plotted using a plus symbol (+).

ML threshold at 8 and 10.1 kHz, and associated correlations accounted for 28% and 43%, respectively, of the total variance. This supports the view that a reduction in the cochlear source strength of the emission was associated with an elevated threshold. While similar trends were evident at 12.7 and 16 kHz, no significant correlation was found at either frequency. One contributing factor to the absence of such a relationship may be the low SNRs at 12.7 and 16 kHz (see Figs. 4 and 5).

VI. DISCUSSION

A. CEOAE spectra

The problem of calibrating the sound stimulus at high audio frequencies was successfully addressed by measurements in a long smooth-walled rigid tube. Such a tube functioned in an anechoic manner over sufficiently long recording times that it was possible to measure the incident sound level without the contaminating effects of acoustic standing waves. This enabled specification of the stimulus SPL over a broad frequency range, 0.25–16 kHz. An efficient ML procedure to measure audiometric threshold based on a yes-no task resulted in similar results to clinical audiometry up to 8 kHz and provided audiometric results at higher frequencies that were qualitatively similar to those in previous reports. An advantage in our procedures was that it was based on the use of insert earphones, which are routinely used to measure audiograms at lower frequencies and which are used to measure OAEs.

The experiments demonstrated that CEOAEs can be recorded in some otologically normal adult ears up to 16 kHz. Such responses serve as a noninvasive probe of cochlear mechanics at the base of the basilar membrane. Previous CEOAE measurements are restricted to an upper frequency of approximately 5 kHz. This frequency limitation was a property of the recording technique (and any probe transducer limitations) rather than any limitation in the ability of the human cochlea to produce a CEOAE response at higher frequencies. The ability to measure a CEOAE response up to 16 kHz has potential clinical applications as an objective physiological screening test for high-frequency sensorineural hearing loss.

While the methods described in this paper resulted in measurable high-frequency CEOAEs in many subjects, the SPL of the emissions was reduced at frequencies above 8 kHz (see median responses in Fig. 5). There are several possible reasons for this. First, the microphone sensitivity was reduced at 16 kHz, so that the stimulus SPL was reduced

at the highest frequencies. Second, subject inclusion criterion specified that hearing thresholds be within normal limits only up to 8 kHz. Our results suggest that subjects had varying degrees of hearing sensitivity above 8 kHz. Stelmachowicz *et al.* (1989a, 1989b) compiled normative thresholds between 8 and 20 kHz as a function of age. They showed that age-related threshold shifts were greatest above 8 kHz and tended to begin starting in the second decade of life. The participants in this study had a mean (SD of) age of 20.5 +3.4 years, with ages symmetrically distributed around the mean. Thus, many subjects could be expected to have had a hearing loss above 8 kHz. The increased measurement variability in the high-frequency thresholds shown in Fig. 11 supports this expectation. This variability probably reflects the changing thresholds within a subject pool spanning 15 years.

A third reason that may account for the decrease in CEOAE SPL above 8 kHz relates to the relative ability of the middle ear to transmit sound energy at higher frequencies. Middle-ear transmission measurements in human temporal bones show a decrease with frequency above about 1 kHz for both forward and reverse transmission (Puria *et al.*, 1997; Puria, 2003). Interestingly, CEOAE SPL spectra do not closely resemble the round-trip middle-ear pressure transfer function, at least between 2 and 6 kHz where they appear to decline with frequency at a slower rate (Smurzynski and Kim, 1992). One explanation is that the mechanisms generating the CEOAEs are frequency dependent (Puria, 2003). Alternatively, indirect measurements using OAEs suggest a broader middle-ear transfer function, more consistent with the levels of the emissions, as described in Keefe (2007). These indirect measurements showed a rolloff between 4 and 8 kHz. Such indirect middle-ear transmission estimates have not been reported for frequencies above 8 kHz. Additional direct measurements of middle-ear transmission in human temporal bones are also needed to better understand the high-frequency regime relevant to the CEOAE measurements in the present study.

The sound level of the stimulus used in this experiment was relatively flat in frequency up to approximately 11 kHz and rolled off at higher frequencies (Fig. 1, top panel). Due to an additional rolloff in the forward middle-ear transfer function at high frequencies, stimulation of the cochlea at places corresponding to frequencies above 8 kHz may not have been as great as stimulation at places corresponding to lower frequencies. It might be preferable to increase the stimulus energies at and above 8 kHz in future studies.

The incident power in the ear canal, which is the integral of the incident acoustic intensity flowing through the cross-sectional area at a given location and proportional to the square of the incident pressure, is not influenced by standing waves and thus is well suited as a calibration standard for audiometric and CEOAE measurements. Nevertheless, the acoustic pressure is markedly influenced by these standing waves depending on the measurement location of the microphone within the ear canal. This has been a topic of concern in studies of OAEs at higher frequencies (Siegel, 1994; Whitehead *et al.*, 1995b). Siegel (2007) reviewed issues related to OAE probe calibration and measurements at high frequencies. The distinctive feature of the incident-pressure approach used in the present study to measure OAEs at high frequencies is that it does not require a measurement of an aural acoustic transfer function such as admittance or reflectance. At high frequencies in the ear canal, a small shift in the OAE probe sound source or microphone can produce a large shift near a frequency at which a standing-wave minimum is present. However a small shift in the OAE probe produces no change in the incident pressure, at least to the extent that variations in the cross-section area of the ear canal are small. In any case, such area changes might produce effects on the order of a couple dB, whereas standing waves may produce SPL variations as large as 20 dB (Siegel, 1994). Studies of the repeatability of CEOAEs at high frequencies need to be performed, but specifying the stimulus level as incident pressure may reduce the variability of CEOAE levels, especially after frequency averaging over third octaves. The variation in the incident-pressure SPL was small below 11 kHz (see the No ext. curve in the top plot of Fig. 1).

Measuring an aural acoustic transfer function would allow calculation of the power absorbed from the OAE stimulus by the middle ear (and by ear-canal walls, particularly in young infants) (Keefe *et al.*, 1993). This power absorbtion is closely related to acoustic intensity flow within the ear canal (Neely and Gorga, 1998; Farmer-Fedor and Rabbitt, 2002). It is outside the scope of the present work to consider the additional information that may be gained by combined measurements of OAEs and aural acoustic transfer functions in the same ear, but it is an area of current interest (Keefe, 2007; Scheperle *et al.*, 2008).

B. CEOAE delays

Individual subjects' data suggested that there are at least two components to the CEOAEs: a compressive-growth, longer-delay component that tends to dominate at lower stimulus levels and a linear-growth, shorter-delay component that dominates at higher levels (Fig. 8). Some subjects showed additional early and/or late peaks (e.g., Fig. 8, 4 kHz panel and 12.5 kHz panel).

The growth pattern and delay of the later components are mainly consistent with a coherent reflection mechanism (Zweig and Shera, 1995; Prieve *et al.*, 1996; Kalluri and Shera, 2007). This mechanism incorporates a linear, spatially distributed set of reflections of the basilar-membrane traveling wave that are coherently filtered at a given frequency by the tall broad peak in the basilar-membrane displacement envelope near the tonotopic place (i.e., the place on the basilar membrane at which its displacement for a tone at a particular frequency is maximal). The saturating nonlinearity in the CEOAE components associated with coherent reflection arises from the saturating gain characteristics of outer hair cell motion.

Although there were small decreases in latency as a function of stimulus level for the early component, a much larger effect was the shift in dominance from later to earlier components as stimulus level increased. No change in latency was found for later components. This finding is con-

sistent with Konrad-Martin and Keefe (2005) (their Fig. 6) and Carvalho *et al.* (2003), who reported similar results for the later-occurring component. Thus, the CEOAE latency based on the entire response decreased with increasing stimulus level, but this was due to a change in the relative proportion of the energies of the short-latency and long-latency components. These results extend the work of Carvalho *et al.* (2003) by reporting that phase changes coupled with near-peak invariance is seen for both early and late components, as well as at frequencies up to 12 kHz. Above 12 kHz, a component with a longer compressive-growth delay was not identified. This may have been due to poorer SNRs, a difference in CEOAE generation at these higher frequencies, or a temporal overlap between the early and late components.

These results have implications for theories concerning the generation of CEOAEs. Kalluri and Shera (2007) suggested that CEOAEs (at least those occurring later than 5 ms poststimulus delivery) are generated by a coherent reflection mechanism present in independent channels, as opposed to being generated by intermodulation distortion between channels. A nonlinear generalization of the coherent reflection mechanism would predict that the delay of the (later) component should decrease with increasing stimulus level due to compressive growth of the traveling wave on the basilar membrane with increasing stimulus level, whereas the present finding is that this delay was approximately constant with increasing level. While the theory of coherent reflection mechanism, as originally proposed (Zweig and Shera, 1995), was used to interpret SFOAE data only at a low stimulus level (40 dB SPL), it is widely thought that this theory may apply to OAEs at higher stimulus levels. The idea is that the peak of the excitation pattern on the basilar membrane is shifted basally and the phase delay to tonotopic place is reduced with increasing stimulus level (Recio and Rhode, 2000; Lin and Guinan, 2000; de Boer and Nuttall, 2000; though see also Moore and Glasberg, 2003). The net result from both effects would be a reduction in the OAE delay with increasing level. The finding that the delay of the later CEOAE component is approximately constant is not in accord with this nonlinear generalization of coherent reflection theory.

Schairer *et al.* (2006) reported that SFOAE delays decreased approximately uniformly over frequencies of 0.5–4 kHz by a factor of 1/2 with increasing stimulus level over a 20 dB range. This is in apparent agreement with the coherent reflection theory, as generalized to include the compressive nonlinear of basilar-membrane mechanics. It is unknown the extent to which undetected multiple components may have contributed to the calculated SFOAE delays, which were calculated using the phase gradient at the frequency of the tonal stimulus. The phase-gradient method requires the assumption of a single-component SFOAE. If, as suggested by Kalluri and Shera (2007), CEOAEs and SFOAEs share a common mechanism, this assumption might not hold for SFOAEs at higher stimulus levels. Future studies should examine SFOAEs for the presence of multiple components, but the relative latency changes with increasing stimulus level of individual CEOAE components in the present study appear smaller than the relative latency changes of SFOAEs reported in Schairer *et al.* (2006). While multiple components of CEOAE latency were identified through differing nonlinear rates of growth, the present measurements did not allow identification of the origination regions within the cochlea of the components that contributed to our ear-canal measurements.

The approximately linear growth of the earlier component of the CEOAE nonlinear residual is consistent with a nonlinear distortion source. Such a linear growth in the CEOAE residual corresponds to a quadratic growth in the total CEOAE and is not consistent with the saturating growth of a coherent reflection model. The presence of a nonlinear distortion component in the early CEOAE is in agreement with recent work by Withnell *et al.* (2008), who used the relatively shallow phase gradient of the early time-windowed response to infer a distortion-generated component. Our data suggest some potential complications with the phase-gradient method. Some subjects showed evidence of multiple early linear—or expansive-growth peaks (e.g., Fig. 8, 4 and 12.5 kHz panels). The relationship between the relatively invariant envelope peak latencies in the present study and the time shifts inferred from the phase-gradient method has not been well characterized. These issues raise questions regarding absolute delay values obtained by the phase-gradient method and suggest that further study is needed. Nevertheless, regardless of these measurement issues, the two studies agree that a distortion component is present in human CEOAEs.

Basilar-membrane mechanical responses to clicks measured at the cochlear base in chinchilla (Recio *et al.*, 1998) show a "two-lobed" waveform envelope that mainly grows compressively with level, consistent with the dominant CEOAE response growth properties in the present study. However, at high click levels, the earliest envelope peak on the basilar membrane has a faster, almost linear rate of growth than later envelope peaks. Basilar-membrane responses are measured on a small region of the basilar membrane, whereas a CEOAE, even after bandpass filtering, is likely to integrate over a spatially extended region of the basilar membrane. Nevertheless, a high-frequency OAE response is likely generated from sources near the basal end of the cochlea. The variations in rates of growth of response in the basal region of the basilar membrane may be related to the level dependence observed in our CEOAE latency results. Dominance of a shorter-latency source region in basilar-membrane mechanics would shift the CEOAE envelope peak to shorter latencies, as was observed in the present study. Further research is needed to better understand relationships between click-evoked OAE responses and basilar-membrane responses.

C. Behavioral threshold prediction using CEOAEs

Having established that high-frequency CEOAE responses can be measured in some otologically normal ears, it remains to be determined whether such responses can be used to detect the presence of a high-frequency hearing loss in a population of subjects with varying degrees of senso-

rineural hearing loss. Analyses at 8 and 10.1 kHz showed that CEOAE SEL was reduced in ears with slightly elevated high-frequency thresholds. While this is a promising result, more studies are needed in subjects with a broader range of thresholds.

An important use for high-frequency audiometry has been monitoring for ototoxic hearing loss. Since ototoxic hearing loss affects the high frequencies first, an OAE detection paradigm measuring OAEs at frequencies >8 kHz has the potential to lead to earlier detection. Such a paradigm might not be effective for older patients who already have high-frequency hearing loss but may be advantageous for younger patients. If future studies show that high-frequency CEOAEs can detect a high-frequency sensorineural hearing loss, related future studies might examine the feasibility of monitoring high-frequency CEOAEs in patients receiving ototoxic medications who are at risk for hearing loss and compare the relative efficacy of high-frequency CEOAEs with DPOAEs and SFOAEs.

VII. CONCLUSIONS

The problem of characterizing the acoustic source level of a probe designed for ear-canal measurements at high frequencies was solved by an incident-pressure calibration procedure based on reference measurements in a long cylindrical tube, which functioned as an acoustically anechoic termination. High-frequency behavioral thresholds up to 16 kHz were measured using this calibrated source based on an efficient ML procedure. The results confirmed previous research showing a wide variation in audiometric levels above 8 kHz in a population of subjects with normal audiometry at lower frequencies. In teenage and young-adult subjects with normal hearing up to 8 kHz, CEOAEs were obtained at frequencies up to 16 kHz. The CEOAE SEL spectrum decreased with increasing audiometric threshold in the third octaves centered at 8 and 10.1 kHz, which suggests the possibility that high-frequency CEOAEs may be useful in a test to detect high-frequency sensorineural hearing loss. Third-octave filtered CEOAE residual waveform envelopes showed an earlier linear-growth component, which implies a quadratic growth in the total CEOAE waveform, and a later compressive-growth component consistent with nonlinear compression acting on basilar-membrane mechanics. The envelope delays of earlier components showed small but statistically significant decreases in latency with increases in level. The envelope delays of later components were invariant with level. CEOAE latency based on the entire response decreased with increasing stimulus level, but this was mostly due to a change in the relative proportion of energies of earlier and later components.

ACKNOWLEDGMENTS

The authors thank two anonymous reviewers for their helpful critiques of a previous version of this report and thank Jonathan K. Stewart of Etymotic Research, Inc. for detailed information regarding the probe-microphone calibration. This research was supported by NIH Grant Nos. DC07023 and DC03784, with core support from Grant No. DC04662.

APPENDIX: SOUND LEVEL SPECIFICATION FOR TRANSIENT-EVOKED OTOACOUSTIC EMISSIONS

The specification of the sound level of the stimulus and response for transient-evoked OAEs, which include CEOAEs, differs somewhat from that for tonal-evoked OAEs such as SFOAEs and DPOAEs. For example, the measured SPL of a sinusoidal tone is substantially independent of its measurement duration, whereas the SPL of a transient such as a click decreases with increasing measurement duration. This appendix describes how sound levels are reported for transient responses measured using discrete-time signal processing based on the DFT. These generally correspond to definitions of sound levels for transient responses measured using continuous-time signal processing based on the Fourier transform (Young, 1970; Pierce, 1989). These relationships have significance for understanding differences in sound levels reported for transient-evoked and tone-evoked OAEs.

The sound exposure E from a transient pressure waveform $p(t)$ in a continuous-time representation and from a pressure finite sequence $p[n]$ in a discrete-time representation with sample period T and buffer length of N samples is

$$E = \int_0^{NT} dt|p(t)|^2 = T\sum_{n=0}^{N-1} |p[n]|^2, \tag{A1}$$

with the integral or sum extending over the time duration NT of the measurement and with n representing the sample number. The effective duration of the transient is assumed less than NT (else a larger N would be used). A SEL L_E is defined as

$$L_E = 10\log_{10}\left(\frac{E}{P_{\text{ref}}^2 T_{\text{ref}}}\right), \tag{A2}$$

in which the reference pressure is $P_{\text{ref}}=2\times10^{-5}$ Pa and the reference averaging time T_{ref} may be arbitrarily specified with a default value of 1 s in continuous-time analysis. The SEL quantifies the level (in dB) of a transient sound.

$P[k]$ denotes the DFT of $p[n]$ at the kth frequency bin with center frequency $f_k=k/(NT)$, e.g., as defined in Oppenheim and Schafer (1989). Because $p[n]$ is real, the so-called Parseval relation associated with the DFT for even N is

$$\sum_{n=0}^{N-1} |p[n]|^2 = \frac{1}{N}\sum_{k=0}^{N-1} |P[k]|^2 = \frac{1}{N}\left(\sum_{k=1}^{N/2-1} 2|P[k]|^2 + |P[0]|^2 + |P[N/2]|^2\right). \tag{A3}$$

The spectral terms at $k=0$ and $k=N/2$ are outside the measurement bandwidth and are thus discarded in the following. Using Eqs. (A1)–(A3), the SEL within the measurement bandwidth is

$$L_E = 10\log_{10}\left(\frac{\frac{T}{N}\sum_{k=1}^{N/2-1} 2|P[k]|^2}{P_{\text{ref}}^2 T_{\text{ref}}}\right). \quad \text{(A4)}$$

Generalizing the above equation to the kth spectral frequency, a band sound-exposure spectrum level L_{Eb}, or SEL spectrum, is defined by

$$L_{Eb} = 10\log_{10}\left(\frac{\frac{T}{N}2|P[k]|^2}{P_{\text{ref}}^2 T_{\text{ref}}}\right) = 10\log_{10}\left(\frac{\frac{2}{N^2}|P[k]|^2/\Delta f}{P_{\text{ref}}^2 T_{\text{ref}}}\right), \quad \text{(A5)}$$

with the latter defined for reasons described below. The DFT bandwidth of each spectral bin is $\Delta f=(NT)^{-1}$. Other band SELs can be defined over octaves or other averaging bandwidths using an appropriate partial sum over multiple DFT frequency bins in place of the sum used in Eq. (A4), as was used in the present work to calculate the third-octave band SEL spectra.

The SEL spectrum of a single transient is related to the SPL spectrum of a periodic sequence of transients, which has a period equal to the duration NT of the single transient. Alternatively, the finite sequence $p[n]$ of length N might be sampled from an underlying random signal, and a periodic sequence might be formed to facilitate use of the DFT in periodogram analysis (Oppenheim and Schafer, 1989). Using a terminology similar to that in ANSI S1.1 (1994) for the continuous-time signal analysis, a band sound pressure spectrum level L_{pbs} is

$$L_{pbs} = 10\log_{10}\left(\frac{\frac{2}{N^2}|P[k]|^2/\Delta f}{P_{\text{ref}}^2/\Delta_{\text{ref}} f}\right), \quad \text{(A6)}$$

in which the reference bandwidth is $\Delta_{\text{ref}} f$ with a default value of 1 Hz. The L_{pbs} with this default is the SPL spectrum (re 1 Hz bandwidth). In the numerator, the kth spectral band component $\frac{2}{N^2}|P[k]|^2$ may be produced by an underlying continuous distribution of frequency components so that its spectral density over the bandwidth Δf is $\frac{2}{N^2}|P[k]|^2/\Delta f$. For a sinusoid of frequency f_k, the SPL L_p can be written as

$$L_p = 10\log_{10}\left(\frac{\frac{2}{N^2}|P[k]|^2}{P_{\text{ref}}^2}\right), \quad \text{(A7)}$$

so that $L_p=L_{pbs}+10\log_{10}(\Delta f/\Delta_{\text{ref}} f)$.

A comparison of Eqs. (A5) and (A6) shows that the SEL spectrum (re 1 s averaging time) of a transient sound is numerically equal to the SPL spectrum (re 1 Hz bandwidth) of the repeated transient sound. This would require that the duration NT of the underlying DFT be 1 s, which is not typically the case in practical measurements. A convenient choice for discrete-time measurements using the DFT is that the reference averaging time is equal to the DFT buffer length, i.e., $T_{\text{ref}}=NT$. For example, this is appropriate for a click train with period NT. The SEL spectrum for this choice of T_{ref} simplifies to

$$L_{Eb} = 10\log_{10}\left(\frac{\frac{2}{N^2}|P[k]|^2}{P_{\text{ref}}^2}\right). \quad \text{(A8)}$$

The relationship between the SEL spectrum of the single transient and the SPL spectrum of the periodic sequence of this transient is

$$L_{Eb} = L_{pbs} - 10\log_{10}(NT). \quad \text{(A9)}$$

The observation that the right-hand sides of Eqs. (A7) and (A8) are equal carries no special significance because they correspond to different types of measurements.

The actual stimulus presentation in the CEOAE measurement was not a periodic sequence of equal-amplitude clicks but, as described in Sec. IV D, used three click stimuli of varying amplitude with an interclick interval of 25.5 ms. Specifying the transient sound level using the SPL spectrum must also include the conditions under which the SPL should be interpreted. The simpler choice used in the present study specifies the transient sound level using the SEL spectrum based on Eq. (A8).

[1]At frequencies above 8 kHz, the sensitivity level of the Etymotic ER-10B+ microphone had a maximum of −18.8 dB (re 1 V/Pa) at 11.7 kHz with a 3 dB quality factor (Q) of 8. It had a minimum sensitivity level of −31.2 dB at 16.1 kHz with a Q of 3.6. The sensitivity levels at third-octave frequencies ⩾8 kHz were −27.5 dB at 8 kHz, −24.6 at 10.1 kHz, −24.3 dB at 12.7 kHz, and −31.2 dB at 16 kHz, which were relative to the nominal sensitivity level of −26 dB at lower frequencies.

[2]Consistent with Sec. III, this peak-to-peak pressure amplitude was calculated in terms of the peak-to-peak voltage amplitude in the ADC recording and the nominal microphone sensitivity. This involves some error because of the unknown phase sensitivity of the probe microphone above 8 kHz. Nor was the level sensitivity of the probe microphone applied, which varied from the nominal sensitivity above 8 kHz, because in the absence of a phase calibration, an accurate peak-to-peak pressure amplitude cannot be fully specified. Nevertheless, calculating peSPL based on the nominal microphone sensitivity was sufficient for the goals of this study.

Agullo, J., Cardona, S., and Keefe, D. H. (**1995**). "Time-domain deconvolution to measure reflection functions from discontinuities in waveguides," J. Acoust. Soc. Am. **97**, 1950–1957.

ANSI S1.1 (**1994**). Acoustical Terminology (American National Standards Institute, New York).

ANSI S1.11 (**2004**). Specification for Octave-Band and Fractional-Octave-Band Analog and Digital Filters (American National Standards Institute, New York).

ANSI S3.6 (**2004**). Specification for Audiometers (American National Standards Institute, New York).

ANSI S3.7 (**1995**). Methods for Coupler Calibration of Earphones (American National Standards Institute, New York).

Benjamini, Y., and Hochberg, Y. (**1995**). "Controlling the false discovery rate—A practical and powerful approach to multiple testing," J. R. Stat. Soc. Ser. B (Methodol.) **57**, 289–300.

Brass, D., and Kemp, D. T. (**1991**). "Time-domain observation of otoacoustic emissions during constant stimulation," J. Acoust. Soc. Am. **90**, 2415–2427.

Brummett, R. E. (**1980**). "Drug-induced ototoxicity," Drugs **19**, 412–428.

Burkhard, M. D., and Sachs, R. M. (**1977**). "Sound pressure in insert earphone couplers and real ears," J. Speech Hear. Res. **20**, 799–807.

Carvalho, S., Buki, B., Bonfils, P., and Avan, P. (**2003**). "Effect of click intensity on click-evoked otoacoustic emission waveforms: Implications for the origin of emissions," Hear. Res. **175**, 215–225.

Chan, C. K., and Geisler, C. D. (**1990**). "Estimation of eardrum acoustic pressure and of ear canal length from remote points in the canal," J. Acoust. Soc. Am. **87**, 1237–1247.

de Boer, E., and Nuttall, A. (**2000**). "The mechanical waveform of the

basilar membrane. III. Intensity effects," J. Acoust. Soc. Am. **107**, 1494–1507.
Dreisbach, L. E., Long, K. M., and Lees, S. E. (**2006**). "Repeatability of high-frequency distortion-product otoacoustic emissions in normal-hearing adults," Ear Hear. **27**, 466–479.
Dreisbach, L. E., and Siegel, J. H. (**2001**). "Distortion-product otoacoustic emissions measured at high frequencies in humans," J. Acoust. Soc. Am. **110**, 2456–2469.
Dreisbach, L. E., and Siegel, J. H. (**2005**). "Level dependence of distortion-product otoacoustic emissions measured at high frequencies in humans," J. Acoust. Soc. Am. **117**, 2980–2988.
Dreisbach, L. E., Siegel, J. H., and Chen, W. (**1998**). "Stimulus frequency otoacoustic emissions measured at low- and high-frequencies in untrained human subjects," Abstracts of the Twenty-First Annual Midwinter Research Meeting of the Association for Research in Otolaryngology, p. 88 (abstract).
Dreschler, W. A., van der Hulst, R. J., Tange, R. A., and Urbanus, N. A. (**1989**). "Role of high-frequency audiometry in the early detection of ototoxicity. II. Clinical aspects," Audiology **28**, 211–220.
Farmer-Fedor, B. L., and Rabbitt, R. D. (**2002**). "Acoustic intensity, impedance and reflection coefficient in the human ear canal," J. Acoust. Soc. Am. **112**, 600–620.
Fausti, S. A., Erickson, D., Frey, R., Rappaport, B. Z., and Schechter, M. A. (**1981**). "The effects of noise upon human hearing sensitivity from 8000–20000 Hz," J. Acoust. Soc. Am. **69**, 1343–1349.
Fausti, S. A., Frey, R. H., Erickson, D., Rappaport, B. Z., Cleary, R. E., and Brummet, R. E. (**1979**). "A system for evaluating auditory function from 8000–20000 Hz," J. Acoust. Soc. Am. **66**, 1713–1718.
Fausti, S. A., Helt, W. J., Phillips, D. S., Gordon, J. S., Bratt, G. W., Sugiura, K. M., and Noffsiger, D. (**2003**). "Early detection of ototoxicity using 1/6th-octave steps," J. Am. Acad. Audiol **14**, 444–450.
Fausti, S. A., Henry, J. A., Helt, W. J., Phillips, D. S., Frey, R. H., Noffsiger, D., Larson, V. D., and Fowler, C. G. (**1999**). "An individualized, sensitive frequency range for early detection of ototoxicity," Ear Hear. **20**, 497–505.
Gilman, S., and Dirks, D. D. (**1986**). "Acoustics of ear canal measurement of eardrum SPL in simulators," J. Acoust. Soc. Am. **80**, 783–793.
Glattke, T. J., and Robinette, M. S. (**2007**). "Transient evoked otoacoustic emissions in populations with normal hearing sensitivity," in *Otoacoustic Emissions: Clinical Applications*, 3rd ed., edited by M. S. Robinette and T. J. Glattke (Thieme, New York).
Green, D. M. (**1993**). "A maximum-likelihood method for estimating thresholds in a yes-no task," J. Acoust. Soc. Am. **93**, 2096–2105.
Green, D. M., Kidd, G., Jr., and Stevens, K. N. (**1987**). "High-frequency audiometric assessment of a young adult population," J. Acoust. Soc. Am. **81**, 485–494.
Gu, X., and Green, D. M. (**1994**). "Further studies of a maximum-likelihood yes-no procedure," J. Acoust. Soc. Am. **96**, 93–101.
Hansen, P. C. (**2001**). Regularization Tools: A Matlab Package for Analysis and Solution of Discrete Ill-Posed Problems. Version 3.1 for MATLAB 6.0 written documentation and version 3.2 MATLAB toolkit code (URL: http://www2.imm.dtu.dk/~pch/). Last viewed 1/9/2009.
Hoaglin, D., Mosteller, F., and Tukey, J. W. (**1983**). *Understanding Robust and Exploratory Data Analysis* (Wiley, New York).
Huang, G. T., Rosowski, J. J., Puria, S., and Peake, W. T. (**1998**). "Noninvasive technique for estimating acoustic impedance at the tympanic membrane (TM) in ear canals of different size," Assoc. Res. Otolaryngol. Abstr. **21**, 487.
Huang, G. T., Rosowski, J. J., Puria, S., and Peake, W. T. (**2000**). "A noninvasive method for estimating acoustic admittance at the tympanic membrane," J. Acoust. Soc. Am. **108**, 1128–1146.
Hunter, L. L., Margolis, R. H., Rykken, J. R., Le, C. T., Daly, K. A., and Giebink, G. S. (**1996**). "High frequency hearing loss associated with otitis media," Ear Hear. **17**, 1–11.
Kalluri, R., and Shera, C. A. (**2007**). "Near equivalence of human click-evoked and stimulus-frequency otoacoustic emissions," J. Acoust. Soc. Am. **121**, 2097–2110.
Keefe, D. H. (**1997**). "Otoreflectance of the cochlea and middle ear," J. Acoust. Soc. Am. **102**, 2849–2859.
Keefe, D. H. (**1998**). "Double-evoked otoacoustic emissions: I. Measurement theory and nonlinear coherence," J. Acoust. Soc. Am. **103**, 3489–3498.
Keefe, D. H. (**2007**). "Influence of middle-ear function and pathology on otoacoustic emissions," in *Otoacoustic Emissions: Clinical Applications*, 3rd ed., edited by M. R. Robinette and T. J. Glattke (Thieme, New York), Chap. 7.
Keefe, D. H., and Benade, A. H. (**1980**). "Impedance measurement source and microphone proximity effects," J. Acoust. Soc. Am. **69**, 1489–1495.
Keefe, D. H., Bulen, J. C., Arehart, K. H., and Burns, E. M. (**1993**). "Ear-canal impedance and reflection coefficient in human infants and adults," J. Acoust. Soc. Am. **94**, 2617–2638.
Keefe, D. H., Ellison, J. C., Fitzpatrick, D. F., Jesteadt, W., and Schairer, K. S. (**2009**). "Is temporal overshoot present in stimulus-frequency otoacoustic emission responses to tone bursts in noise?," J. Acoust. Soc. Am., to be published.
Keefe, D. H., and Ling, R. (**1998**). "Double-evoked otoacoustic emissions: II. Intermittent noise rejection, calibration and ear-canal measurements," J. Acoust. Soc. Am. **103**, 3499–3508.
Keefe, D. H., and Simmons, J. L. (**2003**). "Energy transmittance predicts conductive hearing loss in older children and adults," J. Acoust. Soc. Am. **114**, 3217–3238.
Kemp, D. T. (**1978**). "Stimulated acoustic emissions from within the human auditory system," J. Acoust. Soc. Am. **64**, 1386–1391.
Kemp, D. T., and Chum, R. A. (**1980**). "Observations on the generator mechanism of stimulus frequency acoustic emissions—Two tone suppression," in *Physiological Basis and Psychophysics*, edited by R. Klinke and R. Hartman (Springer, Berlin).
Kemp, D. T., Ryan, S., and Bray, P. (**1990**). "A guide to the effective use of otoacoustic emissions," Ear Hear. **11**, 93–105.
Komune, S., Asakuma, S., and Snow, J. B., Jr. (**1981**). "Pathophysiology of the ototoxicity of cis-diamminedichloroplatinum," Otolaryngol.-Head Neck Surg. **89**, 275–282.
Konishi, T., Gupta, B. N., and Prazma, J. (**1983**). "Ototoxicity of cis-dichlorodiammine platinum (II) in guinea pigs," Am. J. Otolaryngol. **4**, 18–26.
Konrad-Martin, D., and Keefe, D. H. (**2005**). "Transient-evoked stimulus-frequency and distortion-product otoacoustic emssions in normal and impaired ears," J. Acoust. Soc. Am. **117**, 3799–3815.
Kruger, B., and Rubin, R. J. (**1987**). "The acoustic properties of the infant ear," Acta Oto-Laryngol. **103**, 578–585.
Kuronen, P., Sorri, M. J., Pääkkönen, R., and Muhli, A. (**2003**). "Temporary threshold shift in military pilots measured using conventional and extended high-frequency audiometry after one flight," Int. J. Audiol. **42**, 29–33.
Lee, F. S., Matthews, L. J., Dubno, J. R., and Mills, J. H. (**2005**). "Longitudinal study of pure-tone thresholds in older persons," Ear Hear. **26**, 1–11.
Lin, T., and Guinan, J. J. (**2000**) "Auditory-nerve-fiber responses to high-level clicks: interference patterns indicate that excitation is due to the combination of multiple drives," J. Acoust. Soc. Am. **107**, 2615–2630.
Margolis, R. H., Saly, G. L., and Hunter, L. L. (**2000**). "High-frequency hearing loss and wideband middle ear impedance in children with otitis media histories," Ear Hear. **21**, 206–211.
Matthews, L. J., Lee, F. S., Mills, J. H., and Dubno, J. R. (**1997**). "Extended high-frequency thresholds in older adults," J. Speech Lang. Hear. Res. **40**, 208–214.
Moore, B. C. J., and Glasberg, B. R. (**2003**). "Behavioural measurement of level-dependent shifts in the vibration pattern on the basilar membrane at 1 and 2 kHz," Hear. Res. **175**, 66–74.
Mulheran, M., and Degg, C. (**1997**). "Comparison of distortion product OAE generation between a patient group requiring frequent gentamicin therapy and control subjects," Br. J. Audiol. **31**, 5–9.
Nakai, Y., Konishi, K., Chang, K. C., Ohashi, K., Morisaki, N., Minowa, Y., and Morimoto, A. (**1982**). "Ototoxicity of the anticancer drug cisplatin. An experimental study," Acta Oto-Laryngol. **93**, 227–232.
Neely, S. T., and Gorga, M. P. (**1998**). "Comparisons between intensity and pressure as measures of sound level in the ear canal," J. Acoust. Soc. Am. **104**, 2925–2934.
Oppenheim, A. V., and Schafer, R. W. (**1989**). *Discrete-Time Signal Processing* (Prentice-Hall, Englewood Cliffs, NJ).
Pierce, A. D. (**1989**). *Acoustics: An Introduction to Its Physical Principles and Applications* (Acoustical Society of America, Woodbury).
Prieve, B. A., Gorga, M. P., and Neely, S. T. (**1996**). "Click- and tone-burst evoked otoacoustic emissions in normal-hearing and hearing-impaired ears," J. Acoust. Soc. Am. **99**, 3077–3086.
Probst, R., Lonsbury-Martin, B. L., and Martin, G. K. (**1991**). "A review of otoacoustic emissions," J. Acoust. Soc. Am. **89**, 2027–2067.
Puria, S. (**2003**). "Measurements of human middle ear forward and reverse

acoustics: implications for otoacoustic emissions," J. Acoust. Soc. Am. **113**, 2773–2789.
Puria, S., Peake, W. T., and Rosowski, J. J. (**1997**). "Sound-pressure measurements in the cochlear vestibule of human-cadaver ears," J. Acoust. Soc. Am. **101**, 2754–2770.
Rabinowitz, W. M. (**1981**). "Measurement of the acoustic input immittance of the human ear," J. Acoust. Soc. Am. **89**, 2379–2390.
Recio, A., and Rhode, W. S. (**2000**) "Basilar membrane response to broadband stimuli," J. Acoust. Soc. Am. **108**, 2281–2298.
Recio, A., Rich, N. C., Narayan, S. S., and Ruggero, M. A. (**1998**). "Basilar-membrane responses to clicks at the base of the chinchilla cochlea," J. Acoust. Soc. Am. **103**, 1972–1989.
Ress, B. D., Sridhar, K. S., Balkany, T. J., Waxman, G. M., Stagner, B. B., and Lonsbury-Martin, B. L. (**1999**). "Effects of cis-platinum chemotherapy on otoacoustic emissions: The development of an objective screening protocol," Otolaryngol.-Head Neck Surg. **121**, 693–701.
Schairer, K. S., Ellison, J. C., Fitzpatrick, D. F., and Keefe, D. H. (**2006**). "Use of stimulus-frequency otoacoustic emission latency and level to investigate cochlear and middle-ear mechanics in human ears," J. Acoust. Soc. Am. **120**, 901–914.
Schairer, K. S., Fitzpatrick, D. F., and Keefe, D. H. (**2003**). "Input-output functions for stimulus-frequency otoacoustic emissions in normal-hearing adult ears," J. Acoust. Soc. Am. **114**, 944–966.
Scheperle, R. A., Neely, S. T., Kopun, J. G., and Gorga, M. P. (**2008**). "Influence of in situ, sound-level calibration on distortion-product otoacoustic emission variability," J. Acoust. Soc. Am. **124**, 288–300.
Schweitzer, V. G., Hawkins, J. E., Lilly, D. J., Litterst, C. J., Abrams, G., Davis, J. A., and Christy, M. (**1984**). "Ototoxic and nephrotoxic effects of combined treatment with cis-diamminedichloroplatinum and kanamycin in the guinea pig," Otolaryngol.-Head Neck Surg. **92**, 38–49.
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., Guinan, J. J., Jr., and Oxenham, A. J. (**2002**). "Revised estimates of human cochlear tuning from otoacoustic and behavioral measurements," Proc. Natl. Acad. Sci. U.S.A. **99**, 3318–3323.
Shera, C. A., Tubis, A., Talmadge, C. L., de Boer, E., Fahe, P. F., and Guinan, J. J. (**2007**). "Allen-Fahey and related experiments support the predominance of cochlear slow-wave otoacoustic emissions," J. Acoust. Soc. Am. **121**, 1564–1575.
Siegel, J. H. (**1994**). "Ear-canal standing waves and high-frequency sound calibration using otoacoustic emission probes," J. Acoust. Soc. Am. **95**, 2589–2597.
Siegel, J. H. (**2007**). "Calibrating otoacoustic emission probes," *Otoacoustic Emissions: Clinical Applications*, 3rd ed., edited by M. S. Robinette and T. J. Glattke (Thieme Medical, New York), pp. 403–427.
Smurzynski, J., and Kim, D. O. (**1992**). "Distortion-product and click-evoked otoacoustic emissions of normally-hearing adults," Hear. Res. **58**, 227–240.
Stavroulaki, P., Apostolopoulos, N., Segas, J., Tsakanikos, M., and Adamopoulos, G. (**2001**). "Evoked otoacoustic emissions—An approach for monitoring cisplatin induced ototoxicity in children," Int. J. Pediatr. Otorhinolaryngol. **59**, 47–57.
Stavroulaki, P., Vossinakis, I. C., Dinopoulou, D., Doudounakis, S., Adamopoulos, G., and Apostolopoulos, N. (**2002**). "Otoacoustic emissions for monitoring aminoglycoside-induced ototoxicity in children with cystic fibrosis," Arch. Otolaryngol. Head Neck Surg. **128**, 150–155.
Stelmachowicz, P. G., Beauchaine, K. A., Kalberer, A., and Jesteadt, W. (**1989a**). "Normative thresholds in the 8-to 20-kHz range as a function of age," J. Acoust. Soc. Am. **86**, 1384–1391.
Stelmachowicz, P. G., Beauchaine, K. A., Kalberer, A., Kelly, W. J., and Jesteadt, W. (**1989b**). "High-frequency audiometry: Test reliability and procedural considerations," J. Acoust. Soc. Am. **85**, 879–887.
Stelmachowicz, P. G., Beauchaine, K. A., Kalberer, A., Langer, T., and Jesteadt, W. (**1988**). "The reliability of auditory thresholds in the 8-to 2-kHz range using a prototype audiometer," J. Acoust. Soc. Am. **83**, 1528–1535.
Stevens, K. N., Berkovitz, R., Kidd, G., Jr., Green, D. M. (**1987**). "Calibration of ear canals for audiometry at high frequencies," J. Acoust. Soc. Am. **81**, 470–484.
Stinson, M. R., and Lawton, B. W. (**1989**). "Specification of the geometry of the human ear canal for the prediction of sound-pressure level distribution," J. Acoust. Soc. Am. **85**, 2492–2503.
Stinson, M. R., Shaw, E. A. G., and Lawton, B. W. (**1982**). "Estimation of acoustical energy reflectance at the eardrum from measurements of pressure distribution in the human ear canal," J. Acoust. Soc. Am. **72**, 766–773.
Tange, R. A., Dreschler, W. A., and van der Hulst, R. J. (**1985**). "The importance of high-tone audiometry in monitoring for ototoxicity," Arch. Oto-Rhino-Laryngol. **242**, 77–81.
Tognola, G., Grandori, F., and Ravazzani, P. (**1997**). "Time-frequency distributions of click-evoked otoacoustic emissions," Hear. Res. **106**, 112–122.
van der Hulst, R. J., Dreschler, W. A., and Urbanus, N. A. (**1988**). "High frequency audiometry in prospective clinical research of ototoxicity due to platinum derivatives," Ann. Otol. Rhinol. Laryngol. **97**, 133–137.
Whitehead, M. L., Jimenez, A. M., Stagner, B. B., McCoy, M. J., Lonsbury-Martin, B. L., and Martin, G. K. (**1995a**). "Time windowing of click-evoked otoacoustic emissions to increase signal-to-noise ratio," Ear Hear. **16**, 599–611.
Whitehead, M. L., Stagner, B. B., Lonsbury-Martin, B. L., and Martin, G. K. (**1995b**). "Effects of ear-canal standing waves on measurements of distortion-product otoacoustic emissions," J. Acoust. Soc. Am. **98**, 3200–3214.
Withnell, R. H., Hazlewood, C., and Knowlton, A. (**2008**). "Reconciling the origin of the transient evoked otoacoustic emission in humans," J. Acoust. Soc. Am. **123**, 212–221.
Young, R. W. (**1970**). "On the energy transported with a sound pulse," J. Acoust. Soc. Am. **47**, 441–442.
Zemplenyi, J., Gilman, S., and Dirks, D. (**1985**). "Optical method for measurement of ear canal length," J. Acoust. Soc. Am. **78**, 2146–2148.
Zweig, G., and Shera, C. A. (**1995**). "The origin of periodicity in the spectrum of evoked otoacoustic emissions," J. Acoust. Soc. Am. **98**, 2018–2047.
Zwicker, E., and Schloth, E. (**1984**). "Interrelation of different otoacoustic emissions," J. Acoust. Soc. Am. **75**, 1148–1154.

A functional-magnetic-resonance-imaging investigation of cortical activation from moving vibrotactile stimuli on the fingertip

Ian R. Summers[a)]
Biomedical Physics Group, School of Physics, University of Exeter, Exeter EX4 4QL, United Kingdom

Susan T. Francis and Richard W. Bowtell
Magnetic Resonance Centre, School of Physics and Astronomy, University of Nottingham, NG7 2RD, United Kingdom

Francis P. McGlone
Magnetic Resonance Centre, School of Physics and Astronomy, University of Nottingham, NG7 2RD, United Kingdom and Department of Neurological Sciences, School of Medicine, Liverpool University, L69 3GE, United Kingdom

Matthew Clemence
Biomedical Physics Group, School of Physics, University of Exeter, Exeter EX4 4QL, United Kingdom

(Received 18 June 2008; revised 18 November 2008; accepted 26 November 2008)

Using a 100-element tactile stimulator on the fingertip during functional-magnetic-resonance imaging, brain areas were identified that were selectively activated by a moving vibrotactile stimulus (the sensation of a moving line being dragged over the fingertip). Activation patterns elicited by tactile motion, contrasted to an equivalent stationary stimulus, were compared in six human subjects with those generated by a moving visual stimulus, contrasted to an equivalent stationary stimulus. Results provide further evidence for a neuroanatomical convergence of tactile-motion processing and visual-motion processing in humans. The sites of this convergence are found to lie in the middle temporal complex (hMT+/V5), an area with known specialization for visual-motion processing, and in the intraparietal area of the posterior parietal cortex. In an advance on previous studies, the present study includes separate delineation of activations for moving tactile stimuli and activations for moving visual stimuli. Results suggest that the two sets of activations are not entirely collocated. Compared to the visual-motion activations, the tactile-motion activations are found to lie nearer the midline of the brain and further superior.
 [DOI: 10.1121/1.3056399]

PACS number(s): 43.64.Vm, 43.66.Wv [WPS] Pages: 1033–1039

I. INTRODUCTION

In many perceptual operations, sensations from two or more sensory modalities combine to produce a unified percept. Information must be allocated between and within modalities to ensure that the appropriate task-relevant information is passed on to decision-making and behavioral-control systems. In the particular case of sensing object motion, the perceptual operation may involve sight, hearing, and touch. However, it is not well understood how and where the cross-modal interaction of motion systems for different modalities takes place. In this study, functional magnetic resonance imaging (fMRI) has been used to investigate the cortical regions associated with tactile-motion processing and their relation to the cortical regions associated with visual-motion processing. The fMRI technique has been widely used in previous studies of auditory perception (e.g., Mohr *et al.*, 1999; Whalen *et al.*, 2006) but its application to previous studies of tactile perception has been more limited, partly due to the technical difficulties of providing controlled tactile stimulation in the high-magnetic-field environment of a magnetic resonance imager.

Relatively little is known of the pathways involved in tactile-motion processing. In contrast, there has been much investigation into visual-motion perception. In an early visual study, Dubner and Zeki (1971) identified a motion-sensitive region in the middle temporal/V5 (MT/V5) area in the macaque. This region and adjacent motion-sensitive regions, including the medial superior temporal region (MST), may be collectively labeled as the middle temporal complex (MT+/V5). An equivalent area hMT+/V5 has been identified in humans using positron-emission tomography (PET) (Watson *et al.*, 1993; Zeki *et al.*, 1991) and fMRI (Tootell *et al.*, 1995). In a study of postmortem brains, Malikovic *et al.* (2007) recently identified a structural correlate of hMT+/V5. A fMRI study by Poirier *et al.* (2005) indicates that the hMT+/V5 area is also associated with auditory motion processing, although Lewis *et al.* (2000) observed negative activation in hMT+/V5 during an auditory motion discrimination task.

[a)]Electronic mail: i.r.summers@exeter.ac.uk

Correspondence between motion perception in the visual and tactile modalities has been investigated by Hagen *et al.* (2002) using PET. Their results suggest that the visual area hMT+/V5 is also involved in tactile-motion processing, Similarly, in a fMRI study involving visual and tactile detections of the motion of a rotating sphere, Blake *et al.* (2004) observed activation in hMT+/V5 in response to both visual and tactile motion. Vanello *et al.* (2004) observed activation in hMT+/V5 in response to both a moving visual stimulus (gray dots moving on a black background) and a moving tactile stimulus (Braille-like dots moving across the fingertips). Ricciardi *et al.* (2007), using similar stimuli, reported fMRI activations in the region of hMT+/V5 but with different localizations in the visual-motion and tactile-motion cases. However, in that experiment, the identification of tactile-motion regions may be compromised—the stationary and moving tactile stimuli (embossed dots on a plastic surface) are expected to produce different levels of excitation in the skin mechanoreceptors, because of the nature of the mechanoreceptor response; hence some differences in activation between stationary and moving stimuli may be associated with differences in perceived stimulus intensity rather than stimulus movement.

In contrast, Bodegård *et al.* (2000) reported a PET study in which the principal tactile-motion activation was observed in the primary somatosensory area. In a similar PET study (Burton *et al.*, 1999), activation of both primary and secondary somatosensory areas (SI and SII) was observed. In neither of these two PET studies was activation of visual-motion areas observed in response to tactile motion—this may be related to the nature of the moving tactile stimuli, which in both cases continually contacted the same small area of skin with a brushing/rubbing action (i.e., there was no coherent variation of the stimulus location on the skin).

Beauchamp *et al.* (2007), using *non-moving* tactile stimuli on the palms of the hands or the soles of the feet, observed activation in the MST area of hMT+/V5 but not in the MT area. They also suggested that the hMT+/V5 activation reported by Ricciardi *et al.* (2007) may be confined to the MST area.

Multimodal motion processing of visual, tactile, and auditory motion has been investigated using fMRI (Bremmer *et al.*, 2001). Activation was observed in hMT+/V5 for visual motion, but not for auditory or tactile motion. However, the moving tactile stimulus used in that study—air flowing continuously across the subject's forehead—was far from ideal.

II. METHODS

In the present study a vibrotactile stimulator—an array of 100 contactors over the fingertip—is used to produce spatiotemporal patterns of mechanical disturbance on the fingertip during fMRI at a magnetic field strength of 3.0 T. In addition, a visual-motion condition is included, so as to allow the observed locations of tactile-motion centers to be compared with the locations of visual-motion centers. A simple visual-motion stimulus—a bar pattern moving coherently in the vertical direction—allows reliable location of the visual-motion centers.

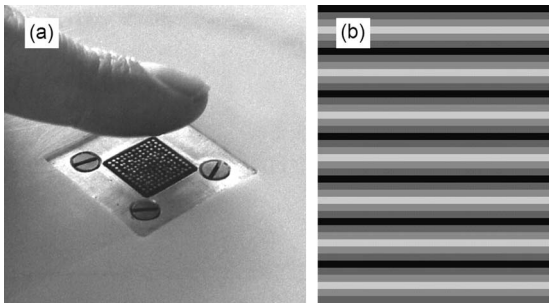


FIG. 1. (a) The tactile stimulator with 100 independently controlled contactors over an area of 1 cm^2; (b) the pattern used for the visual stimulus with seven cycles of a sinusoidal intensity variation (each cycle subtending a visual angle of approximately 3°).

A. Subjects

Six healthy volunteer subjects (all male, aged 20–33 years) took part in the study. All were right handed and all gave informed consent. The study conformed with the Code of Ethics of the World Medical Association (Declaration of Helsinki) and was approved by the relevant local Ethics Committee.

B. Tactile and visual stimuli

A range of tactile stimulators has been employed in fMRI studies (e.g., Francis *et al.*, 2000; Deuchert *et al.*, 2002; Briggs *et al.*, 2004). Most of these devices are designed to provide tactile stimulation at a fixed location on the skin, or quasistatic patterns of stimulation at one of a number of skin sites (Zappe *et al.*, 2004; Overduin and Servos, 2004; Dresela *et al.*, 2008). In order to produce the sensation of tactile movement over the skin, a device that provides sequential stimulation over a distribution of sites is required.

The tactile stimulator used in the present study was originally designed as a psychophysical tool, to simulate sensations of "natural" touch on the fingertip by generating appropriate patterns of mechanical disturbance at the skin surface. The stimulator (Summers and Chanter, 2002) has an array of 100 circular contactors arranged on a 1 mm square matrix (ten rows and ten columns) over an area of 1×1 cm^2 [see Fig. 1(a)]. The contactors are each driven by a piezoelectric bimorph cantilever, under individual software control, with a working bandwidth of 25–400 Hz. Each contactor has a diameter of 0.6 mm; vibrotactile stimulation is produced by movement in a direction normal to the skin surface. [A more recent version of this device in a virtual-reality application is described by Magnenat-Thalmann *et al.* (2007).] To allow compatibility with fMRI experiments, the device was mounted in a nonmagnetic box and modified to allow remote driving. For the tactile experiments in the present study, the stimulator was placed in contact with the fingertip of digit 2 (right hand).

The tactile fMRI experiment contrasts moving and stationary stimuli. Tactile stimulus generation is based on vibra-

tion of the ten contactors in a particular row of the array, at a frequency of 40 Hz and a peak-to-peak amplitude of approximately 50 μm. This produces the sensation of a "line" on the fingertip at a comfortable sensation level. For the moving tactile stimulus, a transit over the fingertip is produced by sequentially activating the ten rows of the array, starting from row 1 and ending with row 10. Each row is activated (providing 40 Hz vibration) for 100 ms, giving a total transit time of 1000 ms. This produces the sensation of a moving line being dragged over the fingertip, from the tip to the base. The complete moving stimulus comprises seven transits, separated by 100 ms periods of no stimulation, giving an overall stimulus duration of 7.6 s. For the stationary tactile stimulus, the sensation of a static line on the fingertip is produced by activating a single row of the array (providing 40 Hz vibration) for 1000 ms. The complete stationary stimulus comprises seven repetitions of this 1000 ms event, separated by 100 ms periods of no stimulation, again giving an overall stimulus duration of 7.6 s. The row selected for activation remains constant over the seven repetitions within a single stimulus, but is randomly varied (over all ten rows) across multiple stimulus cycles within the overall measurement period. In this way, over the whole experiment, the fingertip area that is stimulated by the stationary stimulus is the same as that stimulated by the moving stimulus. The overall subjective intensity of the moving and stationary stimuli is also approximately matched. (The intention is to produce differences in activation that are associated with stimulus movement while avoiding differences in activation that are associated with stimulus intensity. In the event, it is unlikely that the observed activation differences are entirely motion related. However, the stimuli used in the present experiment are expected to offer a significant advantage over stimuli produced by a real surface on the fingertip—in that case, as mentioned above, moving and stationary stimuli are likely to differ markedly in intensity.)

The visual stimulus is based on a sinusoidal variation of intensity—seven cycles over a rectangular area within the subject's field of view, as shown in Fig. 1(b). For the moving visual stimulus, the sinusoidal variation moves as a wave in the vertical direction, so that the total pattern traverses the field of view completely in 8.4 s, which is the overall stimulus duration. At the eye, this corresponds to an angular speed of around 2.5° per second. For the stationary visual stimulus, the sinusoidal variation is presented as a static pattern, again with a duration of 8.4 s.

C. Procedure

Tactile and visual fMRI experiments were performed on each subject in a single imaging session. The first investigation performed was always the tactile, followed by the visual. This order was chosen to avoid imagined motion in the tactile investigation, deriving from the visual experiment. The subjects were unaware of the object of the experiment.

Within both the tactile and visual investigations, stimuli were delivered in a block design. The tactile investigation (duration of 21 min) comprised 40 stimulus cycles, each including 7.6 s of tactile stimulation ("on" period) followed by 24 s rest ("off" period). The cycles alternated between presentation of the moving stimulus and presentation of the stationary stimulus. The subject was asked to visually fixate while attending to the tactile stimulus. The visual investigation (duration of 4 min) comprised 8 cycles, each including 8.4 s of visual stimulation (on period) followed by 20 s rest (off period). The cycles alternated between presentation of the moving stimulus and presentation of the stationary stimulus.

During each fMRI investigation, images of 12 coronal slices within the subject's brain were acquired every 2 s using echoplanar imaging (EPI). These slices (field of view 192×384 mm^2, in-plane resolution 3×3 mm^2, and slice thickness 8 mm) were positioned to cover the posterior brain, including the visual and somatosensory areas. The subject's head was stabilized throughout using inflatable pillows to minimize any movement. Following the fMRI measurements, EPI and inversion-recovery EPI data sets were acquired at higher resolution (64 slices covering the whole brain, 3 mm isotropic resolution) to aid normalization of the fMRI data sets to a standard brain template (see below).

D. fMRI data analysis

Cortical activity is identified from changes in image grayscale values, which result from the blood-oxygen-level-dependent effect in the EPI images. The fMRI data sets were processed using statistical parametric mapping software (version SPM2, Wellcome Department of Imaging Neuroscience, University of London, UK). To minimize the effect of subject movement during the measurement, each data set was realigned to images acquired at the half-way point of the measurement period. The data from each experiment were then coregistered to the corresponding higher resolution data set (64-slice whole brain), and the higher resolution data set was normalized (Friston *et al.*, 1995) to fit a standard brain template (ICBM152 template; McGill University, Montreal, Canada). The coregistration and normalization parameters were subsequently used to spatially normalize the fMRI data to fit the standard template, which has a voxel size of $2 \times 2 \times 2$ mm^3.

Prior to a group analysis (see below), individual-subject analysis was performed on the tactile and visual data sets, using an anatomical marker for hMT+/V5 to ensure that motion areas were correctly attributed and not lost at the group level (Dumoulin *et al.*, 2000; Hagen *et al.*, 2002; Poirier *et al.*, 2005). For the individual analysis, the data were spatially smoothed with a Gaussian kernel (full width at half maximum of 5 mm). For the data from the tactile investigations, statistical parametric maps (based on the *t* statistic) were formed for each individual to identify areas that were (i) more activated by the stationary tactile stimulus than by the off (no stimulus) condition, (ii) more activated by the moving tactile stimulus than by the off condition, and (iii) more activated by the moving tactile stimulus than by the stationary tactile stimulus. (In the statistical analysis, the activation was modeled by convolving the timecourse of stimulus "on" periods with a canonical hemodynamic response function.) A similar procedure was applied to the data from

TABLE I. Group analysis: regions activated more by the moving tactile stimulus than by the 'off' (no stimulus) condition; all regions with a corrected probability $p<0.05$ are tabulated.

Region	Most activated voxel in cluster (Talairach: x y z)	Cluster size k [$\times(2\ \text{mm})^3$]	Corrected probability p
SI (left)	−56 −22 46	123	<0.001
SII (left)	−46 −22 18	1249	<0.001
SII (right)	52 −34 24	86	0.002
SII/insula (right)	54 −16 22	28	0.004
hMT+/V (left)	−32 −70 2	83	0.003
hMT+/V5 (right)	40 −64 0	40	0.005
Intraparietal (left)	−30 −48 42	237	<0.001
Intraparietal (right)	46 −52 44	156	0.003
Cerebellum (right)	26 −68 −34	302	0.003

TABLE II. Group analysis: regions activated more by the moving tactile stimulus than by the stationary tactile stimulus; all regions with a corrected probability $p<0.05$ are tabulated.

Region	Most activated voxel in cluster (Talairach: x y z)	Cluster size k [$\times(2\ \text{mm})^3$]	Corrected probability p
SI (left)	−58 −22 42	63	0.035
SII (left)	−47 −21 14	78	0.002
SII (right)	53 −38 25	4	0.047
hMT+/V5 (left)	−31 −70 2	45	0.004
hMT+/V5 (right)	41 −64 0	35	0.008
Intraparietal (left)	−30 −50 45	95	0.012
Intraparietal (right)	42 −51 49	83	0.014

the visual investigations. Activated regions were identified on the basis of a significance level of $p<0.05$, corrected for familywise error. To identify areas common to both visual- and tactile-motion processing, a conjunction analysis (Price and Friston, 1997) was performed, again thresholded at a corrected probability of $p<0.05$.

A fixed-effects group analysis was then performed. For this the data were spatially smoothed with a 10 mm Gaussian kernel to account for individual-subject cortical variability (in place of the 5 mm kernel used when processing the data on a single-subject basis). The same contrasts were investigated as in the individual-subject analysis described above, including testing for the main effects of tactile movement and visual movement, as well as application to the group data of the conjunction analysis for tactile movement and visual movement. A conservative small volume correction (Worsley *et al.*, 1996) was applied, using a mask of all regions showing a response to both the moving and the stationary condition at $p<0.001$. The p values were corrected for false discovery rate (Genovese *et al.*, 2002) and regions with a corrected $p<0.05$ were tabulated. In the tables (see Sec. III) positions in the brain are reported in terms of Talairach coordinates [(x, y, z) coordinates in mm], converted from the MNI coordinates of the ICBM152 template (2 mm resolution) using the mni2tal script written by Slotnick (Department of Psychology, Boston College, Boston, MA).

III. RESULTS

Table I (group data) lists regions activated more by the moving tactile stimulus than by the off condition. Table II (group data) lists regions activated more by the moving tactile stimulus than by the stationary tactile stimulus; these regions are also shown in Fig. 2(a). (Data are not presented from the analysis of regions activated more by the stationary tactile stimulus than by the off condition.) Both the moving and stationary tactile stimuli produce significant activation of the primary somatosensory cortex SI (contralaterally, i.e., the right-side stimulation produces left-side activation), the secondary somatosensory cortex SII (bilaterally), and the intraparietal area of the posterior parietal cortex. This is in agreement with previous studies of vibrotactile stimulation (Francis *et al.*, 2000; McGlone *et al.*, 2002). Compared to the stationary tactile stimulus, the moving tactile stimulus is found to generate significantly increased activity in SI (contralaterally), SII (bilaterally), in the vicinity of hMT+/V5 (bilaterally), and in the intraparietal area (bilaterally).

Table III group data) lists regions activated more by the moving visual stimulus than by the stationary visual stimulus; these regions are also shown in Fig. 2(b). (Data are not

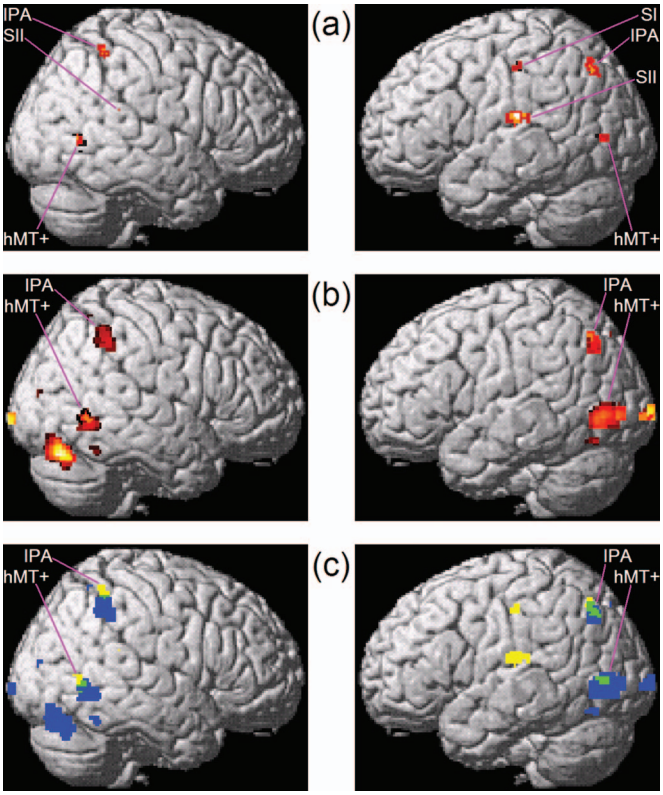


FIG. 2. (a) False-color (thermal scale) map of the group statistic for moving tactile stimulus vs stationary tactile stimulus; (b) false-color (thermal scale) map of the group statistic for moving visual stimulus vs stationary visual stimulus; (c) map of regions activated more by the moving tactile stimulus than by the stationary tactile stimulus [taken from (a), shown in yellow] and regions activated more by the moving visual stimulus than by the stationary visual stimulus [taken from (b), shown in blue]; the four areas common to both sets of activations are shown in green—two lower in the brain (bilateral hMT+/V5, labeled as hMT+) and two higher in the brain (bilateral intraparietal area, labeled as IPA). In all cases the regions are defined by a threshold corresponding to a corrected probability $p<0.05$; they are rendered onto a representation of the standard brain template from the SPM2 software.

TABLE III. Group analysis: regions activated more by the moving visual stimulus than by the stationary visual stimulus; all regions with a corrected probability $p < 0.05$ are tabulated.

Region	Most activated voxel in cluster (Talairach: *x y z*)	Cluster size *k* [×(2 mm)^3]	Corrected probability *p*
V1 (bilateral)	4 −87 2	456	<0.001
V3/V3a (left)	−26 −72 36	43	0.006
V3/V3a (right)	42 −88 14	29	0.007
hMT+/V5 (left)	−56 −72 −12	76	0.005
hMT+/V5 (right)	55 −52 −10	48	0.004
Intraparietal (left)	−46 −58 39	167	0.003
Intraparietal (right)	58 −44 42	224	0.002
Middle occipital gyrus (left)	−48 −70 −16	17	0.008
Middle occipital gyrus (right)	44 −76 −24	296	<0.001
Middle occipital gyrus (right)	20 −84 −22	2	0.034
Parietal (superior, left)	−12 −76 50	96	0.006
Parietal (inferior, right)	38 −58 60	3	0.049

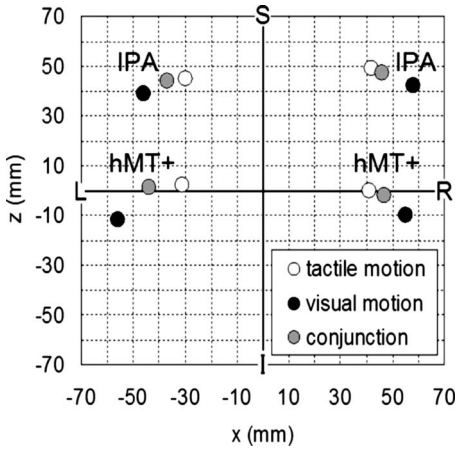


FIG. 3. Positions of the activation centers associated with moving tactile stimulation and with moving visual stimulation (group analysis), projected onto a coronal plane (i.e., showing the *x* and *z* Talairach coordinates). The activation centers from the group conjunction analysis, associated with both moving tactile stimulation and moving visual stimulation, are also shown. Centers in the hMT+/V5 complex are labeled hMT+ and centers in the intraparietal area are labeled IPA. The axis labels indicate the inferior-superior (I-S) and left-right (L-R) directions.

presented from the analysis of regions activated more by the stationary visual stimulus than by the off condition, or from the analysis of regions activated more by the moving visual stimulus than by the off condition.) The moving/stationary-visual contrast reveals significantly increased activity in regions including the primary visual area V1 (bilaterally), the tertiary visual area V3/V3a (bilaterally), an area with coordinates corresponding to hMT+/V5 (bilaterally), and the intraparietal area of the posterior parietal cortex (bilaterally).

Figure 2(c) shows the moving versus stationary contrast for tactile stimuli [in yellow: regions taken from Fig. 2(a) data], and the moving versus stationary contrast for visual stimuli [in blue: regions taken from Fig. 2(b) data], with common areas shown in green. There is a significant overlap between the visual and tactile areas. However, compared to the regions of moving-visual activation (Table III), the regions of moving-tactile activation (Table II) appear to be nearer the midline and further superior—some voxels activated by the moving tactile stimulus fall outside the clusters associated with moving visual stimulation, and some voxels activated by the moving visual stimulus fall outside the clusters associated with moving tactile stimulation. Figure 3 shows the positions of the activation centers, projected onto a coronal plane. These observations are confirmed by results from the group conjunction analysis: regions in hMT+/V5 [activation centers with Talairach coordinates (−44, −71, 1) and (47, −60, −2)] and in the intraparietal area [activation centers with Talairach coordinates (−37, −53, 44) and (46, −47, 47)] are found to be associated with both tactile motion and visual motion; these activation centers are also shown in Fig. 3. The cluster sizes from the group conjunction analysis for activations in hMT+/V5 (left, right) and the intraparietal area (left, right) are 26, 12, 76, and 12, respectively [units of (2 mm)^3], representing 58%, 34%, 80%, and 14% of the volume of the corresponding tactile-motion clusters (see Table II) and 34%, 25%, 46%, and 5% of the volume of the corresponding visual-motion clusters (see Table III). In the conjunction analyses for individual subjects, all but one show results in agreement with the results of the group conjunction analysis, i.e., some overlap of tactile and visual motion-sensitive areas in hMT+/V5 and the intraparietal area.

IV. DISCUSSION AND CONCLUSION

This study confirms that there are brain areas in common for motion processing in the visual and tactile modalities. These are found to be located in the middle temporal complex, hMT+/V5, and in the intraparietal area of the posterior parietal cortex [in an area thought to be the human equivalent of the ventral intraparietal area VIP, which has been identified in the monkey as sensitive to visual motion (Colby *et al.*, 1993; Cook and Maunsell, 2002)]. The hMT+/V5 area has been considered to be primarily associated with visual-motion processing but, in the present study, activation is observed for tactile motion in the absence of visual motion. This finding supports that of the PET study of Hagen *et al.*, (2002), which demonstrated hMT+/V5 activation due to a moving tactile stimulus produced by stroking the forearm. On the basis of the results of Beauchamp *et al.*, (2007) it might be expected that the present study would show differences in activation between the MT and MST regions of hMT+/V5, which could be distinguished experimentally (Dukelow *et al.*, 2001) on the basis of their *y* coordinates in the Talairach system (anterior/posterior direction). However, in the present experiment, the spatial accuracy of the data in this direction is lower than in the other two directions, and not sufficient to resolve these regions within hMT+/V5.

The results of the present study and Hagen *et al.* (2002), together with the results of Blake *et al.* (2004) and Vanello *et al.* (2004) for tactile motion, and Poirier *et al.* (2005) for auditory motion, suggest that the specialized processing operations available in hMT+/V5 for visual motion are also

available for both tactile and auditory modalities. As discussed by Hagen *et al.* (2002), hMT+/V5 appears to be a unimodal area, with direct sensory input from the visual modality, but not from the somatosensory modality. However, activation of hMT+/V5 by somatosensory input may be explained by links between hMT+/V5 and multimodal centers in the intraparietal area (Driver and Spence, 2000).

In the PET study by Hagen *et al.* (2002), with the limited spatial resolution of that imaging modality, it was not possible to establish if there was a unimodal tactile-motion-only region within the hMT+/V5 complex. In the fMRI study by Blake *et al.* (2004), the methodology did not involve separate delineation of the tactile-motion and visual-motion regions. However, in an advance on these previous studies, the present fMRI study includes separate delineation of these regions. The observed spatial offset between tactile-motion and visual-motion regions may be interpreted as evidence for spatial organization of these activations according to modality, thus providing further evidence for a specific tactile-motion-processing capability within the hMT+/V5 complex. [As mentioned above, Ricciardi *et al.* (2007) also reported fMRI activations in the region of hMT+/V5 with different locations for the visual-motion and tactile-motion cases, although their tactile-motion regions may include intensity-related activations—the tactile-motion regions identified in the present study are less extensive.] However, there is also a possibility that, within hMT+/V5 and the intraparietal area, activations may be spatially organized according to the nature of the movement—location, speed, direction, etc.—so by a suitable choice of tactile and visual stimuli it might be possible to achieve a collocation of activity, in which case the observed mismatch of cortical locations in the present study might result from mismatched stimuli rather than from a modality effect. There is some evidence in literature (e.g., Dukelow *et al.*, 2001) for such spatial organization in the case of visual motion and, although literature is lacking in comparable results from tactile studies, this alternative interpretation of the results from the present study cannot be ruled out.

The results of the present study may also be compared to those obtained by Bremmer *et al.* (2001), who located three brain areas associated with multimodal (visual, auditory, and tactile) motion processing: in the intraparietal area, the ventral premotor cortex, and the lateral inferior postcentral cortex. In the present study bilateral activation has been observed at a similar location in the intraparietal area.

As already stated, results from the present study and previous studies by Hagen *et al.* (2002), Blake *et al.* (2004), and Vanello *et al.* (2004) suggest that some brain areas associated with visual-motion processing are also involved with tactile-motion processing. Because these studies involved only sighted subjects, a possible confound is that a moving tactile stimulus may provoke reflex visual imagery of that stimulus. However, this issue has been addressed by Blake *et al.* (2004) who, for imagined visual motion, did not find the significant activation in the hMT+/V5 complex that they observed for tactile motion. Similarly, in relation to the study by Vanello *et al.* (2004), subsequent results obtained by the same research group (Ricciardi *et al.*, 2007) suggest that the activations in the tactile case were not due to visual imagery: Similar bilateral activation in the hMT+/V5 complex and the intraparietal area was observed in congenitally blind subjects for the moving tactile stimulus. In addition, results from another investigation support the interpretation that the observed response to tactile stimuli relates to processing capabilities that are intrinsically tactile: Pietrini *et al.* (2004), investigating tactile object recognition by congenitally blind subjects, observed category-related activity in brain areas related to visual object categorization, thus establishing that category-related representations develop in these areas despite the total absence of visual experience. Similarly, for the case of auditory movement, Poirier *et al.* (2006) reported that brain areas relating to visual-motion processing, including hMT+/V5, are activated in early blind subjects by moving stimuli presented through the auditory modality.

An attractive possibility for future work is to investigate the suggestion (as discussed above) that cortical activations in response to moving tactile stimuli may be spatially organized according to the nature of the stimulus movement. The 100-element stimulator described in this paper has the potential to generate moving stimuli on the fingertip with a wide range of speeds, directions, and geometric forms, as required for such an experiment. (This aspect has not been investigated in the present study, which involves only one type of moving tactile stimulus.)

Similarly, further information on the distribution of activation within the hMT+/V5 complex might be obtained in a future experiment by providing visual stimuli separately to the left and right visual fields (Beauchamp *et al.*, 2007), allowing separate identification of the MST region (since the MT region does not respond to ipsilateral visual stimuli).

As mentioned above, activation in hMT+/V5 has been observed in the present study, which involves "virtual stroking" over the fingertip, and in the study by Hagen *et al.* (2002), which involved stroking along the forearm. However, this activation was not observed for the air-jet stimulus used by Bremmer *et al.* (2001), which produced no coherent variation of the stimulus location. Similarly, activation in hMT+/V5 was not observed in the studies by Burton *et al.* (1999) and Bodegård *et al.* (2000), which involved brushing or rubbing, again with no coherent variation of the stimulus location. It may be conjectured that hMT+/V5 activation by a touch stimulus is associated with coherent variation of stimulus location. The 100-element stimulator used in the present study can generate either coherent or incoherent movement across the contactor array, allowing this conjecture also to be tested in a future experiment.

ACKNOWLEDGMENTS

The authors thank the subjects for their important contribution to this study. Thanks are also due to Abdelmalek Benattayallah and Jon Fulford for helpful discussions, and to two anonymous reviewers for valuable suggestions. This work was supported by the Medical Research Council (UK), Program Grant No. G9900259.

Beauchamp, M. S., Yasar, N. E., Kishan, N., and Ro, T. (**2007**). "Human MST but not MT responds to tactile stimulation," J. Neurosci. **27**, 8261–8267.

Blake, R., Sobel, K. V., and James, T. W. (**2004**). "Neural synergy between kinetic vision and touch," Psychol. Sci. **15**, 397–402.

Bodegård, A., Geyer, S., Naito, E., Zilles, K., and Roland, P. E. (**2000**). "Somatosensory areas in man activated by moving stimuli: Cytoarchitectonic mapping and PET," NeuroReport **11**, 187–191.

Bremmer, F., Schlack, A., Shah, N. J., Zafiris, O., Kubischik, M., Hoffmann, K.-P., Zilles, K., and Fink, G. R. (**2001**). "Polymodal motion processing in posterior parietal and premotor cortex: A human fMRI study strongly implies equivalencies between humans and monkeys," Neuron **29**, 287–296.

Briggs, R. W., Dy-Liacco, I., Malcolm, M. P., Lee, H., Peck, K. K., Gopinath, K. S., Himes, N. C., Soltysik, D. A., Browne, P., and Tran-Soy-Tay, R. (**2004**). "A pneumatic vibrotactile stimulation device for fMRI," Magn. Reson. Med. **51**, 640–643.

Burton, H., Abend, N. S., MacLeod, A. M. K., Sinclair, R. J., Snyder, A. Z., and Raichle, M. E. (**1999**). "Tactile attention tasks enhance activation in somatosensory regions of parietal cortex: A positron emission tomography study," Cereb. Cortex **9**, 662–674.

Colby, C. L., Duhamel, J. R., and Goldberg, M. E. (**1993**). "Ventral intraparietal area of the macaque: Anatomic location and visual response properties," J. Neurophysiol. **69**, 902–914.

Cook, E. P., and Maunsell, J. H. R. (**2002**). "Dynamics of neuronal responses in macaque MT and VIP during motion detection," Nat. Neurosci. **5**, 985–994.

Dresela, C., Andreas Parzingera, A., Rimpauc, C., Zimmerb, C., Ceballos-Baumannd, A. O., and Haslingera, B. (**2008**). "A new device for tactile stimulation during fMRI," Neuroimage **39**, 1094–1103.

Deuchert, M., Ruben, J., Schwiemann, J., Meyer, R., Thees, S., Krause, T., Blankenburg, F., Villringer, K., Kurth, R., Curio, G., and Villringer, A. (**2002**). "Event-related fMRI of the somatosensory system using electrical finger stimulation," NeuroReport **13**, 365–369.

Driver, J., and Spence, C. (**2000**). "Multisensory perception: Beyond modularity and convergence," Curr. Biol. **10**, R731-R735.

Dubner, R., and Zeki, S. (**1971**). "Response properties and receptive fields of cells in an anatomically defined region of the superior temporal sulcus in the monkey," Brain Res. **35**, 528–532.

Dukelow, S. P., DeSouza, J. F. X., Culham, J. C., van den Berg, A. V., Menon, R. S., and Vilis, T. (**2001**). "Distinguishing subregions of the human MT+ complex using visual fields and pursuit eye movements," J. Neurophysiol. **86**, 1991–2000.

Dumoulin, S. O., Bittar, R. G., Kabani, N. J., Baker, C. L., Le Goualher, G., Pike, G. B., and Evans, A. C. (**2000**). "A new anatomical landmark for reliable identification of human area V5/MT: A quantitative analysis of sulcal patterning," Cereb. Cortex **10**, 454–463.

Francis, S. T., Kelly, E. F., Bowtell, R., Dunseath, W. J. R., Folger, S. E., and McGlone, F. (**2000**). "FMRI of the responses to vibratory stimulation of digit tips," Neuroimage **11**, 188–202.

Friston, K. J., Ashburner, J., Frith, C. D., Poline, J. B., Heather, J. D., and Frackowiak, R. S. J. (**1995**). "Spatial registration and normalization of images," Hum. Brain Mapp **3**, 165–189.

Genovese, C. R., Lazar, N. A., and Nichols, T. (**2002**). "Thresholding of statistical maps in functional neuroimaging using the false discovery rate," Neuroimage **15**, 870–878.

Hagen, M. C., Franzen, O., McGlone, F., Essick, G., Dancer, C., and Pardo, J. V. (**2002**). "Tactile motion activates the human middle temporal/V5 (MT/V5) complex," Eur. J. Neurosci. **16**, 957–964.

Lewis, J. W., Beauchamp, M. S., and DeYoe, E. A. (**2000**). "A comparison of visual and auditory motion processing in human cerebral cortex," Cereb. Cortex **10**, 873–888.

Magnenat-Thalmann, N., Volino, P., Bonanni, U., Summers, I. R., Bergamasco, M., Salsedo, F., and Wolter, F.-E. (**2007**). "From physics-based simulation to the touching of textiles: The HAPTEX project," Int. J. Virtual Reality **6**, 35–44.

Malikovic, A., Amunts, K., Schleicher, A., Mohlberg, H., Eickhoff, S. B., Wilms, M., Palomero-Gallagher, N., Armstrong, E., and Zilles, K. (**2007**). "Cytoarchitectonic analysis of the human extrastriate cortex in the region of V5/MT+: A probabilistic, stereotaxic map of area hOc5," Cereb. Cortex **17**, 562–574.

McGlone, F., Kelly, E. F., Trulsson, M., Franci, S. T., Westling, G., and Bowtell, R. (**2002**). "Functional neuroimaging studies of human somatosensory cortex," Behav. Brain Res. **135**, 147–158.

Mohr, C. M., King, W. M., Freeman, A. J., Briggs, R. W., and Leonard, C. M. (**1999**). "Influence of speech stimuli intensity on the activation of auditory cortex investigated with functional magnetic resonance imaging," J. Acoust. Soc. Am. **105**, 2738–2745.

Overduin, S. A., and Servos, P. (**2004**). "Distributed digit somatotopy in primary somatosensory cortex," Neuroimage **23**, 462–472.

Pietrini, P., Furey, M. L., Ricciardi, E., Gobbini, M. I., Wu, W. H. C., Cohen, L., Guazzelli, M., and Haxby, J. V. (**2004**). "Beyond sensory images: Object-based representation in the human ventral pathway," Proc. Natl. Acad. Sci. U.S.A. **101**, 5658–5663.

Poirier, C., Collignon, O., De Volder, A. G., Renier, L., Vanlierde, A., Tranduy, D., and Scheiber, C. (**2005**). "Specific activation of the V5 brain area by auditory motion processing: An fMRI study," Brain Res. Cognit. Brain Res. **25**, 650–658.

Poirier, C., Collignon, O., Scheiber, C., Renier, L., Vanlierde, A., Tranduy, D., Veraart, C., and De Volder, A. G. (**2006**). "Auditory motion perception activates visual motion areas in early blind subjects," Neuroimage **31**, 279–285.

Price, C. J., and Friston, K. J. (**1997**). "Cognitive conjunction: A new approach to brain activation experiments," Neuroimage **5**, 261–270.

Ricciardi, E., Vanello, N., Sani, L., Gentili, C., Scilingo, E. P., Landini, L., Guazzelli, M., Bicchi, A., Haxby, J. V., and Pietrini, P. (**2007**). "The effect of visual experience on the development of functional architecture in hMT+," Cereb. Cortex **17**, 2933–2939.

Summers, I. R., and Chanter, C. M. (**2002**). "A broadband tactile array on the fingertip," J. Acoust. Soc. Am. **112**, 2118–2126.

Tootell, R. B. H., Reppas, J. B., Kwong, K. K., Malach, R., Born, R. T., Brady, T. J., Rosen, B. R., and Belliveau, J. W. (**1995**). "Functional-analysis of human MT and related visual cortical areas using magnetic-resonance-imaging," J. Neurosci. **15**, 3215–3230.

Vanello, N., Ricciardi, E., Dente, D., Sgambelluri, N., Scilingo, E. P., Gentili, C., Sani, L., Positano, V., Santarelli, F. M., Guazzelli, M., Haxby, J. V., Landini, L., Bicchi, A., and Pietrini, P. (**2004**). "Perception of optic and tactile flow both activate V5/MT cortical complex in the human brain," Proceedings of the Tenth Annual Meeting of the Organization for Human Brain Mapping, Budapest, Hungary, Poster No. TU 323.

Watson, J. D. G., Myers, R., Frackowiak, R. S. J., Hajnal, J. V., Woods, R. P., Mazziotta, J. C., Shipp, S., and Zeki, S. (**1993**). "Area V5 of the human brain: Evidence from a combined study using positron emission tomography and magnetic resonance imaging," Cereb. Cortex **3**, 79–94.

Whalen, D. H., Benson, R. R., Richardson, M., Swainson, B., Clark, V. P., Lai, S., Mencl, W. E., Fulbright, R. K., Constable, R. T., and Liberman, A. M. (**2006**). "Differentiation of speech and nonspeech processing within primary auditory cortex," J. Acoust. Soc. Am. **119**, 575–581.

Worsley, K. J., Marrett, S., Neelin, P., Vandal, A. C., Friston, K. J., and Evans, A. C. (**1996**). "A unified statistical approach for determining significant signals in images of cerebral activation," Hum. Brain Mapp **4**, 58–73.

Zappe, A. C., Maucher, T., Meier, K., and Scheiber, C. (**2004**). "Evaluation of a pneumatically driven tactile stimulator device for vision substitution during fMRI studies," Magn. Reson. Med. **51**, 828–834.

Zeki, S., Watson, J. D., Lueck, C. J., Friston, K. J., Kennard, C., and Frackowiak, R. S. (**1991**). "A direct demonstration of functional specialization in human visual cortex," J. Neurosci. **11**, 641–649.

Effects of temporal uncertainty and temporal expectancy on infants' auditory sensitivity[a)]

Lynne A. Werner,[b)] Heather K. Parrish, and Nicole M. Holmer
Department of Speech and Hearing Sciences, University of Washington, Seattle, Washington 98105-6246

(Received 19 September 2007; revised 8 October 2008; accepted 13 November 2008)

Adults are more sensitive to a sound if they know when the sound will occur. In the present experiment, the effects of temporal uncertainty and temporal expectancy on infants' and adults' detection of a 1 kHz tone in a broadband noise were examined. In one experiment, masked sensitivity was measured with an acoustic cue and without an acoustic cue to possible tone presentation times. Adults' sensitivity was greater for the cue than for the no-cue condition, while infants' sensitivity did not differ significantly between the cue and no-cue conditions. In a second experiment, the effect of temporal expectancy was investigated. The detection advantage for sounds occurring at an expected (most frequent) time, over sounds occurring at unexpected (less frequent) times, was examined. Both infants and adults detected a tone better when it occurred before or at an expected time following a cue than when it occurred at a later time. Thus, despite the fact that the auditory cue did not improve infants' sensitivity, it nonetheless provided the basis for temporal expectancies. Infants, like adults, are more sensitive to sounds that are consistent with temporal expectancy. © *2009 Acoustical Society of America.*. [DOI: 10.1121/1.3050254]

PACS number(s): 43.66.Dc, 43.66.Mk [RYL] Pages: 1040–1049

I. INTRODUCTION

It is well established that infants have higher detection thresholds than adults (e.g., Schneider *et al.*, 1989; Bargones *et al.*, 1995; Berg and Boswell, 1999). Several contributors to infants' immature sensitivity have been suggested, including inattention and unselective listening across frequency (Bargones and Werner, 1994; Bargones *et al.*, 1995; Werner and Boike, 2001). Another possible contributor is unselective listening in time.

In a typical infant test procedure, the listener hears a continuous background noise that starts at the beginning of the test session. At certain points during the session the target tone is presented. There is no explicit cue informing the listener of when a tone might be presented, although the experimenter observing the listener's response is aware that a trial is in progress. The listener is thus uncertain about the timing of the signal tone. Temporal uncertainty is a feature of most, if not all, psychoacoustical procedures applied to infant listeners (e.g., Schneider and Trehub, 1985; Werner, 1995; Berg and Boswell, 1999).

It is generally accepted that adults listen selectively in time to optimize sensitivity. Temporal uncertainty reduces adults' auditory sensitivity (Egan *et al.*, 1961; Lappin and Disch, 1973; Lisper *et al.*, 1977). For example, Egan *et al.* (1961) examined listeners' ability to detect a tone in broadband noise when the time at which the tone could be presented was known and when the presentation time varied over intervals of 1–8 s. Their results showed that a signal that was detected with a d' of about 1.5, when presentation time was known, was detected with a d' of only 0.75 when the interval of uncertainty was 8 s. The results of Egan *et al.* (1961) indicate that such an increase in uncertainty would be equivalent to a 9 dB decrease in signal-to-noise ratio.

Another indication that adults listen selectively in time is that adult listeners detect tones better at expected presentation times than at unexpected presentation times (Leis Rossio, 1986; Chang, 1991; Chang and Viemeister, 1991). On a majority of trials in these studies, the signal occurred at a fixed time during the observation interval, the "expected" time. On the remaining trials, the signal occurred either before or after the expected time; these signals occurred at "unexpected" times. Leis Rossio (1986) measured adults' hit rate for a click in noise when the expected presentation time was 500 ms into the observation interval with unexpected presentation times varying between 100 and 1100 ms. A single-interval, yes-no procedure was used in that study. Chang (1991) and Chang and Viemeister (1991) used a two-interval forced-choice method and a 20 ms tone as a signal. Beside a visual indicator of each observation interval, a click was presented in the contralateral ear to indicate precisely the expected presentation time within the observation interval. Unexpected presentation times varied between 100 and 900 ms. Although the details of the procedures and the gradient of performance over time differed between the studies, both showed that as the presentation time of the signal deviated from the expected time, detection of the signal grew poorer. The results of these studies further support the benefit of knowing when to listen for a sound.

The effects of temporal uncertainty on detection have not been studied developmentally. If infants' detection is more disrupted by temporal uncertainty than adults', that

[a)]Portions of this work are based on a M.S. thesis submitted by H. Parrish to the University of Washington, Seattle, WA. Portions of this work were presented at the 1994, 1999, and 2003 Midwinter Meeting of the Association for Research in Otolaryngology and at the 2004 Meeting of the American Auditory Society.

[b)]Electronic mail: lawerner@u.washington.edu

could at least partially explain why infants' thresholds are higher than adults' in a temporally uncertain test procedure.

The effects of frequency uncertainty have been examined in children. Allen and Wightman (1995) found that children's detection was less affected than adults' by uncertainty about signal frequency, suggesting that children did not focus on a particular frequency when the signal frequency was known. Other results support the idea that infants do not listen selectively in frequency. For example, while adults detect tones at an expected frequency better than those at unexpected frequencies, infants detect expected and unexpected frequency tones equally well (Bargones and Werner, 1994). If infants do not focus on the time when signals are expected to occur, then decreasing temporal uncertainty may produce little change in infants' sensitivity. If that were the case, the difference between infants' and adults' sensitivity would be greater when temporal uncertainty is reduced, because adults' sensitivity would improve, but infants' would not.

The goal of the present experiments was to determine how infants' and adults' detection of a tone in noise is affected by temporal uncertainty and temporal expectancy in an infant test procedure. First, detection in the typical, temporally uncertain, infant procedure was compared to detection when a cue to the timing of signal presentation was provided. Second, detection of tones that occurred at expected times was compared to that of tones that occurred at unexpected times.

II. EXPERIMENT 1: EFFECTS OF TEMPORAL CUES ON SENSITIVITY

A. Method

1. Subjects

The data were provided by 98 infants and 93 adults. The age of the infants ranged from 29 to 40 wk (M=33.5 wk; SD=3.3 wk). The age of the adults ranged from 19 to 31 yr (M=24 yr; SD=3 yr). All subjects had normal hearing, as assessed by parent report or self-report. None had any risk factors associated with hearing loss, and all subjects passed screening typanometry on the test day. All infants were full term, healthy, and developing normally by parent report.

2. Stimuli

Subjects detected a 1 kHz tone, 300 ms in duration, with 16 ms rise and fall times, in the presence of a 2500 Hz low-pass noise. The noise was presented continuously throughout the session. The level of the tone was 50 dB Sound Pressure Level (SPL) for the infants and 42 dB SPL for the adults. The spectrum level of the noise was always 20 dB SPL during trials. These levels were chosen to allow detection of the tone with a $d'=1$, based on the results of a previous study (Bargones *et al.*, 1995).

In the cue conditions, the cue indicated when the trial began; when the tone was presented, its onset was at a fixed interval after the cue. The cues were always acoustic cues. Acoustic cues were chosen, because even young infants focus attention within a sensory modality and respond less to stimulation in a modality other than the one on which they are focused (Richards, 2000). Thus, it seemed preferable to present the cue in the same modality as the target stimulus. Two different cue conditions and a no-cue condition were tested. Each subject was tested in only one of these conditions. Data collection for one cue condition was completed before data collection for the other cue condition. To make sure that any effect of the cue type was not due to changes in testers or instrumentation, a separate group of listeners was tested in the no-cue condition for each cue condition.

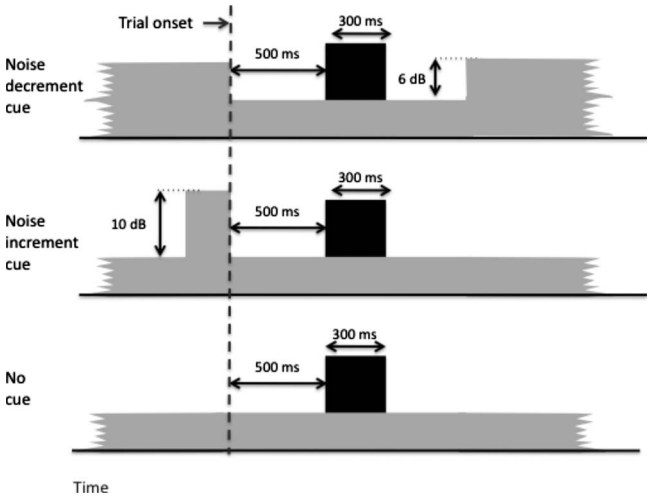


FIG. 1. Stimulus configurations in experiment 1. A broadband noise is presented continuously (gray shading). Tones (black rectangles) are presented on tone trials. In the no-cue condition (bottom panel), a tone is presented 500 ms after the observer starts the trial, with no indication to the listener that the trial is underway. In the noise-decrement cue condition, the spectrum level of the background noise drops from 26 to 20 dB SPL, 500 ms before the tone and remains at 20 dB for the duration of the trial. In the noise-increment cue condition, the spectrum level of the background noise increases from 20 to 30 dB for 200 ms, and then returns to 20 dB, 500 ms before the tone. The trial configurations on no-tone trials are the same as the tone trial in each condition, except that no tone is presented.

The stimulus conditions are depicted in Fig. 1. In the noise-decrement-cue condition, a reduction in background noise level was the cue. The spectrum level of the noise was 26 dB SPL until trial onset. At trial onset, the spectrum level dropped to 20 dB SPL. When the tone was presented, its onset was 500 ms after trial onset. Thus, the cue was the drop in the level of the background noise. In the noise-increment-cue conditions, the cue to trial onset was a 200 ms, 10 dB increment in the background noise. When the tone was presented, its onset was 500 ms after the offset of the noise increment. In the corresponding no-cue conditions, no cue was presented to the listener to mark the onset of the trial, but when the tone was presented its onset was 500 ms after trial onset. Previous studies indicate that infants of this age can easily detect noise level changes of the magnitudes used in the noise cue conditions (Berg and Boswell, 1998; Werner and Boike, 2001).

Data collection in the noise-decrement-cue conditions was completed first. A noise decrement, rather than an increment, was chosen as the cue to ensure that the cue did not mask the signal tone. Subsequent work showed that forward masking of the tone by the cue would not be expected in this condition (Werner, 1999). An unexpected result in the noise-decrement-cue conditions led us to repeat the study using the noise-increment cue. The number of subjects tested in the

noise-decrement cue, and no-cue conditions were 26 and 32, respectively, for the infants, and 28 and 25, respectively, for the adults. The number of subjects tested in the noise-increment cue, and no-cue conditions were 21 and 19, respectively, for the infants, and 20 and 20, respectively, for the adults.

The stimuli were presented to the subject's right ear using an Etymotic ER-1 insert earphone. A computer controlled the presentation of the stimulus and stored the results on each trial. Testing took place in a sound-attenuating booth.

3. Procedure

Infants' detection of the tone was measured using the observer-based psychoacoustic procedure (Werner, 1995). The infant, with ear tip in place, was seated on a parent's lap in the booth. An assistant, seated to the infant's left, manipulated toys on a table in front of the infant to maintain the infant's gaze forward. Both the parent and assistant listened to masking sounds over circumaural headphones so that they could not hear any of the sounds presented to the infant. Two mechanical toys in dark Plexiglas boxes with lights were placed to the infant's right; these toys were activated to reinforce the infants' response to the tone as described below. An observer watched the infant through a one-way window and on a video monitor. The observer pushed a button interfaced to the computer to begin a trial when the infant was quiet and attentive, without knowing whether a tone would be presented or not.[1] Both "tone trials" and "no-tone trials" were presented. Trials were 4 s in duration. If the observer judged on the basis of the infant's behavior that a tone had occurred, she pushed a button to indicate a "yes" response. If the observer was correct in judging that a tone had occurred, one of the mechanical toys in the test booth was illuminated and activated as reinforcement for the infant. The observer received feedback at the conclusion of all trials. The same general procedure was used to test adults. The adult subject was told to respond "when you hear the sound that will make the toy come on." An assistant outside the booth recorded the adult's responses, and a mechanical toy was activated when a response was recorded during a tone trial.

At the beginning of each session, a brief (approximately five trials) training phase was completed during which the tone was clearly audible and the reinforcer toy was turned on after every tone trial. This procedure demonstrated to the infant that the tone (or cue+tone) was associated with the toy. The toy was never turned on after no-tone (or cue alone) trials. In the second training phase, the tone remained clearly audible, tone and no-tone trials were equally probable, and the reinforcer toy only came on if the observer correctly identified a tone trial. The infant/observer team or the adult subject was required to achieve 80% correct on both tone and no-tone trials. Thus, in the cue conditions, the infant learned to respond to cue+tone, but not to the cue alone. Similarly, the observer learned to differentiate the infant's response to the cue+tone from the infant's response to the cue alone. This phase took about 22 trials to complete in all conditions. Once training criterion had been met, 35 test trials were presented, including 15 tone trials, 15 no-tone trials, and 5 probe trials. On probe trials, the level of the tone was chosen to be readily detectable, 51 dB SPL for adults and 60 dB SPL for infants. A subject's data were only used if at least three of the five probe tones were detected. This provided a check that the subject was "on task."

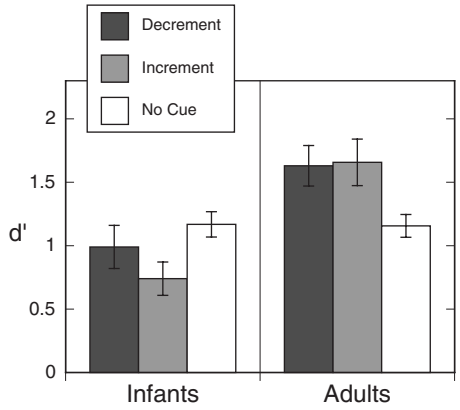


FIG. 2. Mean d' as a function of cue condition in experiment 1 for infants and adults, ± 1 SEM.

If a subject reached training criterion but did not complete all test trials, a new block of test trials was completed in a subsequent visit after an abbreviated training procedure.

Sensitivity was expressed as d'. Hit or false alarm rates of 1 or 0 were adjusted by $1/2n$ where n is the number of trials (Macmillan and Creelman, 2005). Levene's test of homogeneity of variance was significant in the dataset as a whole (with both infants and adults, $p<0.0001$), but it was not significant within age groups (both $p>0.4$). For that reason, the effect of the cues on d' was analyzed within age groups, and the pattern of effects compared between age groups.

B. Results

In the no-cue conditions, both infants and adults generally achieved a d' around 1.0, as expected (e.g., Bargones *et al.*, 1995). Mean d' in the noise-decrement no-cue group was 1.28 (SD=0.83) for infants and 1.16 (SD=0.64) for adults; in the noise-increment no-cue group mean d' was 0.98 (SD =0.60) for infants and 1.15 (SD=0.70) for adults. The differences between the two no-cue groups were not statistically significant by t-test [$t(49)=1.4$, $p=0.17$, $d=0.4$) for infants; $t(43)=0.03$, $p=0.98$, $d=0.01$ for adults]. The data of the two no-cue groups were therefore pooled within age groups in the remainder of the analyses.

Average d' in the noise-decrement-cue (dark gray bars), noise-increment-cue (light gray bars), and no-cue (white bars) conditions is plotted in Fig. 2, with infants' data on the left and adults' data on the right. In each cue condition, adults' d' was greater than in the no-cue condition. One-way analysis of variance (ANOVA) indicated a significant effect of cue type (noise-decrement cue, noise-increment cue, no cue) for the adults [$F(2,90)=4.81$, $p=0.1$, $\eta^2=0.10$]. Bonferroni *post hoc* pairwise comparisons showed that d' was significantly higher in each of the cue conditions than in the no-cue condition (both $p<0.04$). The two cue conditions were not significantly different for adults ($p>0.99$).

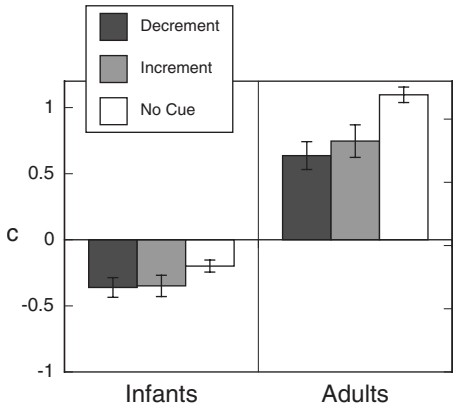


FIG. 3. Mean c as a function of cue condition in experiment 1 for infants and adults, ± 1 SEM.

Infants' d' in each cue condition, however, was actually a little lower than that in the no-cue condition; clearly neither cue improved infants' detection of the tone. For the infants, the effect of cue type was only marginally significant by one-way ANOVA [$F(2,95)=2.70$, $p=0.7$, $\eta^2=0.05$) Bonferroni *post hoc* pairwise comparisons showed a marginally significant difference in d' between the noise-increment-cue and no-cue condition ($p=0.083$). The noise-decrement-cue and no-cue conditions were not statistically different ($p=0.55$), and the two cue conditions were not significantly different ($p>0.99$).

Adults are typically very conservative in their response bias in an "infant procedure," while infants/observers tend to be unbiased or a little liberal in their response bias in the same procedure (e.g., Werner and Marean, 1991). A cue might be expected to change response bias, although it is not clear that infants and adults would be affected in the same way. To examine the effect of the cue on response bias, bias was described as

$$c = 0.5[z(\text{hit rate}) + z(\text{false alarm rate})]$$

(Macmillan and Creelman, 2005). Hit or false alarm rates of 1 or 0 were adjusted by $1/2n$, where n is the number of trials. Positive values of c indicate a conservative bias, while negative values indicate a liberal bias. In the no-cue conditions, infants were somewhat liberal responders, while adults were quite conservative, as expected. Mean c in the noise-decrement no-cue group were -0.21 (SD=0.33) for infants and 1.05 (SD=0.39) for adults; in the noise-increment no-cue group -0.18 (SD=0.37) for infants and 1.15 (SD=0.37) for adults. The differences between the two no-cue groups were not statistically significant by t-test [$t(49)=-0.32$, $p=0.74$, $d=0.1$ for infants; $t(43)=-0.88$, $p=0.38$, $d=0.26$ for adults]. The data of the two groups were therefore pooled within age groups in the remainder of the analyses.

Average c in the noise-decrement-cue (dark gray bars), noise-increment-cue (light gray bars), and no-cue (white bars) conditions is plotted in Fig. 3, with infants' data plotted on the left and adults' data plotted on the right. As noted, infants tended to be a little liberal, while adults tended to be conservative. Both infants and adults tended to respond more liberally when a cue was provided, although the effect appears smaller for infants than for adults.

One-way cue-type ANOVA of c indicated a significant effect for adults [$F(2,90)=8.87$, $p=0.0003$, $\eta^2=0.16$]. Bonferroni *post hoc* pairwise comparisons indicated that both types of cues made adults significantly more liberal ($p<0.0001$ for noise-decrement cue, $p=0.025$ for noise-increment cue). For infants, the one-way cue-type ANOVA of c was marginally significant [$F(2,95)=0.246$, $p=0.09$, $\eta^2=0.05$), but the Bonferroni *post hoc* pairwise comparisons were not significant ($p=0.173$ for noise-decrement cue, $p=0.305$ for noise increment cue). Thus, the cues clearly made adults more liberal in their response bias, and although infants' bias changed in the same direction with the cues, the effect was not statistically significant.

C. Discussion

The results of Experiment 1 indicate that a cue to trial onset led to improved performance in adult listeners, but not in infant listeners. For adults, this result held whether the cue was a decrement in the level of the background noise or an increment in the level of the background noise. Cue type also made little difference to infants, although some weak evidence suggested that the noise-increment cue could be detrimental to infants' performance.

As expected, adults' pure-tone detection was better when temporal uncertainty was reduced. This result is qualitatively consistent with previous reports (Egan *et al.*, 1961; Lappin and Disch, 1973; Lisper *et al.*, 1977).

The results for infant listeners suggest that infants do not benefit from a reduction in temporal uncertainty. It is possible that some other sort of cue might improve infants' detection performance. We avoided using a visual cue in the current experiments so that infants would not be required to divide attention between sensory modalities (e.g., Richards, 2000). However, a recent study suggests that visual information can facilitate infants' ability to separate an auditory target from a masker. Hollich *et al.* (2005) tested infant's ability to recognize a word in a background of competing speech. If the target word was paired with a video of a face saying the word, or even with an "oscilloscopelike trace" that was temporally synchronized with the word, infants recognized the word at a lower signal-to-background ratio than they did when no visual information was provided, or if the visual display was not synchronized with the target word. This suggests that a visual cue could improve infants' detection of a tone, even if an auditory cue does not.

To benefit from any cue, the listener must (1) learn that the cue predicts the possible occurrence of the signal, (2) learn and remember when the signal could occur following the cue, and (3) be able to listen selectively at the predicted time. One explanation for the cue's failure to improve infant's detection is that infants do not form expectancies that one event will follow another. Casual observation of infants suggests this is unlikely. Furthermore, it is well established that infants develop expectations that one visual event will follow another (e.g., Haith *et al.*, 1988). Another explanation is that while infants develop expectancies and attempt to direct listening to the appropriate time, their ability to estimate or to remember the interval between events is highly

inaccurate. In that case, their uncertainty about the timing of the signal might not be reduced by a cue. A final explanation is that infants are not able to listen at a particular time for an expected event. That would be consistent with their listening along the frequency dimension: Infants detect expected and unexpected frequencies equally well, while adults detect expected frequencies better than unexpected frequencies (Bargones and Werner, 1994).

Experiment 2 was a more direct test of infants' ability to form and use temporal expectancies about sounds. The probe-signal method (e.g., Greenberg and Larkin, 1968; Scharf, 1987; Schlauch and Hafter, 1991; Dai and Wright, 1995; Arbogast and Kidd, 2000) was used to determine whether infants detect sounds presented at expected times better than they detect sounds presented at unexpected times. In this method, listeners detect a tone. On 75% of the trials, the tone is presented at one temporal position in the observation interval; on the remaining trials, the tone is presented before or after that time. The level of the tone is set so that it is detectable on, perhaps, 80% of the trials if it is presented at a fixed time. Adults detect the tone presented at the more common temporal position more often than they detect the tones at the other temporal positions (Leis Rossio, 1986; Chang, 1991; Chang and Viemeister, 1991), just as they detect tones at a more common frequency (e.g., Schlauch and Hafter, 1991), duration (Wright and Dai, 1994; Dai and Wright, 1995), or spatial position (Arbogast and Kidd, 2000) more often than they detect tones at other frequencies, durations, or spatial positions. We have previously used the probe-signal method to examine infants' "listening bands" in frequency (Bargones and Werner, 1994). We refer to the effect of temporally selective listening as a "listening window."

III. EXPERIMENT 2: PROBE-SIGNAL STUDY OF TEMPORAL SELECTIVITY

A. Method

1. Subjects

The final sample included 14 28–36 wk old infants and 19 18–30 yr old adults. The average age of infants was 35.5 wk (SD=2.2 wk). The average age of adults was 22.3 yr (SD=2.8 yr). The inclusion criteria were the same as for experiment 1. Sixteen other infants met the training criteria, but did not provide a complete data set in three visits to the laboratory. Fifteen infants did not meet training criteria. Four adults were excluded because they did not provide a complete data set in a 1 h test session. The high exclusion rate reflects the difficulties of this paradigm, in which the detectability of the stimulus must be controlled for each subject individually. If it takes several attempts to find an appropriate test level for an infant, time and patience often run out before all data have been obtained.

2. Stimuli

The stimuli were a 1 kHz tone and a broadband noise low-pass filtered at 2500 Hz. The duration of the tone was 150 ms, with 15 ms rise and fall times. The noise was presented continuously at a spectrum level of 30 dB SPL. The 10 dB noise increment used as a cue in experiment 2 was used to mark the beginning of an observation interval. On signal trials, the tone was presented either 200, 500, or 800 ms following the offset of the cue. The level of the tone was set for each listener to produce a correct detection rate of 70%–85%.

The duration of the tone burst is much longer than the stimuli used in previous studies of adults' listening windows (20 ms, Chang, 1991; 0.5 ms, Leis Rossio, 1986). A short-duration stimulus is desirable to obtain a narrow listening window. The longer duration was chosen, because infants' thresholds for short-duration sounds tend to be relatively worse, compared to adults, than their thresholds for a longer duration sound (Bargones *et al.*, 1995; Berg and Boswell, 1995). Thus, for this first attempt to examine temporal selectivity, we sacrificed precise estimation of the duration of the listening window to be able to obtain reasonable data from infants. Moreover, pilot testing indicated that adults demonstrated temporal selectivity with the 150 ms tone.

3. Procedure

The observer-based procedure was used to obtain these data, as in the earlier experiment. Each subject was tested in two conditions. In the "fixed" condition, the tone was presented at the same presentation time on all trials, either 200, 500, or 800 ms following the cue. Presumably, the assigned presentation time would be the expected presentation time in this condition. Subsequently, in the "mixed" condition, the tone was presented at 500 ms on 75% of the trials, at 200 ms on 12.5% of the trials, and at 800 ms on 12.5% of the trials. In this condition, the 500 ms presentation time is presumably the expected time and the other times, unexpected. If listeners are temporally selective, each of the presentation times should be detected equally well in the fixed conditions. In the mixed condition, tones presented at the unexpected times should not be detected as well as the tone at the expected time, and tones at the unexpected times should not be detected as well as tones presented at the same times in the fixed condition. Although there are no comparable data on temporal expectancy, several papers that have examined the effect of frequency expectancy have demonstrated that expectancy is built up quickly and is relatively robust with respect to the proportions of expected and unexpected frequencies in the mixed block (e.g., Scharf, 1987). Pilot testing with adults in our laboratory indicated that the same is true of temporal expectancy.

The training procedure was the same as in experiment 1. The level of the tone during training was 70 dB SPL for infants and 60 dB SPL for adults. The training procedure was completed twice, prior to testing in each condition. The presentation time in training matched that in testing in the fixed condition; in the mixed condition, the presentation time in training was 500 ms for all subjects.

The fixed condition was tested first. The purpose of the fixed condition was, first, to identify a tone level that the subject could detect 70%–85% of the time, and second, to assess performance when the tone was only presented at one of the three presentation times used in the mixed condition (i.e., when each was the expected time). Each subject was

randomly assigned to the 200, 500, or 800 ms fixed presentation time condition. After reaching the training criterion of at least 80% correct, the listener completed a block of 32 test trials, with equal numbers of tone and no-tone trials. In the initial test block, the level of the tone was set at 62 dB SPL for the infants and at 49 dB for the adults. These levels were chosen on the basis of performance with the noise-increment cue in experiment 1. Fixed test blocks were repeated, adjusting the tone level as needed, until a level was identified at which the listener achieved a hit rate between 70% and 85% and a false alarm rate of no more than 40% on the test block. The average intensity levels used for infants and adults were 64.1 dB (SD=0.52 dB) and 48.8 dB (SD=1.0 dB), respectively.

In the mixed condition, the listener completed another training phase, with a presentation time of 500 ms. After reaching the training criterion of 80% correct, the listener completed a block of 32 test trials. The block contained 16 no-tone trials, 12 tone trials with a 500 ms presentation time, 2 tone trials with a 200 ms presentation time, and 2 tone trials with an 800 ms presentation time. The level of the tone was the level identified in the fixed condition as yielding a hit rate between 70% and 85% and a false alarm rate no greater than 40%. If the listener did not reach training criterion, training was reattempted in the next session. If the listener's false alarm rate was greater than 40% in the mixed test block, the same block was retested in subsequent sessions.

Whenever a fixed or mixed test block was repeated, the listener was given a few reminder trials prior to beginning the testing phase in subsequent test sessions. Adults completed all conditions in one 1 h session. Infants were scheduled for three sessions; a few infants required a fourth session to complete all conditions. Typically, an infant who did not complete all conditions in three sessions was excluded from the study.

The observer-based method used to test listeners in this study is a single-interval procedure. Single-interval procedures have been used in probe-signal experiments (e.g., Greenberg and Larkin, 1968) and other studies of the effects of uncertainty on detection (e.g., Richards and Neff, 2004; Scharf *et al.*, 2007). When the frequency of the signal varies in a block of trials, false alarm rate cannot be estimated independently for each signal, because it is not clear which no-signal trials should be assigned to each signal. In some cases (e.g., Greenberg and Larkin, 1968), hit rate has been used as the metric of performance for that reason. In the course of data collection in this experiment, hit rate was used as the primary performance measure. However, preliminary analyses revealed apparent differences in response bias that could influence the interpretation of the results. Infants had higher false alarm rates than adults (infants M=0.31, SD =0.06; adults M=0.07, SD=0.08). For that reason, d' was analyzed rather than hit rate. Hit or false alarm rates of 1 or 0 were corrected to $1-1/2n$ or $1/2n$, respectively, where n is the number of trials contributing to the rate (Macmillan and Creelman, 2005).

Two issues arise in the application of d' in this context. As discussed above, the first is the calculation of d' in the mixed condition, in which the temporal position of the signal in the observation interval varied within a block of trials. We calculated d' following the approach of Macmillan and Creelman (2005) using a single composite false alarm rate. This method has been shown to produce estimates of d' similar to those estimated with separate false alarm rates for each signal or in a two-alternative forced-choice procedure (Richards and Neff, 2004; Scharf *et al.*, 2007).

The second issue is the calculation of d' for individual listeners in different conditions, because the number of trials was very small. Small n would increase variability and statistical bias in d' estimates. More importantly, because the number of signal trials at unexpected times was, of necessity, much smaller than that at expected times, the maximum achievable hit rate in the unexpected conditions, after correction for rates of 0 or 1, would be much smaller than that in the expected condition. Consequently, d' would be lower in the unexpected than in the expected condition on that basis alone. Macmillan and Creelman (2005) recommended calculating d' for the group, essentially calculating d' from the average hit and false alarm rates in each condition, when the number of trials is small. This procedure has the additional advantage, in this case, of reducing the potential bias in the estimate of d' in the unexpected conditions, because the difference between the maximum hit rates in the unexpected and expected conditions would be much smaller, and because the average hit rates are not near the maximum. One value of d' was obtained, using the average hit and false alarm rate, for each age × presentation time × condition (fixed versus mixed) combination. These are referred to as group d'. To allow comparison of group d' between conditions, 95% confidence intervals around each d' were calculated using the exact calculation method described by Miller (1996) (see also Macmillan and Creelman, 2005, Chap. 13).[2] When this method is used, the size of the confidence interval above d' could differ from that below d'. Two group d' were considered different if the 95% confidence intervals did not overlap.[3]

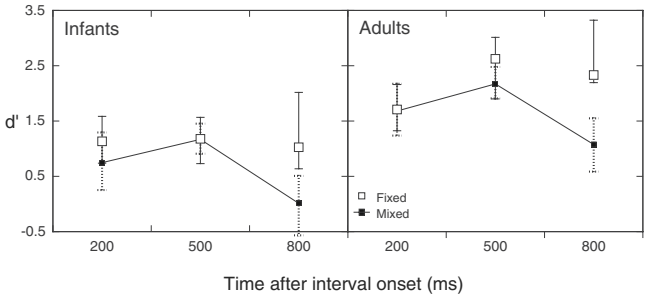


FIG. 4. Group d' as a function of presentation time in experiment 2 for infants and adults. The filled symbols represent the mixed condition; the unfilled symbols represent the fixed condition. The error bars represent the 95% confidence interval calculated using a procedure described by Miller (1996).

B. Results

Figure 4 shows group d' as a function of presentation time, in the fixed (unfilled symbols) and mixed (filled symbols) conditions, for infants (left panel) and adults (right

panel). The 95% confidence intervals in the fixed condition are drawn as dashed lines, while those in the mixed condition are drawn as solid lines. Each subject contributed to the data at all presentation times in the mixed condition, but each subject contributed to the data at only one presentation time in the fixed condition. It is clear that adults were more sensitive to the tone than the infants were in all conditions, although the age groups had equivalent hit rates (infants M =0.76, SD=0.04; adults M=0.78, SD=0.04) in the fixed condition. The important question here, however, was not the age difference in sensitivity, but the effect of presentation time on sensitivity within age groups.

As Fig. 4 indicates, infants and adults showed similar patterns of sensitivity across presentation times in the mixed condition, with apparently higher group d' at the expected 500 ms presentation time than at either 200 or 800 ms unexpected presentation times. Judging from the overlap in confidence intervals, group d' was higher at the 500 ms presentation time than at the 800 ms presentation time in the mixed condition. However, by the same criterion, group d' was not significantly higher at the 500 ms presentation time than at the 200 ms presentation time. Thus, it appears that both infants and adults were significantly more sensitive to a tone that occurred at an expected time than to a tone that occurred at a later unexpected time.

Interpretation of the effects of temporal expectancy depends on the idea that the signal is equally detectable when it is presented at any of the possible presentation times alone (i.e., when it is expected). Thus, it is important to verify that presentation time did not significantly affect sensitivity in the fixed condition. For infants, group d' was about the same when the fixed signal was presented at any of the temporal positions, and the confidence interval (solid lines) at each presentation time overlapped with those at the others. For adults, on the other hand, it appears that group d' is somewhat lower in the fixed condition when presentation time was 200 ms. (Although adults' hit rate was about the same at all presentation times, their false alarm rate was somewhat higher at the 200 ms presentation time.) The confidence interval around group d' at 200 ms overlaps with that at 500 ms, but not with that at 800 ms. Thus, adults were somewhat less sensitive to the signal when it was presented only at 200 ms than they were when it was presented at 800 ms. More importantly, however, they were no more sensitive to the signal when it was presented at 500 ms than they were when it was presented at either of the other times.

Finally, if the listener is more sensitive to the tone at an expected time than at an unexpected time, d' should be higher in the fixed condition than in the mixed condition for the mixed-condition-unexpected presentation times (200 and 800 ms), and d' should be the same in the fixed and mixed conditions for the mixed-condition-expected time (500 ms). In both age groups, it appears that sensitivity to the tones presented 200 or 500 ms after the cue was about the same in the fixed and mixed conditions, but that it was poorer in the mixed condition than in the fixed condition at 800 ms.

Thus, both age groups detected sounds that occurred before the expected presentation time as well as they detected sounds that occurred at the expected presentation time. They detected sounds that occurred after the expected presentation time more poorly than they detected sounds at the expected presentation time.

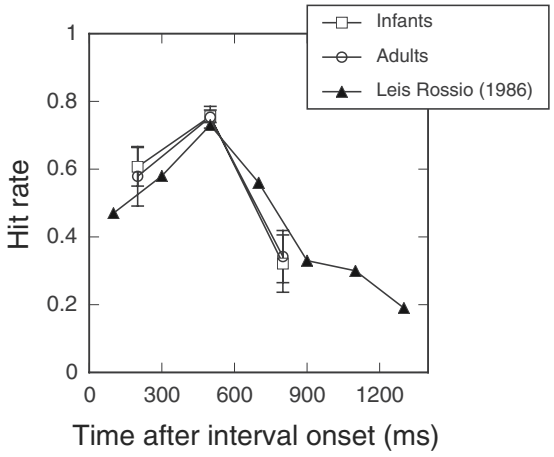


FIG. 5. Hit rate as a function of presentation time in experiment 2 compared to results of Leis Rossio (1986). Mean hit rate ±1 SEM is plotted for infants (squares) and adults (circles). Mean hit rate is plotted for adults tested by Leis Rossio's (1986) (triangles).

C. Discussion

The most notable result of experiment 2 was that infants and adults are similar in their auditory temporal selectivity. Infants, like adults, detect sounds at (and earlier than) an expected time better than they detect sounds at later times. In fact, this effect was remarkably similar in infants and adults in both its size and its dependence on presentation time.

Two previous studies have examined listening windows in adult listeners. In general, the current results are similar to those reported in those two studies. Figure 5 compares the results of Leis Rossio (1986) to those of the current experiment; Leis Rossio (1986) reported hit rate. The current results and those of Leis Rossio (1986) are similar, despite the fact that the signal duration was much longer in the current study and that the subjects in the current study received relatively little training and completed fewer trials. Thus, the listening window effect appears to be relatively robust with respect to variations in stimulus duration and some procedural details. Chang's (1991) results are plotted with the current results in Fig. 6; Chang (1991) reported d'. Note that the decrease in d' when the signal occurs at a time that is later than the expected presentation time is very similar in these and Chang's (1991) results. However, Chang's (1991) subjects show a much steeper drop-off in performance for unexpected presentation times that precede the expected time. Recall that in Chang's (1991) study a contralateral click was presented to indicate the expected presentation time within the observation interval; thus, in that study the subject would not be required to remember the expected presentation time from trial to trial. It is possible that in the absence of the contralateral click used by Chang (1991) to indicate the expected presentation time, listeners open their listening window earlier in the observation interval.

The conditions used in this experiment—long tones and three widely spaced presentation times—do not allow the

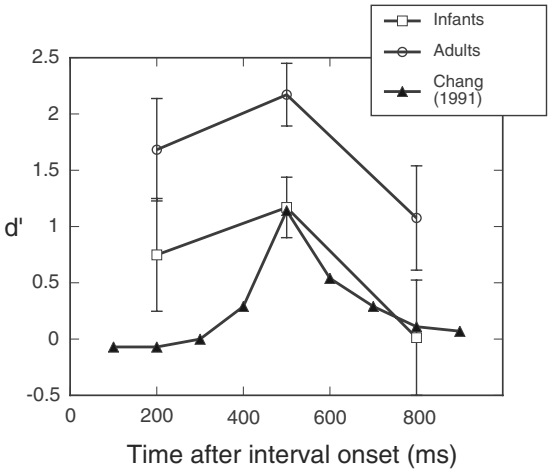


FIG. 6. d' as a function of presentation time in experiment 2 compared to results of Chang (1991). Group d' with 95% confidence interval is plotted for infants (squares) and adults (circles). Mean d' is plotted for adults tested by Chang (1991) (triangles).

duration of the listening window to be precisely estimated for either infants or adults. It may be that infants are broadly similar to adults in the ability to listen selectively in time but still have longer listening windows.

IV. GENERAL DISCUSSION

Infants' ability to listen selectively in time may seem surprising, given their failure to listen selectively in the frequency dimension, as measured using the probe-signal method (Bargones and Werner, 1994), as well as their failure to detect sounds more often when provided with a temporal cue in experiment 1. However, consideration of the literature on the development of visual expectancies makes this finding seem less surprising. In a long series of studies, Haith and co-workers demonstrated that if infants are presented with a series of pictures that occur in different, but predictable locations, infants will quickly begin to look at the expected location of the next picture in the series before the next picture is presented (e.g., Haith *et al.*, 1988; Haith, 1990; Adler and Haith, 2003). In fact, the infants in these studies were 3 months old, younger than the infants tested in the current studies. These studies, among others, suggest that infants develop expectancies about events and that their looking behavior reflects those expectancies. Because the auditory system begins to function prior to the visual system during early development and because auditory function is generally more mature than visual function at any point during development, one would expect infants' ability to use auditory information to be more mature than their ability to use visual information. Thus, the current findings are consistent with the findings of studies of visual development.

The experiment on temporal expectancy indicates that infants can listen selectively in time, but the experiment on cuing effects indicates that a cue does not improve infants' sensitivity. How can these two observations be reconciled? If infants are listening selectively for the tone at the expected time, why does the cue not improve their detection of the tone? No definitive answer to that question is possible on the basis of the current results. We might speculate that the cue has counteractive positive and negative effects on infants' detection. Knowing when to listen would have a positive effect on sensitivity. Three possible negative effects are (1) masking, (2) observer difficulty in distinguishing cue-alone from cue+tone trials, and (3) a type of "distraction" or informational masking.

Werner (1999) reported no forward masking of a short-duration tone by a broadband noise at a masker-tone interval of 200 ms among 6 month old listeners. Thus, it is unlikely that the noise increment would mask a longer duration tone presented 500 ms later. Clearly, the noise-decrement cue could not have forward-masked the tone. If the higher level of the background noise in the noise-decrement cue condition led to neural adaptation, detection of a signal in the period just after the noise level decreased could be difficult, but probably not for 500 ms. Thus, it is unlikely that a sensory masking effect is negatively affecting infant detection in the cue conditions.

Another possibility is that the cue led to a change in the infants' behavior that made it difficult for the observer to distinguish cue-alone trials from cue+tone trials. If, for example, the infant made a clear response to the cue on every trial, that response may have obscured any response to the tone. Recall, however, that the infant/observer team reached a training criterion of 80% correct in both the cue and no-cue conditions. Similarly, on probe trials during the testing phase, the infant/observer team achieved an average hit rate of about 86% in both cue and no-cue conditions.[4] That means that the observer was able to distinguish cue-alone and cue+tone trials reliably, at least when the tone was clearly audible to the infant. Furthermore, the number of trials required to reach training criterion was no different in the cue and no-cue conditions. Is it possible that the infants' response to the tone became less salient when the level of the tone was reduced to near-threshold level? Although that possibility cannot be dismissed, observers, anecdotally, do not report a strong correlation between the salience of the infants' response and the level of the stimulus. Moreover, while infant/observer response bias tended to be somewhat more liberal in the cue conditions, that tendency was not statistically significant. An increase in yes responses during testing would be expected, if the observer were responding on the basis of the infant's response to the cue. Of course, it is still possible that the observer's response bias changed over the course of testing, if the infant's response became an unreliable indicator of cue+tone trials. Thus, while we do not believe that the negative effect of the cue on infants' sensitivity results from the methodology used to measure infants' sensitivity, we cannot eliminate the possibility.

Informational masking has been defined as masking that cannot be explained in terms of peripheral, "energetic," masking (e.g., Durlach *et al.*, 2003). Informational masking can be demonstrated in many adults when the masker frequency is randomly varied over presentations (e.g., Neff and Green, 1987), and in infants and young children even when the masker does not vary in frequency (Leibold and Neff, 2007; Leibold and Werner, 2007). In adults, informational masking is not reported for forward maskers (Neff, 1991), and as previously noted, Werner (1999) reported no forward

masking among 6 month old listeners at a 200 ms masker-tone interval. However, Werner (1999) presented the forward masker and tone several times on each trial, while in the current experiment the tone was presented only once on each trial. It may be that the repeated tone reduced informational masking for the infants in the forward masking experiment (e.g., Kidd *et al.*, 1994). Furthermore, infants and children exhibit informational masking under conditions that do not produce a similar effect in adults (e.g., Werner and Bargones, 1991; Leibold and Neff, 2007; Leibold and Werner, 2007). Thus, the informational masking explanation remains tenable.

The development of frequency selective listening provides an interesting contrast to that of temporally selective listening. Infants and children's pure-tone threshold is higher when other sounds, remote in frequency from the target tone, are presented simultaneously with the tone (e.g., Werner and Bargones, 1991; Leibold and Neff, 2007; Leibold and Werner, 2007). It has been argued that this sort of informational masking is related to a lack of frequency selective listening, as demonstrated in a probe-signal procedure (Bargones and Werner, 1994). That infants' thresholds are affected by remote frequency sounds seems consistent with the idea that infants listen broadly across frequency, even for a narrowband sound. In the temporal domain, it appears that infants can listen selectively for a sound at a specific time. It would be interesting should it prove that they are nonetheless "distracted" by temporally remote sounds.

To return to the question posed at the beginning of this paper, do the effects of temporal uncertainty contribute to the age differences in threshold typically observed in an infant psychoacoustics task? On the basis of the results of experiment 2, the answer would be "no."

V. SUMMARY AND CONCLUSIONS

Providing listeners with an auditory cue indicating the beginning of the observation interval improves adults' tone detection in noise, but not infants'. This pattern of results is observed whether the cue is a decrement in the noise level or a brief increment in the noise level. At the same time, infants, like adults, are more likely to detect a tone that follows the cue by an expected interval than by a longer unexpected interval. Thus, despite the fact that the auditory cue does not increase infants' detection efficiency, it appears that infants develop expectancies about when the signal tone will occur relative to the cue and that these expectancies guide infants' listening. Both infants and adults listen selectively in time.

ACKNOWLEDGMENTS

The authors thank the participants and their parents. Jill Bargones and Jude Steinberg assisted in data collection. The work was supported by grants from NIH, R01 DC00396 and P30 DC04661.

[1]Both the noise increment and the noise decrement began as soon as the observer pushed the button to start a trial. Because the trial actually began with the offset of the noise increment, the delay between the button press and the tone, if it occurred, was 200 ms longer in the noise-increment condition. The performance of a separate group of subjects tested with a 200 ms interval between the offset of the noise increment and the onset of the tone, however, did not differ from that of either group of subjects tested with a 500 ms trial-onset-signal interval.

[2]The calculation of confidence intervals around d' assumes that the outcome on each trial is independent of that on other trials. Our data violate that assumption in the sense that multiple trials come from each subject. However, the violation of this assumption is likely to have a small effect on the calculation of confidence intervals, compared to the potential bias in d' calculation that results from differences in the number of trials available at each presentation time for individual subjects in the mixed condition.

[3]In a parallel analysis, d' was calculated for each subject in each condition, and differences in mean d' were assessed using separate analyses of variance for infants and adults. The results of these analyses were the same as those reported for the group d' analysis, except that for both infants and adults, group d' in the mixed condition at the 200 ms presentation time was significantly better than that at the 800 ms presentation time. This difference would not have influenced interpretation of the results.

[4]Infants' hit rate on probe trials averaged 0.86 (SD=0.16) for the noise-decrement cue, 0.86 (SD=0.17) for the noise-increment cue, and 0.87 (SD=0.15) for no cue. The differences between these means was not statistically significant by one-way ANOVA [$F(2,95)=0.03, p=0.97$].

Adler, S. A., and Haith, M. M. (**2003**). "The nature of infants' visual expectations for event content," Infancy **4**, 389–421.
Allen, P., and Wightman, F. (**1995**). "Effects of signal and masker uncertainty on children's detection," J. Speech Hear. Res. **38**, 503–511.
Arbogast, T. L., and Kidd, G.. (**2000**). "Evidence for spatial tuning in informational masking using the probe-signal method," J. Acoust. Soc. Am. **108**, 1803–1810.
Bargones, J. Y., and Werner, L. A. (**1994**). "Adults listen selectively; Infants do not," Psychol. Sci. **5**, 170–174.
Bargones, J. Y., Werner, L. A., and Marean, G. C. (**1995**). "Infant psychometric functions for detection: Mechanisms of immature sensitivity," J. Acoust. Soc. Am. **98**, 99–111.
Berg, K. M., and Boswell, A. E.. (**1995**). "Temporal summation of 500-Hz tones and octave-band noise bursts in infants and adults," Percept. Psychophys. **57**, 183–190.
Berg, K. M., and Boswell, A. E.. (**1998**). "Infants' detection of increments in low- and high-frequency noise," Percept. Psychophys. **60**, 1044–1051.
Berg, K. M., and Boswell, A. E. (**1999**). "Effect of masker level on Infants' detection of tones in noise," Percept. Psychophys. **61**, 80–86.
Chang, P. (**1991**). "Temporal windows for signals presented at uncertain times," thesis, University of Minnesota, Minneapolis, MN.
Chang, P., and Viemeister, N. F.. (**1991**). "Temporal windows for signals presented at uncertain times." J. Acoust. Soc. Am. **90**, 2248.
Dai, H., and Wright, B. A. (**1995**). "Detecting signals of unexpected or uncertain durations," J. Acoust. Soc. Am. **98**, 798–806.
Durlach, N. I., Mason, C. R., Kidd, G., Arbogast, T. L., Colburn, H. S., and Shinn-Cunningham, B. G. (**2003**). "Note on informational masking (L)," J. Acoust. Soc. Am. **113**, 2984–2987.
Egan, J. P., Greenberg, G. Z., and Schulman, A. I. (**1961**). "Interval of the time uncertainty in auditory detection," J. Acoust. Soc. Am. **33**, 771–778.
Greenberg, G. Z., and Larkin, W. D. (**1968**). "The frequency response characteristic of auditory observers detecting signals of a single frequency in noise: The probe-signal method," J. Acoust. Soc. Am. **44**, 1513–1523.
Haith, M. M. (**1990**). "Progress in the understanding of sensory and perceptual processes in early infancy," Merrill-Palmer Q. **36**, 1–26.
Haith, M. M., Hazan, C., and Goodman, G. S. (**1988**). "Expectation and anticipation of dynamic visual events by 3.5-month-old babies," Child Dev. **59**, 467–479.
Hollich, G., Newman, R. S., and Jusczyk, P. W. (**2005**). "Infants' use of synchronized visual information to separate streams of speech," Child Dev. **76**, 598–613.
Kidd, G., Mason, C. R., Deliwala, P. S., Woods, W. S., and Colburn, H. S. (**1994**). "Reducing informational masking by sound segregation," J. Acoust. Soc. Am. **95**, 3475–3480.
Lappin, J. S., and Disch, K. (**1973**). "Latency operating characteristic: III. Temporal uncertainty effects," J. Exp. Psychol. **98**, 279–285.
Leibold, L., and Neff, D. L. (**2007**). "Effects of masker-spectral variability and masker fringes in children and adults," J. Acoust. Soc. Am. **121**, 3666–3676.
Leibold, L. J., and Werner, L. A. (**2007**). "The effect of masker-frequency variability on the detection performance of infants and adults," J. Acoust.

Soc. Am. **119**, 3960–3970.
Leis Rossio, B. (**1986**). "Temporal specificity: Signal detection as a function of temporal position," thesis, University of Iowa, Iowa City, IA.
Lisper, H. O., Melin, L., Sjoden, P. O., and Fagerstrom, K. O. (**1977**). "Temporal uncertainty of auditory signals in a monitoring task: Effects of inter-signal interval length and regularity on increase in reaction time," Acta Psychol. **41**, 183–190.
Macmillan, N. A., and Creelman, C. D. (**2005**). *Detection Theory: A User's Guide* (Lawrence Erlbaum Associates, Hillsdale, NJ).
Miller, J. O. (**1996**) "The sampling distribution of d′," Percept. Psychophys. **58**, 65–72.
Neff, D. L. (**1991**). "Forward masking by maskers of uncertain frequency content," J. Acoust. Soc. Am. **89**, 1314–1323.
Neff, D. L., and Green, D. M. (**1987**). "Masking produced by spectral uncertainty with multicomponent maskers," Percept. Psychophys. **41**, 409–415.
Richards, J. (**2000**). "Development of multimodal attention in young infants: Modification of the startle reflex by attention," Psychophysiology **37**, 65–75.
Richards, V. M., and Neff, D. L. (**2004**). "Cuing effects for informational masking," J. Acoust. Soc. Am. **115**, 289–300.
Scharf, B. (**1987**). "Focused auditory attention and frequency selectivity," Percept. Psychophys. **42**, 215–223.
Scharf, B., Reeves, A., and Suciu, J. (**2007**). "The time required to focus on a cued signal frequency," J. Acoust. Soc. Am. **121**, 2149–2157.
Schlauch, R. S., and Hafter, E. R. (**1991**). "Listening bandwidths and frequency uncertainty in pure-tone signal detection," J. Acoust. Soc. Am. **90**, 1332–1339.
Schneider, B. A., and Trehub, S. E. (**1985**). "Behavioral assessment of basic auditory abilities," in *Auditory Development in Infancy*, edited by S. E. Trehub, and B. Schneider (Plenum, New York), pp. 101–113.
Schneider, B. A., Trehub, S. E., Morrongiello, B. A., and Thorpe, L. A. (**1989**). "Developmental changes in masked thresholds," J. Acoust. Soc. Am. **86**, 1733–1742.
Werner, L. A. (**1995**). "Observer-based approaches to human infant psychoacoustics," in *Methods in Comparative Psychoacoustics*, edited by G. M. Klump, R. J. Dooling, R. R. Fay, and W. C. Stebbins (Birkhauser, Boston), pp. 135–146.
Werner, L. A. (**1999**). "Forward masking among infant and adult listeners," J. Acoust. Soc. Am. **105**, 2445–2453.
Werner, L. A., and Bargones, J. Y. (**1991**). "Sources of auditory masking in infants: Distraction effects," Percept. Psychophys. **50**, 405–412.
Werner, L. A., and Boike, K. (**2001**). "Infants' sensitivity to broadband noise," J. Acoust. Soc. Am. **109**, 2101–2111.
Werner, L. A., and Marean, G. C. (**1991**). "Methods for estimating infant thresholds," J. Acoust. Soc. Am. **90**, 1867–1875.
Wright, B. A., and Dai, H. (**1994**). "Detection of unexpected tones with short and long durations," J. Acoust. Soc. Am. **95**, 931–938.

Psychometric functions for pure tone intensity discrimination: Slope differences in school-aged children and adults

Emily Buss,[a] Joseph W. Hall III, and John H. Grose
Department of Otolaryngology/Head and Neck Surgery, University of North Carolina School of Medicine, Chapel Hill, North Carolina 27599

(Received 9 May 2008; revised 19 September 2008; accepted 17 November 2008)

Previous work on pure tone intensity discrimination in school-aged children concluded that children might have higher levels of internal noise than adults for this task [J. Acoust. Soc. Am. **120**, 2777–2788 (2006)]. If true, this would imply that psychometric function slopes are shallower for children than adults, a prediction that was tested in the present experiment. Normal hearing children (5–9 yr) and adults were tested in a two-stage protocol. The first stage used a tracking procedure to estimate 71% correct for intensity discrimination with a gated 500 Hz pure tone and a 65 dB sound pressure level standard level. The mean and standard deviation of these tracks were used to identify a set of five signal levels for each observer. In the second stage of the experiment percent correct was estimated at these five levels. Psychometric functions fitted to these data were significantly shallower for children than adults, as predicted by the internal noise hypothesis. Data from both stages of testing are consistent with a model wherein performance is based on a stable psychometric function, with sensitivity limited by psychometric function slope. Across observers the relationship between slope and threshold conformed closely to predictions of a simple signal detection model.
 [DOI: 10.1121/1.3050273]

PACS number(s): 43.66.Fe, 43.66.Ba [RYL] Pages: 1050–1058

I. INTRODUCTION

School-aged children perform more poorly than adults in a wide range of psychoacoustic paradigms including simple tasks, such as detection of a tone (Fior, 1972; Maxon and Hochberg, 1982; Allen and Wightman, 1994), frequency discrimination (Maxon and Hochberg, 1982; Jensen and Neff, 1993), and gap detection (Irwin *et al.*, 1985; Wightman *et al.*, 1989; Trehub *et al.*, 1995), as well as more complex tasks, such as speech recognition under challenging listening conditions (Elliott, 1979; Nabelek and Robinson, 1982) and informational masking (Allen and Wightman, 1995; Oh *et al.*, 2001; Wightman *et al.*, 2003; Hall *et al.*, 2005). While there have been many reports documenting this poorer performance, there is little consensus regarding the underlying cause of elevated thresholds in school-aged children.

Whereas the auditory peripheral physiology is thought to be functionally mature well before the school-aged years (e.g., Abdala and Folsom, 1995), there is evidence of continued development in auditory cortex (Moore and Linthicum, 2007) and more general cognitive development not specifically related to auditory processing (e.g., Gomes *et al.*, 2000) throughout this period. A wide range of nonsensory factors has been proposed to account for the poor psychoacoustic data of school-aged children. Some are based on general cognitive development, such as central noise due to fluctuations in attention, unstable decision criteria or memory limitations (e.g., Allen *et al.*, 1998; Wightman *et al.*, 1989; Allen and Nelles, 1996). Others may be more specific to auditory processing, including an inability to optimally combine or switch among sensory cues (Allen *et al.*, 1989), immature sound source determination (Leibold and Neff, 2007), reduced listening efficiency (Hall and Grose, 1991; 1994), inability to listen in a frequency selective manner (Allen and Wightman, 1994; Stellmack *et al.*, 1997; Lutfi *et al.*, 2003), poor temporal focus of attention (Sutcliffe and Bishop, 2005), and relatively ineffective combination of information distributed over time (Schneider and Trehub, 1992).

Whereas any one or more of these interrelated factors could affect thresholds of school-aged children, some researchers conclude that general developmental factors nonspecific to auditory processing fail to explain the developmental effects observed. For example, Schneider *et al.* (1989) argued that the relative stability of thresholds over time and the small effects of changing the task reward structure indicate that motivation and inattention play little, if any, role in the poorer performance of child observers. While the combination of sensory cues almost certainly plays a role in the performance of young children under some listening conditions, the data of Willihnganz *et al.* (1997) suggest that maturation of listening strategy cannot account for developmental effects observed under complex listening conditions.

Several previous studies have proposed that increased internal noise could account for the reduced sensitivity in psychoacoustic tasks for child observers as compared to adult observers. For example, Schneider and co-workers (1989; 1992) proposed that the sensitivity with which child observers are able to detect a tone in a masking noise could be limited by internal noise. They tested this hypothesis by estimating percent correct for detection of a tone masked by a 1/3-octave band of noise. If higher levels of internal noise were responsible for threshold elevation, this should be re-

[a]Author to whom correspondence should be addressed. Electronic mail: ebuss@med.unc.edu

flected in shallower psychometric function slopes. Contrary to initial predictions, psychometric functions from group data in that study failed to provide compelling evidence of shallower psychometric functions for children as compared to adults (Schneider *et al.*, 1989). Other studies have estimated psychometric functions for individual observers and found the expected increase in slope with age (Allen and Wightman, 1994; Bargones *et al.*, 1995). Whereas the literature is somewhat inconclusive regarding the role of internal noise in the detection of a tone, the present study set out to delineate the role of internal noise for intensity discrimination for a suprathreshold tonal signal.

A recent study by Buss *et al.* (2006a) argued that the poor intensity discrimination thresholds obtained in school-aged children is consistent with increased internal noise. The phrase *internal noise* is sometimes used in a very broad sense, describing any source of inaccuracy in processing that might account for deviations from optimal performance, including factors specific to auditory processing, such as jitter in peripheral encoding of sounds, and other more general functions related to cognitive processing, such as transient attention to the task or variable strategy in utilizing the available sensory cues. Buss *et al.* (2006a) hypothesized that the poor performance of child observers and the inferred elevation of internal noise could be due to variability in the neural representation of intensity rather than nonauditory factors. Briefly, that study showed that stimulus variability within or across intervals affected intensity discrimination thresholds of adults to a greater extent than those of children. This finding is broadly consistent with a simple model incorporating higher levels of internal noise for child than adult observers and predicts that psychometric function slope for intensity discrimination should be shallower for child than adult observers. If this is the case, then the data of children should conform to the basic assumptions of signal detection theory, and, in particular, the assumption that a single psychometric function accurately describes performance over time. In contrast, if variable response strategies or transient inattention were responsible for poor performance, then this would be reflected in greater variability in behavioral data over time.

Stability of the psychometric function underlying performance was assessed in the present study in two ways. The first stage of testing used an adaptive procedure to estimate threshold, incorporating two interleaved tracks that converged on a single value (71% correct). This general strategy was proposed by Leek *et al.* (1991) to identify shifts in the psychometric function characterizing performance over time; using this method, a shift in the psychometric function within a block of trials results in increased statistical dependence across pairs of interleaved tracks. That is, trials from both tracks occurring in close temporal proximity should reflect a slow shift in psychometric function, and therefore be highly correlated. A second approach was to compare replicate estimates of percent correct, with the expectation that variability around those estimates would be elevated if the psychometric function underlying performance shifted over the course of multiple blocks of trials. In both cases, variability of child and adult data was expected to be comparable. If that were not the case, and child data were more variable, this result would undermine the psychometric function fit and complicate interpretation of the slope parameter.

Internal noise in the present study was estimated based on the slope of the psychometric function. Because intensity discrimination for a pure tone signal is thought to be limited by internal noise, the slope of the function directly reflects the magnitude of that noise (Jesteadt *et al.*, 2003). Therefore, if poorer performance of child observers is due to higher levels of internal noise, then this should be reflected in proportionally shallower psychometric function slopes.

II. METHODS

A. Observers

The child group included 16 observers, 6 males and 10 females, ages 4.9–9.4 yr (mean 7.1 yr). The adult group included 11 observers, 5 males and 6 females, ages 19.0–55.2 yr (mean 26.7 yr). All observers had pure tone thresholds equal to or better than 15 dB HL for octave frequencies 250–8000 Hz in the test ear (ANSI, 1996). None of the observers had a history of hearing or ear-related problems.

B. Stimuli

The stimulus was a 500 Hz pure tone, gated on for 500 ms including 50 ms ramps computed as one-half of the period of a raised cosine. In the standard intervals this tone was presented at 65 dB sound pressure level (SPL), and in target intervals it was more intense. The intensity increment in the target interval was defined in units of $10\log(\Delta I/I)$.

C. Procedures

Stimuli were presented in a three-alternative forced-choice procedure, with one randomly selected interval containing the target stimulus. Intervals were separated by 600 ms. An animated sequence marked the three listening intervals, after which the observer was prompted to select the target interval via associated buttons on a touch-screen computer monitor. After each correct answer, a small portion of a cartoon picture was revealed, in the style of a puzzle piece. A progress bar at the top of the screen reflected progress through the run. At the end of a run, the cartoon picture was fully revealed, and it performed a brief animation. Data collection was completed in a single 1 h session for adults, whereas child observers took two 1 h sessions.

Testing included two stages. In the first stage an adaptive 2-down, 1-up tracking paradigm was used to estimate the 71% correct point on the psychometric function (Levitt, 1971). Following the procedures proposed by Leek *et al.* (1991), there was a single tracking rule for all trials up to the second track reversal, wherein the signal level was adjusted in relatively large steps of 4 dB. After the second reversal, the track split into two interleaved tracks, and the stepsize for signal level adjustment was reduced to 2 dB. From that point on, one track determined the course of even numbered trials and the other determined the course of odd numbered trials. These two tracks continued until both had achieved six or more reversals and there was an equal number of trials in

both tracks. Thresholds were calculated as the mean level at reversals 3–6 of each track. The standard deviation across those reversal levels was also recorded for each track. This procedure was performed twice, for a total of four estimates of threshold (m) and four associated standard deviations (s).

Estimates obtained in the first procedure were averaged for each observer, resulting in an overall estimate of threshold ($\bar{m}$) and average standard deviation ($\bar{s}$). These values were interpreted as initial estimates of the 71% correct point and psychometric function slope, respectively. A set of five signal levels was computed for each observer based on these results, defined as $\bar{m}-2\bar{s}$, $\bar{m}-\bar{s}$, $\bar{m}$, $\bar{m}+\bar{s}$, and $\bar{m}+2\bar{s}$. Percent correct was then estimated at each of these five signal levels. Data were collected in ten blocks of 40 trials; within a block there were eight presentations of each of the five signal levels, with those signal levels presented in random order.

III. RESULTS

Prior to data analysis all variables were assessed for normality using a one-sample Komogorov–Smirnov test, both in the aggregate and for the two groups separately. For all parametric statistics reported below, this test failed to reject the null hypothesis that data were normally distributed ($p \geqslant 0.10$). These criteria of normality were also met for statistics on derived parameters (e.g., estimates of psychometric function slope). Data from the adaptive track procedure were analyzed in units of $10\log(\Delta I/I)$, the same units used to adjust signal level. Units of ΔL were used in analysis of the percent correct data in order to facilitate comparison with previous results of Buss *et al.* (2006a). While the best units for representing intensity discrimination is a topic of debate (Buus and Florentine, 1991; Moore *et al.*, 1999), the choice between $10\log(\Delta I/I)$ and ΔL was of little consequence in the present case; repeating analyses in each of these units did not affect the general conclusions of the study.

Statistics below were performed excluding the data of one child observer (C4) whose results were notably poorer than those of her peers. While these data may accurately reflect this observer's psychophysical performance, it was decided to consider the trends exhibited by the other 15 children. Results of the outlier child observer are indicated in the data figures with a filled symbol except where values fell outside reasonable axis limits.

A. Thresholds based on adaptive tracks

Intensity discrimination thresholds obtained in the first stage of testing were higher for child than adult observers, with respective means of 0.20 and −4.96 dB $[10\log(\Delta I/I)]$. This group difference was significant ($t_{24}=5.54$, $p<0.0001$). Within the child data, a regression analysis of the threshold on age resulted in a significant trend for improvement with age ($\beta=-0.86$, $t_{12}=2.20$, $p<0.05$ one-tailed). These results confirm that intensity discrimination is different in adults as compared to school-aged children, with evidence of significant improvement between 5 and 9 yr of age.

Because the adaptive runs split into two interleaved tracks after the second track reversal, it is possible to test the hypothesis that threshold elevation in child observers is due

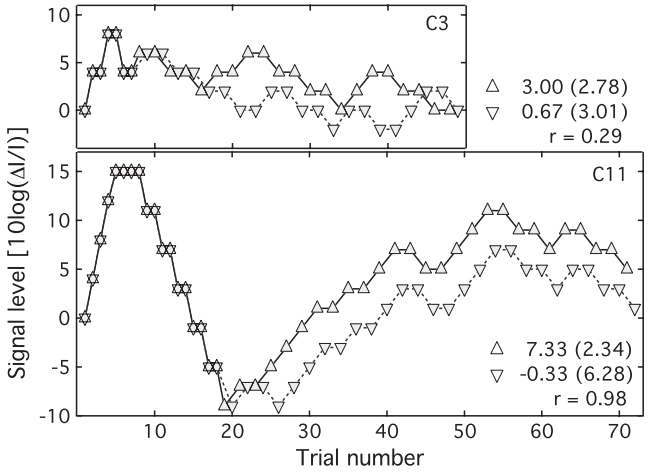


FIG. 1. Signal intensity is plotted as a function of trial number for two observers. The top panel shows data for C3, which is representative of the child data. The bottom panel shows data for C11, a unique case where the two interleaved tracks were highly correlated. Symbols reflect the two tracks, which do not diverge until after the second track reversal. The legend shows the mean and standard deviation associated with each track, as well as the correlation between tracks.

to the instability of the underlying psychometric function, such as might occur with fluctuations in attention or changes in response strategy. Figure 1 shows the example tracks from the child dataset, with signal intensity plotted as a function of trial number. The top panel shows the second adaptive track for observer C3, a 5.7 yr old. In this example, the first phase of the track, characterized by a single track trajectory and large stepsize, lasted for seven trials; after this point, signal level was controlled by the two interleaved tracks. These data reflect some improvement as a function of trial number, but the interleaved tracks are not uniformly parallel, consistent with the interpretation that performance is relatively free from shifts in attention or listening strategy over the course of the trial. The bottom panel shows the second adaptive track for C11, a 7.9 yr old. In this track the signal level rose to the ceiling of 15 dB, fell precipitously, and then gradually increased; the two tracks split apart after the 18th trial but remained largely parallel. It seems likely that this child was confused about the task demands at the outset of the track, adopted a more effective strategy for trials 8–20 and then reverted to a less optimal approach to the task at the end of the track. While the data of C11 are consistent with cognitive dependencies rather than purely auditory limits to performance, this pattern of results was unique among child observers, with the remaining datasets more closely resembling C3 with respect to the relationship of performance across tracks.

The statistical dependence of pairs of interleaved tracks was assessed by comparing the correlation between signal levels in pairs of interleaved tracks for child and adult observers. If instability in the psychometric function were responsible for poor performance in children, and if fluctuations in that function spanned two or more trials, then the intertrack correlation should be greater for child than adult observers. Data were retrieved from disk for the second adaptive run of each observer excluding the last two child observers run on the experiment; files for these two observ-

ers were lost due to computer error. For each dataset, the initial trials using a large stepsize and a single tracking trajectory were excised, leaving only trials in which the small stepsize and the interleaved tracking procedures were used. The first trial from each track was omitted, to allow the two tracks to diverge, and the resulting arrays were used to calculate the Pearson product correlation across tracks. Units of $10\log(\Delta I/I)$ were adopted for this analysis.

The median intertrack correlation was $r=-0.11$ for adults (spanning $r=-0.44$ to 0.58) and $r=0.01$ for children (spanning $r=-0.88$ to 0.98). If the processing underlying performance is variable over the course of a run, this could be reflected in a relatively high correlation across tracks. A t-test assuming unequal variance failed to reject the null hypothesis of equal correlation ($t_{22}=0.42$, $p=0.68$). Within the child group there was no correlation between child age and track correlation ($r=-0.47$, $p=0.10$). These results can be summarized as showing no reliable age effect for intertrack correlation, either across age groups or within the child group, a result that is consistent with comparable stability of the underlying psychometric function over the course of trials.

B. Psychometric function fits

In the second stage of testing, percent correct was measured for a set of five fixed signal levels, based on the mean threshold and standard deviation of each observer's data from the first stage. Results are shown in Fig. 2, where percent correct is plotted as a function of signal intensity in units of ΔL. Circles indicate estimates of percent correct, with 80 trials at each level. Each child observer's data appear in a separate panel, with age of the observer indicated in the lower right corner. The median standard error of the mean (sem) for each group of observers is shown in the legend of Fig. 2. The relationship between sem of child and adult estimates was assessed statistically by applying an arcsine transform to all estimates of percent correct and computing the sem across the ten replicate estimates at each signal level. The resulting values were submitted to a repeated measures analysis of variance (ANOVA), with two levels of GROUP (child, adult) and five levels of SIGNAL (the five signal levels unique to each observer). This analysis revealed a main effect of SIGNAL ($F_{4,96}=4.67$, $p<0.05$), no effect of GROUP ($F_{1,24}=0.19$, $p=0.66$) and no interaction between GROUP and SIGNAL ($F_{4,96}=2.15$, $p=0.08$). This result confirms that the variability in percent correct data of child and adult observers was comparable.

A least squares procedure was used to fit the percent correct data of each observer with a Logit function of the form

$$p(x) = n/3 + (1-n)[(1/3) + (2/3)/(1+e^{-(x-\mu)/k})],$$

where p is the proportion correct (0-1), n is the probability of inattention on any given trial, x is the signal level in ΔL, μ is the midpoint of the function, and k is the slope, with larger values representing shallower functions.[1] Psychometric function fits were good, accounting for 89%–100% of variance in child data and 95%–100% of variance in adult data. Fitted

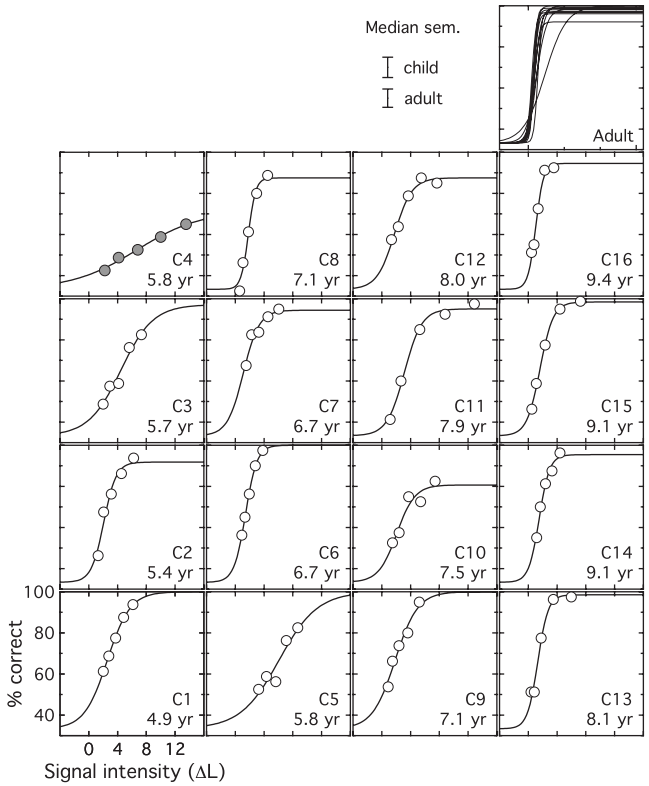


FIG. 2. Percent correct data are shown for individual child observers, plotted in percent correct as a function of intensity increment in units of ΔL. The circles show data, and the lines show data fits. Observer age is indicated in each panel, and panels are arranged with child age increasing from left to right. All adult data fits are shown in the upper right panel. The legend indicates the median standard error of the mean for each group. The filled symbols in the upper left panel highlight the fact that results of observer C4 are markedly poorer than those of other observers, as discussed in the text.

functions are shown with the data for individual child observers, and all adult functions are shown in the upper right panel of Fig. 2. The median value of n was 0.07 for children and 0.02 for adults, associated with upper asymptotes of 95% and 99% correct, respectively. Fitting the data a second time holding n constant at zero reduced the median variance accounted for by 2% in the child dataset and by less than 1% in the adult dataset. Looking across individual data, including the n variable did not significantly increase the quality of the fit for any observer ($p<0.1$, uncorrected), a finding which can be interpreted as a failure to find reliable evidence of random lapses of attention in these data. The decision to retain the variable n in data fits was made to ensure that possible differences in asymptotic performance would not lead to an overestimate of the parameter k (i.e., an excessively shallow fit, as discussed by Allen and Wightman, 1994). Models of inattention and the relationship between inattention and estimated slope will be considered in more detail in Sec. IV.

Consistency of performance over the two stages of testing was systematically evaluated by comparing the 71% point on fitted psychometric functions and 71% thresholds estimated in the adaptive tracking. Figure 3 shows thresholds calculated from psychometric function fits plotted as a function of thresholds based on the adaptive track, both in units of ΔL. Each symbol corresponds to an individual observer's

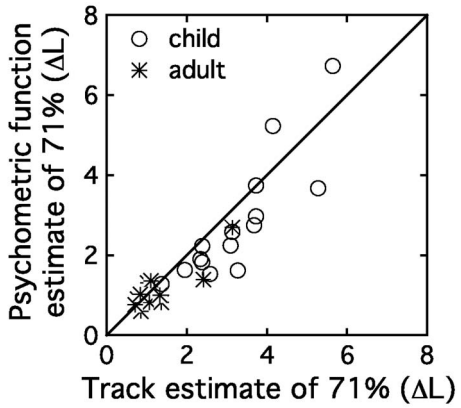


FIG. 3. Threshold estimates based on psychometric function fits are plotted as a function of thresholds estimated in the initial tracking procedure both in units of ΔL. The circles show child data, and the stars indicate adult data. The diagonal line indicates perfect correspondence between these two estimates of 71% correct.

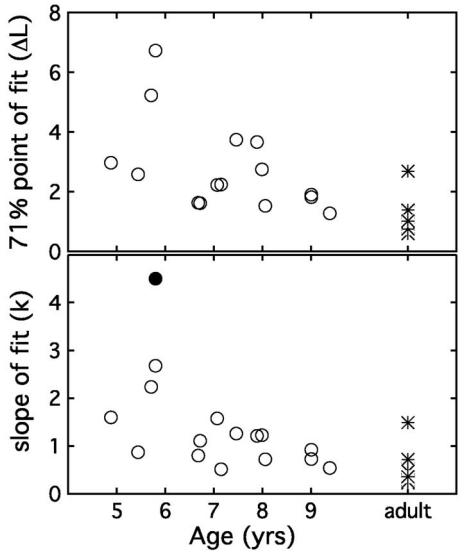


FIG. 4. Two parameters of the fitted psychometric function are plotted with circles as a function of child observer age, with the filled circle indicating the slope estimate of the outlier child observer C4. The threshold value for C4 is not shown because it is greater than the 8 dB axis limit. Adult values are plotted with stars at an arbitrary point on the abscissa. The top panel indicates thresholds for 71% correct in units of ΔL, and the bottom panel shows results for function slope (k).

data, as indicated in the legend, and the diagonal line shows a perfect correspondence between thresholds obtained using the two methods. Thresholds estimated in the two stages of testing are highly correlated when evaluated across all data ($r=0.89$, $p<0.0001$) or just within the child observer group ($r=0.86$, $p<0.0001$). Thresholds were submitted to a repeated measures ANOVA, with two levels of GROUP (child, adult) and two levels of EST (pc, track). There was a main effect of GROUP ($F_{1,24}=17.46$, $p<0.0001$) and a main effect of EST ($F_{1,24}=5.77$, $p<0.05$), but no interaction ($F_{1,24}=1.11$, $p=0.30$). Thresholds based on psychometric function fits were on average 0.3 dB lower than those based on the adaptive tracking procedure; simulations indicate that at least part of this effect could be attributed to the choice of units used to adaptively vary signal level in the track. More important in the present analysis is the fact that there was no interaction between threshold estimation procedure and group. This result supports the assumption that data from these two paradigms both reflect the age effect of interest. Threshold estimates based on psychometric function fits will be used as an index of sensitivity in further analyses, as these estimates are based on a greater number of trials.

Figure 4 shows two parameters based on psychometric function fits plotted as a function of observer age, with adult data plotted at a single arbitrary point on the abscissa. The top panel shows estimates of 71% correct. Child thresholds fall as a function of age, suggesting improvement in performance between 4.9 and 9.4 yr of age. This trend is significant ($\beta=-0.56$, $t_{13}=2.14$, $p<0.05$ one-tailed). The bottom panel of Fig. 4 shows individual estimates of psychometric function slope. Fitted values of slope differ across groups by an average of 0.72, a difference that is significant ($t_{24}=3.45$, $p<0.01$). This effect of age on slope is also evident within the child observer group ($\beta=0.26$, $t_{13}=2.59$, $p<0.05$ one-tailed). We also examined the association between slope and the standard deviation of the track reversal values obtained in the first phase of testing, as it is possible that both could be affected by a common underlying factor. However, the correlation between these metrics did not approach significance ($r=0.001$, $p=0.99$), indicating that this track statistic serves as a very poor predictor of function slope under conditions tested here.

C. Slope as a predictor of threshold

One motivation for the present study was to determine whether psychometric functions for intensity discrimination conform to the prediction of shallower slope in child than adult observers. Buss *et al.* (2006a) modeled intensity discrimination thresholds and found some evidence for greater internal noise in school-aged children than adults; greater spread in the cue underlying the intensity discrimination task would elevate thresholds and produce a shallower psychometric function. As shown in the Appendix, the standard deviation of the underlying cue distribution (σ), reflecting internal noise, can be estimated based on the fitted Logit function slope (k) and then used to generate threshold predictions.

Figure 5 shows thresholds based on psychometric function fits plotted as a function of σ, with symbols indicating individual observer's results. The line shows predicted thresholds. Estimates of σ differed significantly across groups ($t_{24}=3.46$, $p<0.01$), with average values of 0.76 and 1.91 for adults and children, respectively. Threshold predictions were relatively accurate, accounting for 85% of the variance in both child and adult data ($F_{1,24}=136.34$, $p<0.0001$). This fell to 79% when considering just child data ($F_{1,13}=49.88$, $p<0.0001$).

D. Characterizing the outlier

Apart from very high thresholds, the data of the omitted outlier (C4) are very similar to those of other observers. Despite very poor performance, the sem associated with each

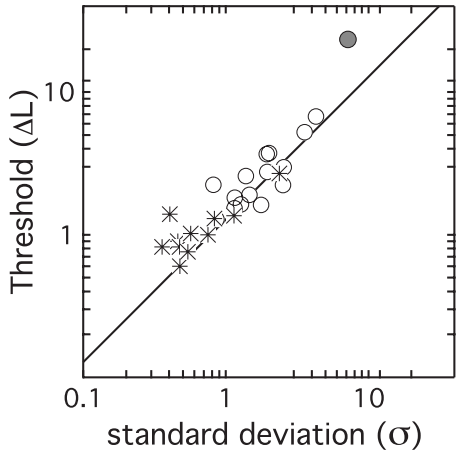


FIG. 5. Thresholds in ΔL are plotted as a function of the standard deviation of the underlying cue distribution (σ), estimated for each observer based on psychometric function slope. Child data are shown with circles, and adult data are shown with stars. The solid line indicates the threshold predicted based on σ, as described in the Appendix.

estimate of percent correct was not unusually large, with values near the 65th percentile relative to other child observers' data. The psychometric function fit was also quite good for this observer, accounting for 98.6% of the variability across the five estimates of percent correct. The relatively low values of percent correct shown in Fig. 2 for this observer stem from a deviation from the testing protocol: adaptive thresholds were judged to be unreasonably high, so signal levels adopted for the second stage of testing were lowered to more typical values in this special case. Assuming the fitted slope estimate of $k=4.5$ is accurate for this observer, internal noise would be estimated as $\sigma=7.18$, corresponding to a threshold prediction of 9.2 dB (ΔL). This result is of interest because, whereas this observer's performance is quite poor, there are no other features of the data that suggest unreliable or unstable underlying psychometric function.

IV. DISCUSSION

The first stage of testing showed a group difference in intensity discrimination thresholds, as well as an effect of age within the child group. Intensity discrimination has previously been reported to be poorer in school-aged children than adults (Fior, 1972; Maxon and Hochberg, 1982; Buss *et al.*, 2006a), though age effects have not been found in all studies (Jensen and Neff, 1993). Examination of the signal levels visited by the interleaved tracks generally supported the assumption that child observers were adopting a reliable listening strategy and using that strategy consistently over the course of an entire track, with one notable exception: the interleaved tracks associated with C11 were highly correlated, highlighting the utility of this procedure. Across observers, however, child data were no more likely to be correlated than those of adults, and the correlation across tracks was not a predictor of threshold. This result is evidence that variability in the psychometric function over the course of a threshold estimation track is probably not responsible for the poor performance of child observers. Previous comparisons of adaptive tracking data of adult and school-aged children also support this conclusion (Allen *et al.*, 1989; Wightman *et al.*, 1989). One caveat to this conclusion is that this method is sensitive only to shifts in the underlying psychometric function which span two or more trials within a track: if these effects were independent across sequential trials, then variability would not be evident in intertrack correlation.

The most interesting aspect of the present data was the finding of shallower psychometric function slopes in child as compared to adult observers. Function fits were quite good in both groups, and variability around each underlying estimate of percent correct was comparable for child and adult observers. These findings are consistent with a simple signal detection model, wherein each observer's performance can be modeled in terms of a single, stable function. Stability of the psychometric function underlying performance over the experiment was further corroborated by the comparison between estimates of 71% correct from the two stages of testing; estimates based on adaptive methods and psychometric function fits were highly correlated both within and across observer groups. The finding of shallower slopes for intensity discrimination in school-aged children than adults is consistent with the shallower slopes for signal detection reported in some (Allen and Wightman, 1994) but not all previous studies (Schneider *et al.*, 1989).

The psychometric function estimates of both threshold and slope were significantly different across groups and were significantly correlated with age within the child observer group. Psychometric function slope in units of ΔL was used to estimate internal noise and generate threshold predictions. These results were broadly consistent with the previous estimates of internal noise determined under comparable stimulus conditions (Buss *et al.*, 2006a). That study reported internal noise estimates of 0.99 and 2.45 dB for intensity discrimination in adult and school-aged child observers, respectively. Estimates based on the present data were 0.76 and 1.91 dB, respectively. Threshold predictions based on estimates of internal noise for individual observers fitted the data quite closely despite substantial individual differences across child observers (see Wightman and Allen, 1992). This result supports the hypothesis that increased internal noise in child observers is responsible for elevated intensity discrimination thresholds, as reflected in shallower psychometric function slopes. The relative stability of performance over time in child data indicates that this age effect is not likely to be due to increased variability in listening strategy or motivation.

One potential difficulty in measuring psychometric function slope is the confounding effect of inattention. Inattention is often modeled as a random, "all-or-none" process. In this model the observer is said to respond randomly without respect to the stimulus on some proportion of randomly distributed trials (n), and on the remainder of trials ($1-n$) to respond according to the psychometric function characterizing sensitivity (e.g., Wightman *et al.*, 1989; Green, 1995). This model produces an upper asymptote below 100% correct. Attempting to fit a function with an asymptote at 100% to data that asymptote below 100% can result in an inaccurately shallow slope estimate as well as threshold elevation

(e.g., Wightman *et al.*, 1989; Wightman and Allen, 1992; Allen and Wightman, 1994). Estimates of n for the present data are consistent with upper asymptotes of 95% and 99% correct for child and adult observers, respectively. These values are consistent with those previously observed under comparable experimental conditions for a detection task (Buss *et al.*, 2001), supporting the use of a three-parameter fit even in the absence of a significant improvement over the two-parameter fit with no inattention ($n=0$). To the extent that the three-parameter fit accounts for individual differences in asymptotic performance, estimates of slope reported here are not affected by all-or-none lapses in attention. Using psychometric function fits with an asymptote at 100% ($n=0$) to estimate internal noise results in larger estimates, with mean values of 0.81 and 2.50 dB for child and adult observers, respectively; this result suggests that any effect of all-or-none inattention may have contributed to, but was not wholly responsible for, the age effect observed in the dataset of Buss *et al.* (2006a).

Wightman and Allen (1992) hypothesized that the psychometric functions characterizing the quality of sensory cues available to adults and children are identical, and that group differences in behavioral thresholds are due solely to random, all-or-none inattention. Modeling in that report indicated that inattention could account for performance if child observers' responses were unrelated to sensory input on about 50% of trials. Pursuing a similar analysis on the present data, psychometric functions fitted to adult data were hypothesized to also characterize sensitivity of child observers, with differences in performance across groups due solely to differences in inattention. For this analysis, values of μ and k were based on mean values fitted to adult data assuming consistent levels of attention ($n=0$, $\mu=1.04$, and $k=0.51$). The value of n was then adjusted to fit the mean child threshold of 2.79 dB. This procedure generated an estimate of $n=0.42$, consistent with an upper asymptote of 72% correct $[n/3+(1-n)]$. Data shown in Fig. 2 are inconsistent with this magnitude of inattention, with the possible exception of observer C4, for whom the best performance measured was 65% correct. This analysis supports the conclusion that all-or-none inattention is insufficient to account for the developmental effects observed in the present experiment. This model also fails to account for the differential development effects observed across auditory tasks (Jensen and Neff, 1993; Dawes and Bishop, 2008).

Several investigators have argued that inattention is unlikely to be random with respect to signal level. A more realistic model might incorporate reduced probability of inattention with increasing signal level (Schneider and Trehub, 1992; Viemeister and Schlauch, 1992; Allen and Wightman, 1994). Whereas signal-dependent attentional effects might be difficult to distinguish from other sources of internal noise (e.g., variability in the neural representation of intensity), it has been argued that signal-dependent inattention could play a larger role in adaptive tracking data than in measurements of percent correct for randomly interspersed signal levels (Schneider and Trehub, 1992; Viemeister and Schlauch, 1992). For example, a series of several "difficult" trials clustered together in time could reduce motivation or confidence in the listening strategy for child observers. If signal-dependent inattention is responsible for the developmental effects in intensity discrimination, and if threshold estimation procedures influence the probability of this type of inattention, then developmental effects might be expected to vary for the two procedures used in the present experiment. Comparison of thresholds across procedures reveals a relatively constant developmental effect, however, a result that fails to support an effect of signal-dependent inattention.

Despite this failure to find evidence of inattention as a source of internal noise, previous work suggests that threshold elevation for intensity discrimination in school-aged children is due to central, nonsensory factors. In a review of the literature, Werner and Marean (1996) argued that the wide variability across individual children and the good performance of some very young observers suggests very early maturation of the sensory representation of intensity cues and implicates nonsensory factors in developmental behavioral effects in school-aged children. This hypothesis is bolstered by the results of Berg and Boswell (2000), who reported adultlike intensity discrimination under some conditions by 3 yr of age. In the present data, the very poor performance of observer C4 is also consistent with nonsensory factors: if a threshold of 9.2 dB were an accurate reflection of sensitivity, one might expect this observer to have significant auditory processing difficulties with naturally occurring sound stimuli, which is not the case. Progressive development of nonsensory factors affecting auditory processing over the school-aged population is supported by both behavioral and physiological studies. A recent paper by Moore and Linthicum (2007) reviews the physiological evidence that, whereas the peripheral auditory system is fully developed early in life, continued maturation of the auditory cortex spans 6–12 yr of age, a process involving plastic changes in response to sensory stimulation and auditory learning. Analogous developmental trends are also evident in the extended developmental effects observed psychophysically (e.g., Dawes and Bishop, 2008). Nonauditory attention also develops extensively over this same age range, with many factors not captured by the all-or-none or even the signal-dependent inattention model considered in analyses of the present data (Gomes *et al.*, 2000).

One aspect of the present paradigm that could highlight the development of nonsensory limits to performance is the choice of a gated stimulus presentation. Durlach and Braida (1969) hypothesized that nonsensory sources of internal noise for gated intensity discrimination can be characterized in terms of memory. For the present three-interval task, an observer might maintain a detailed auditory memory of the stimuli in all three intervals, comparing those memories at the end of the trial; alternatively, the stimulus on each interval could be compared to an internal template of the standard, with the final response based on the interval with stimulus judged to be more intense than the template. It is reasonable to hypothesize that developmental memory effects could limit performance using either of these strategies. If this is the case, then presenting a continuous (as opposed to a gated) standard tone should reduce the developmental effects for intensity discrimination; changing the task in this

way would significantly reduce the memory load, whether child performance was limited by auditory memory for each interval or memory for the standard template. This expectation is consistent with the adultlike performance of 3 yr old observers for intensity discrimination with a continuous standard stimulus observed by Berg and Boswell (2000). Contrary to this expectation, however, initial data suggest that presenting the standard tone continuously may not reduce the developmental effect (Buss *et al.*, 2006b). While memory or some other cognitive factor may play a role in performance, more work is needed to clarify its contribution to the developmental effects observed here.

V. CONCLUSIONS

The present results replicate previous findings of a developmental effect in pure tone intensity discrimination in school-aged children. Psychometric function slopes were shallower in children than adults, consistent with the hypothesis that this age effect is related to internal noise. Predicting performance based on estimates of slope provides a good fit to the data and comparable estimates of internal noise to those published previously (Buss *et al.*, 2006a). Stability of the psychometric function underlying performance appears to be comparable across groups when assessed both via correlation across interleaved adaptive tracks and variability across sequential estimates of percent correct, a result interpreted as showing that fluctuations in attention or listening strategy over blocks of trials is not responsible for the poorer performance of child observers. Future work will evaluate other possible sources of greater internal noise in young school-aged children, including limits to the fidelity of sensory coding as well as memory and signal-dependent attentional factors.

ACKNOWLEDGMENTS

This work was supported by a grant from NIH, R01 DC00397. We thank Lori Leibold for helpful discussions of this work. Ruth Litovsky, Marjorie Leek, and two anonymous reviewers provided helpful comments on a previous draft of this manuscript.

APPENDIX

Many studies report psychometric function slopes in terms of the k-parameter of a Logit fitted to psychophysical estimates of percent correct. The Logit can be defined as

$$p(x) = \frac{1}{1 + e^{-(x-\mu)/k}},$$

where μ is the midpoint of the function, k is the slope parameter, and x is the signal level. In comparison, the cumulative normal distribution can be defined as

$$\Phi(x) = \int_{\infty}^{x} \frac{1}{\sigma\sqrt{2\pi}} e^{-0.5((x-M)/\sigma)^2},$$

where M is the mean of the distribution, σ is the standard deviation of the distribution, and x is the signal level. The Logit is often offered as an approximation of the cumulative normal distribution. Based on Gilchrist *et al.* (2005), the relation between slopes quantified using these two functions can be evaluated as

$$\sigma \approx 4k/\sqrt{2\pi}.$$

As discussed in more details elsewhere (Jesteadt *et al.*, 2003; Buss *et al.*, 2006a), the standard deviation of the underlying cue distribution can be used to generate an estimate of observer sensitivity. Sensitivity is defined as

$$d' = \Delta/\sigma.$$

Because this experiment estimated thresholds for 71% correct using a three-alternative forced-choice task, threshold (Δ) is predicted as

$$\Delta = 1.28\sigma.$$

[1]This form of the Logit function was chosen based on its similarity to the normal cumulative used in previous analysis (Buss *et al.*, 2006a), as described in more details in Appendix. In this form, larger values of k are associated with shallower functions. The Logit can also be expressed in the form

$$p(x) = n/3 + (1-n)[(1/3) + (2/3)/(1 + e^{-k(x-\mu)})],$$

in which case small values of k represent shallower functions.

Abdala, C., and Folsom, RC. (**1995**). "The development of frequency resolution in humans as revealed by the auditory brain-stem response recorded with notched-noise masking," J. Acoust. Soc. Am. **98**, 921–930.

Allen, P., and Nelles, J. (**1996**). "Development of auditory information integration abilities," J. Acoust. Soc. Am. **100**, 1043–1051.

Allen, P., and Wightman, F. (**1994**). "Psychometric functions for children's detection of tones in noise," J. Speech Hear. Res. **37**, 205–215.

Allen, P., and Wightman, F. (**1995**). "Effects of signal and masker uncertainty on children's detection," J. Speech Hear. Res. **38**, 503–511.

Allen, P., Jones, R., and Slaney, P. (**1998**). "The role of level, spectral, and temporal cues in children's detection of masked signals," J. Acoust. Soc. Am. **104**, 2997–3005.

Allen, P., Wightman, F., Kistler, D., and Dolan, T. (**1989**). "Frequency resolution in children," J. Speech Hear. Res. **32**, 317–322.

ANSI (**1996**). *ANSI S3-1996, American National Standards Specification for Audiometers* (American National Standards Institute, New York).

Bargones, J. Y., Werner, L. A., and Marean, G. C. (**1995**). "Infant psychometric functions for detection: Mechanisms of immature sensitivity," J. Acoust. Soc. Am. **98**, 99–111.

Berg, K. M., and Boswell, A. E. (**2000**). "Noise increment detection in children 1 to 3 years of age," Percept. Psychophys. **62**, 868–873.

Buss, E., Hall, J. W., and Grose, J. H. (**2006a**). "Development and the role of internal noise in detection and discrimination thresholds with narrow band stimuli," J. Acoust. Soc. Am. **120**, 2777–2788.

Buss, E., Hall, J. W., and Grose, J. H. (**2006b**). "Processing of intensity in children," paper presented at the Association for Research in Otolaryngology, Baltimore, MD.

Buss, E., Hall, J. W., Grose, J. H., and Dev, M. B. (**2001**). "A comparison of threshold estimation methods in children 6–11 years of age," J. Acoust. Soc. Am. **109**, 727–731.

Buus, S., and Florentine, M. (**1991**). "Psychometric functions for level discrimination," J. Acoust. Soc. Am. **90**, 1371–1380.

Dawes, P., and Bishop, D. V. (**2008**). "Maturation of visual and auditory temporal processing in school-aged children," J. Speech Lang. Hear. Res. **51**, 1002–1015.

Durlach, N. I., and Braida, L. D. (**1969**). "Intensity perception. I. Preliminary theory of intensity resolution," J. Acoust. Soc. Am. **46**, 372–383.

Elliott, L. L. (**1979**). "Performance of children aged 9 to 17 years on a test of speech intelligibility in noise using sentence material with controlled word predictability," J. Acoust. Soc. Am. **66**, 651–653.

Fior, R. (**1972**). "Physiological maturation of auditory function between 3 and 13 years of age," Audiology **11**, 317–321.

Gilchrist, J. M., Jerwood, D., and Ismaiel, H. S. (**2005**). "Comparing and

unifying slope estimates across psychometric function models," Percept. Psychophys. **67**, 1289–1303.
Gomes, H., Molholm, S., Christodoulou, C., Ritter, W., and Cowan, N. (**2000**). "The development of auditory attention in children," Front. Biosci. **5**, D108–D120.
Green, D. M. (**1995**). "Maximum-likelihood procedures and the inattentive observer," J. Acoust. Soc. Am. **97**, 3749–3760.
Hall, J. W., and Grose, J. H. (**1991**). "Notched-noise measures of frequency selectivity in adults and children using fixed-masker-level and fixed-signal-level presentation," J. Speech Hear. Res. **34**, 651–660.
Hall, J. W., and Grose, J. H. (**1994**). "Development of temporal resolution in children as measured by the temporal modulation transfer function," J. Acoust. Soc. Am. **96**, 150–154.
Hall, J. W., III, Buss, E., and Grose, J. H. (**2005**). "Informational masking release in children and adults," J. Acoust. Soc. Am. **118**, 1605–1613.
Irwin, R. J., Ball, A. K., Kay, N., Stillman, J. A., and Rosser, J. (**1985**). "The development of auditory temporal acuity in children," Child Dev. **56**, 614–620.
Jensen, J. K., and Neff, D. L. (**1993**). "Development of basic auditory discrimination in preschool children," Psychol. Sci. **4**, 104–107.
Jesteadt, W., Nizami, L., and Schairer, K. S. (**2003**). "A measure of internal noise based on sample discrimination," J. Acoust. Soc. Am. **114**, 2147–2157.
Leek, M. R., Hanna, T. E., and Marshall, L. (**1991**). "An interleaved tracking procedure to monitor unstable psychometric functions," J. Acoust. Soc. Am. **90**, 1385–1397.
Leibold, L. J., and Neff, D. L. (**2007**). "Effects of masker-spectral variability and masker fringes in children and adults," J. Acoust. Soc. Am. **121**, 3666–3676.
Levitt, H. (**1971**). "Transformed up-down methods in psychoacoustics," J. Acoust. Soc. Am. **49**, 467–477.
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.
Maxon, A. B., and Hochberg, I. (**1982**). "Development of psychoacoustic behavior: Sensitivity and discrimination," Ear Hear. **3**, 301–308.
Moore, J. K., and Linthicum, F. H., Jr. (**2007**). "The human auditory system: A timeline of development," Int. J. Audiol. **46**, 460–478.
Moore, B. C., Peters, R. W., and Glasberg, B. R. (**1999**). "Effects of frequency and duration on psychometric functions for detection of increments and decrements in sinusoids in noise," J. Acoust. Soc. Am. **106**, 3539–3552.
Nabelek, A. K., and Robinson, P. K. (**1982**). "Monaural and binaural speech perception in reverberation for listeners of various ages," J. Acoust. Soc. Am. **71**, 1242–1248.
Oh, E. L., Wightman, F., and Lutfi, R. A. (**2001**). "Children's detection of pure-tone signals with random multitone maskers," J. Acoust. Soc. Am. **109**, 2888–2895.
Schneider, B. A., and Trehub, S. E. (**1992**). "Sources of developmental change in auditory sensitivity," *Developmental Psychoacoustics* (American Psychological Association, Washington, DC), pp. 3–46.
Schneider, B. A., Trehub, S. E., Morrongiello, B. A., and Thorpe, L. A. (**1989**). "Developmental changes in masked thresholds," J. Acoust. Soc. Am. **86**, 1733–1742.
Stellmack, M. A., Willihnganz, M. S., Wightman, F. L., and Lutfi, R. A. (**1997**). "Spectral weights in level discrimination by preschool children: analytic listening conditions," J. Acoust. Soc. Am. **101**, 2811–2821.
Sutcliffe, P., and Bishop, D. (**2005**). "Psychophysical design influences frequency discrimination performance in young children," J. Exp. Child Psychol. **91**, 249–270.
Trehub, S. E., Schneider, B. A., and Henderson, J. L. (**1995**). "Gap detection in infants, children, and adults," J. Acoust. Soc. Am. **98**, 2532–2541.
Viemeister, N. F., and Schlauch, R. S. (**1992**). "Issues in infant psychoacoustics," *Developmental Psychoacoustics* (American Psychological Association, Washington, DC), pp. 191–209.
Werner, L. A., and Marean, G. C. (**1996**). "Factors related to threshold in noise: Intensity, frequency, and temporal resolution," *in Human Auditory Development* (Westview, Boulder, CO), pp. 89–131.
Wightman, F., and Allen, P. (**1992**). "Individual differences," *Developmental psychoacoustics* (American Psychological Association, Washington, DC), pp. 113–133.
Wightman, F., Allen, P., Dolan, T., Kistler, D., and Jamieson, D. (**1989**). "Temporal resolution in children," Child Dev. **60**, 611–624.
Wightman, F. L., Callahan, M. R., Lutfi, R. A., Kistler, D. J., and Oh, E. (**2003**). "Children's detection of pure-tone signals: Informational masking with contralateral maskers," J. Acoust. Soc. Am. **113**, 3297–3305.
Willihnganz, M. S., Stellmack, M. A., Lutfi, R. A., and Wightman, F. L. (**1997**). "Spectral weights in level discrimination by preschool children: Synthetic listening conditions," J. Acoust. Soc. Am. **101**, 2803–2810.

Further examination of pitch discrimination interference between complex tones containing resolved harmonics

Hedwig E. Gockel[a] and Robert P. Carlyon
MRC Cognition and Brain Sciences Unit, 15 Chaucer Road, Cambridge CB2 7EF, United Kingdom

Christopher J. Plack
School of Psychological Sciences, University of Manchester, Manchester M13 9PL, United Kingdom

(Received 25 March 2008; revised 25 November 2008; accepted 5 December 2008)

Pitch discrimination interference (PDI) is an impairment in fundamental frequency (F0) discrimination between two sequentially presented complex (target) tones produced by another complex tone (the interferer) that is filtered into a remote spectral frequency region. Micheyl and Oxenham [J. Acoust. Soc. Am. **121**, 1621–1631 (2007)] reported a modest PDI for target tones and interferers both containing resolved harmonics when the F0 difference between the two target tones (ΔF0) was small. When the interferer was in a lower spectral region than the target, a much larger PDI was observed when ΔF0 was large (14%–20%), and, under these conditions, performance in the presence of an interferer was *worse* than at smaller ΔF0s. The present study replicated the occurrence of PDI for complex tones containing resolved harmonics for small ΔF0s. In contrast to Micheyl and Oxenham's findings, performance in the presence of an interferer always increased monotonically with increasing ΔF0. However, when the interferer was in a lower spectral region than the target (and not vice versa), some subjects needed verbal instructions or modified stimuli to choose the correct cue, indicating an asymmetry in spontaneous obviousness of the correct listening cue across conditions. © *2009 Acoustical Society of America.* [DOI: 10.1121/1.3056568]

PACS number(s): 43.66.Hg, 43.66.Fe, 43.66.Dc, 43.66.Ba [RYL] Pages: 1059–1066

I. INTRODUCTION

Many of the periodic sounds that we encounter in everyday life are broadband. Most current models of pitch perception assume that the derivation of pitch (which, for most harmonic complex tones, corresponds to the fundamental frequency, F0) is based on the combination of information from different frequency regions (Terhardt, 1974; Goldstein, 1973; Moore, 1982; Meddis and Hewitt, 1991; Meddis and O'Mard, 1997). This is consistent with empirical evidence showing that information can be integrated across different frequency regions (Moore *et al.*, 1984; Gockel *et al.*, 2007). However, it has also commonly been assumed that people can listen selectively to specific frequency regions in order to extract the F0s of multiple sources present at the same time. For example, the ability to identify two vowels presented simultaneously is better when the vowels have different F0s than when they have the same F0 (Scheffers, 1983). Although the improvements seen with very small F0 differences (ΔF0s) may be due to within-channel effects (Culling and Darwin, 1994), substantial benefits may be obtained at larger ΔF0s by comparing information from different parts of the spectrum. Models of this effect assume that F0 information can be extracted independently from different parts of the basilar membrane, and are based on the fact that the two vowels have formants at different frequencies, so that the spectrum is dominated at different frequencies by the two different vowels (Assmann and Summerfield, 1990, 1994; Meddis and Hewitt, 1992; de Cheveigné, 1993).

Recently, Gockel *et al.* (2004, 2005a) showed that the ability of listeners to extract the pitch of a harmonic complex can be impaired in the presence of another harmonic complex (the interferer) even when the two tones occupy different frequency regions. This effect was called pitch discrimination interference (PDI) and its existence indicates that listeners have difficulties in independently extracting F0 information from different parts of the spectrum. In these studies, a two-interval two-alternative forced-choice (2I-2AFC) task was used to measure sensitivity (d') to the F0 difference between two target harmonic complexes in the presence of an interferer whose F0 was constant across the two intervals and that was filtered into a remote frequency region. The target contained harmonics that were unresolved by the peripheral auditory system, while the interferer contained resolved harmonics and produced a more salient pitch (see, e.g., Shackleton and Carlyon, 1994). The degree to which F0 discrimination of the target was impaired depended on the similarity of the interferer's F0 and the nominal F0 of the target; the more similar the F0s the worse was performance. The studies also showed that PDI was larger when the interferer contained resolved harmonics than when it contained only unresolved harmonics, suggesting that the relative pitch salience of target and interferer was important. Furthermore, introducing a 400-ms onset and 200-ms offset asynchrony between target and interferer reduced but did not abolish PDI. Even a continuously presented interferer reduced sensitivity to the F0 changes of the target by a small but significant amount.

[a]Author to whom correspondence should be addressed. Electronic mail: hedwig.gockel@mrc-cbu.cam.ac.uk

As mentioned above, in all of these experiments the target complex contained only unresolved harmonics and so its pitch was not very salient. The difference in F0 between the target complexes in the two intervals (ΔF0) was chosen so as to give good but not ceiling performance levels and so was relatively large (mostly between 3.5% and 7.1%). Micheyl and Oxenham (2007) tested whether PDI would occur if both the interferer *and* the target contained resolved harmonics. Each complex was bandpass filtered (8th order Butterworth filter) with fixed 3-dB cutoff frequencies at either 125–625 Hz (LOW region) or 1375–1875 Hz (MID region). The nominal F0 of the target complex was 250 Hz. The interferer F0 was constant across the two intervals of a trial and had an F0 MIDway between the F0s of the target tones. The authors found that, for small ΔF0s (0.875%–1.75%), PDI was observed for targets containing resolved harmonics, irrespective of whether the target was filtered into the LOW and the interferer into the MID region or vice versa. For larger values of ΔF0, no significant PDI was observed when the target was in the LOW and the interferer in the MID region. When the target was in the MID and the interferer was in the LOW region, significantly greater PDI was observed for *large* ΔF0s (14%–20%) than for small ΔF0s. In the absence of the interferer, performance increased monotonically with increasing ΔF0 (being at ceiling at ΔF0s of 14%–20%), but in the presence of the interferer, performance first increased and then, at ΔF0s of 14%–20%, decreased markedly to nearly chance level. For an even larger ΔF0 of 40%, performance recovered but was still somewhat below that observed without an interferer. This nonmonotonic behavior is contrary to our own data obtained in pilot experiments using large ΔF0s values and comparable stimuli and therefore seemed surprising.

Micheyl and Oxenham (2007) suggested two possible explanations for the nonmonotonicity. One is that, in the presence of the LOW-region interferer, subjects listened analytically to the MID-region target tones, i.e., to the frequencies of specific harmonics, rather than listening synthetically to the F0 of the target complexes. The frequency ratio between consecutive audible harmonics of the target complex with a nominal F0 of 250 Hz filtered into the MID region (1375–1875 Hz) was about 1.17. This means that, for ΔF0s of 14% or 20% between the two target tones, the frequencies of harmonics n and $n+1$ in the target complex with the higher F0 were close to the frequencies of harmonics $n+1$ and $n+2$ in the target complex with the lower F0. If subjects listened analytically to shifts in the frequency of an individual harmonic of the target complex across the two intervals, they likely would have compared the "wrong" (noncorresponding) harmonics in the two target complexes. This would lead to systematic errors because, for ΔF0s larger than half the difference between consecutive harmonics, the frequencies of harmonics in the higher-F0 target were *lower* than the frequencies of those harmonics in the lower-F0 target complex that were closest to them in frequency. This could explain the poor performance at ΔF0s of 14%–20% when the target and interferer were filtered into the MID and LOW regions, respectively [for details of the argument see Micheyl and Oxenham (2007)].

The other explanation suggested by Micheyl and Oxenham (2007) was that the pitch of the complex in the LOW region was more salient because it contained lower-numbered and more dominant harmonics and thus suppressed the perception of the pitch of the complex in the MID region. For ΔF0s of 14%–20%, where performance was worst, the F0 difference between the *interferer* and target would be 7%–10%. Micheyl and Oxenham (2007) suggested that a "harmonic sieve" tuned to the F0 of the dominant interferer might be applied and that the harmonics of the targets in the MID region would be suppressed as they "fall between the teeth of the harmonic comb." At smaller ΔF0s the target harmonics fell less precisely between the "teeth," and so this effect would be reduced.

The objectives of the present study were to investigate (i) whether we would replicate the nonmonotonic pattern of results observed by Micheyl and Oxenham (2007) under conditions slightly changed so as to encourage synthetic listening (experiment 1) and, after failing to do so, under identical conditions (experiment 2); (ii) which (if any) of their explanations was more likely to be correct, by increasing the overall number of subjects tested under identical conditions (experiment 2), in order to preclude differences in analytic versus synthetic listening biases across listeners as explanation.

II. EXPERIMENT 1: FILTER REGION VARYING WITH F0

A. Rationale

It is well known that, for complex tones with only a few harmonics excluding the fundamental, many listeners perceive the pitches of individual part-tones rather than the residue pitch (Smoorenburg, 1970; Laguitton *et al.*, 1998; Patel and Balaban, 2001; Schneider *et al.*, 2005). We hypothesized that, in Micheyl and Oxenham's (2007) study, the fixed filter cutoffs (1375 and 1875 Hz) and the large roving range for F0 (six semitones) might have encouraged subjects to listen analytically to the target in the MID region, i.e., to the individual harmonics, rather than listening synthetically to the overall LOW pitch. Considering that the F0 was sometimes as high as 300 Hz, a fixed filter with a bandwidth of 500 Hz did not pass many harmonics and this might well have encouraged subjects to listen to individual harmonics moving in and out of the fixed filter passband when the F0 was changed. To test this hypothesis, we slightly modified the conditions to encourage synthetic listening and also to make them more similar to those used in previous studies on PDI (Gockel *et al.*, 2004, 2005a). There were three changes. First, the mean F0 was varied across trials over a range of of ±10% rather than over the six-semitone range (approximately 41%) used by Micheyl and Oxenham (2007). Second, the cutoff frequencies for the two filter regions were not fixed at absolute frequency values, as in Micheyl and Oxenham (2007), but varied together with, and by the same proportion as, the mean F0 across trials. Third, the spectrum level of a continuous background of pink noise was set to 15 dB (re 20 μPa) at 1 kHz compared to the 12 dB value used by Micheyl and Oxenham (2007).

If Micheyl and Oxenham's (2007) first explanation was correct *and* the present modifications indeed encouraged synthetic listening, one would expect to observe a monotonic relationship between ΔF0 and performance in the presence of an interferer, irrespective of whether the target and interferer were in the MID and LOW regions, respectively, or vice versa. If this were the case, it would also rule out the "suppression" theory suggested by Micheyl and Oxenham (2007). If Micheyl and Oxenham's (2007) first explanation was correct *and* the present modifications were *unsuccessful* in encouraging synthetic listening, or alternatively if their second explanation was correct, then one would expect to observe a similar pattern of results to them.

B. Methods

1. Stimuli

In a 2I-2AFC task, subjects had to discriminate between the F0 of two sequentially presented target complex tones with a nominal F0 of 250 Hz and a fixed difference, ΔF0, between the F0s of the two tones within a block. In one, randomly chosen, interval the target complex had an F0 equal to F0−ΔF0/2, while in the other interval its F0 was F0+ΔF0/2. The values of ΔF0 were 0.225%, 0.45%, 0.9%, 1.8%, 3.6%, 7%, 14%, and 20% of the F0. The target was presented either alone or with a synchronously gated harmonic complex (the interferer) with an F0 that was identical in the two intervals of a trial, and which equaled the arithmetic mean of the F0s of the target in the two intervals. Thus, the interferer's F0 was never identical to the F0 of the simultaneously presented target complex. Each harmonic complex was bandpass filtered (slopes of 48 dB/octave) into one of two frequency regions. The nominal cutoff frequencies (3-dB down points) were 125 and 625 Hz for the LOW region and 1375 and 1875 Hz for the MID region. These frequency regions where chosen following Carlyon and Shackleton (1994) and, according to, e.g., Shackleton and Carlyon (1994), the MID-region complex would contain at least some (the sixth and seventh) resolved harmonics (Plomp, 1964; Plomp and Mimpen, 1968; Moore and Ohgushi, 1993; Shackleton and Carlyon, 1994; Bernstein and Oxenham, 2003). When an interferer was presented with the target, the two complex tones were always filtered into different frequency regions.

The level per component in the passband was always 45 dB sound pressure level (SPL) and components were added in sine phase. The nominal stimulus duration was 400 ms, including 5-ms raised-cosine onset and offset ramps. The silent interval between the two stimuli within a trial was 500 ms. Tones were generated and bandpass filtered digitally; bandpass filtering was achieved with a linear-phase finite impulse response (FIR) filter (order 16 000) implemented in MATLAB with flat passband and linear slopes on a logarithmic frequency scale of 48 dB/octave. They were played out using a 16-bit digital-to-analog converter (CED 1401 plus) with a mean sampling rate of 40 kHz. The actual sampling rate was varied randomly over the range ±10% between trials (this also produced a slight variation in duration and in the filter cutoff frequencies). This was done to discourage subjects from using a long-term memory representation of the F0 of the stimuli and to encourage them to compare the F0 of the target sounds across the two observation intervals within each trial. Stimuli were passed through an antialiasing filter (Kemo 21C30) with a cutoff frequency of 17.2 kHz (slope of 96 dB/octave) and presented monaurally to the subject's left ear, using Sennheiser HD250 headphones. A continuous background of pink noise with a spectrum level of 15 dB (re 20 μPa) at 1 kHz was presented in order to mask possible distortion products and to prevent subjects from relying on possible within-channel cues arising from the interaction of components at the outputs of auditory filters having center frequencies midway between the two spectral regions. Calculation of excitation patterns (following Moore *et al.*, 1997) for the pink noise and for the two complexes together showed that the excitation level of the pink noise at the output of an auditory filter midway between the two regions was about 20 dB above the excitation level of the primary components and thus would mask any interactions between the target and interferer. Subjects were seated individually in a double-walled sound-attenuating booth.

2. Procedure

Subjects had to indicate the interval in which the target sound had the higher F0. They were requested to focus attention on the target and to ignore the interferer as much as possible when it was present. Each interval was marked by a light, and visual feedback was provided after each trial. The value of ΔF0 and the absence or presence of the interferer were fixed within a block of 105 trials. The first five trials were considered as "warm-up" trials and results from those were discarded. For a given condition, usually a block of "target alone" trials was run first, so that subjects were familiar with the timbre of the target. This was followed by two blocks with the interferer present, and another block with the target alone (following an ABBA design), before a new condition was tested. Conditions were run in a pseudorandom order (allowing for the ABBA design within a given value of F0). The frequency regions into which the target and the interferer were filtered were swapped between the first and second halves of a session. The total duration of a session was about 2 h, including rest times. For all subjects, except one (see below), at least 400 trials were run per condition. This corresponded to about six sessions.

3. Subjects

Five subjects participated, one of whom was the first author. They ranged in age from 19 to 44 years, and their absolute thresholds at octave frequencies between 250 and 8000 Hz were within 15 dB of the ISO (2004) standard. All of the subjects had participated in earlier studies on PDI, and all except one had some musical training. Four subjects participated in all conditions, while the fifth was not run for conditions with the smallest and two largest ΔF0s (0.225%, 14% and 20%) due to time constraints. For this fifth subject, 100 trials were collected in the target alone conditions with ΔF0s of 7%, 3.6%, and 1.8%, and in the target plus interferer conditions with ΔF0s of 7% and 3.6% of the F0, for which

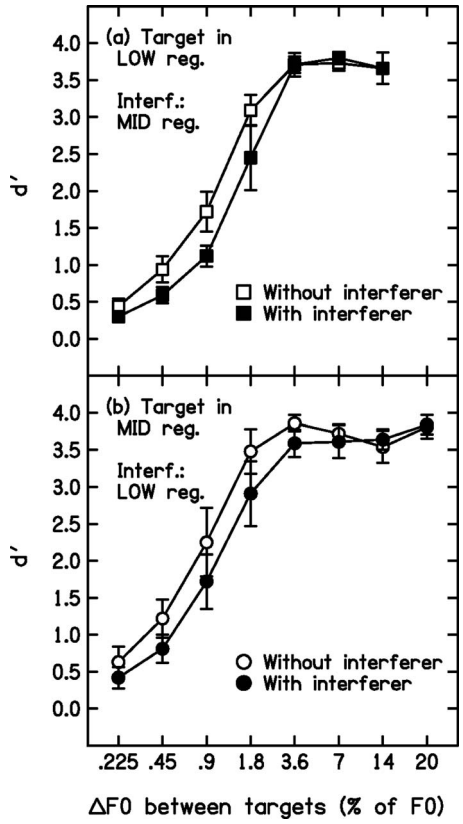


FIG. 1. Mean performance (d') in an F0-discrimination task and the associated standard errors (across subjects) obtained in experiment 1. The nominal F0 of the target tones was 250 Hz. Open and solid symbols show performance in the absence and presence of an additional synchronously presented complex tone (the interferer). Panel (a) shows performance when the target was filtered into a LOW-frequency region from 125 to 625 Hz and the interferer was filtered into a MID-frequency region from 1375 to 1875 Hz. Panel (b) is in (a) but with the target in the MID region and the interferer in the LOW region.

his performance was at ceiling (100% correct). In the remaining conditions at least 200 trials were collected.

C. Results and discussion

Figure 1 shows the mean d' values and the corresponding standard errors averaged across subjects as a function of ΔF0. Open and solid symbols show results in the absence and presence of an interferer, respectively. When the target was in the LOW region and the interferer in the MID region [Fig. 1(a), top], performance increased monotonically with increasing ΔF0, until d′ reached a very high asymptotic value,[1] both in the absence and in the presence of the interferer. Some PDI was observed when performance was above chance for ΔF0 values of 0.45%, 0.9%, and 1.8%. No PDI was observed for ΔF0s of 3.6% and larger. Planned contrasts (one-tailed paired-samples t-tests) indicated that the impairment in the presence of the interferer was significant at ΔF0 values of 0.45% [$t=3.98$, $p<0.01$], 0.9% [$t=3.33$, $p<0.05$] and 1.8% [$t=2.53$, $p<0.05$]. This finding is in good agreement with Micheyl and Oxenham's (2007) data. When the target was in the LOW and the interferer in the MID region, they too reported no significant PDI for ΔF0 values of 3.5% and larger.

When the frequency regions of target and interferer were exchanged, so that the target was in the MID region and the interferer in the LOW region [Fig. 1(b), bottom], the same pattern of results was observed. Performance increased monotonically with increasing ΔF0, both in the absence and in the presence of the interferer. Again, significant PDI was observed when performance was above chance for ΔF0 values of 0.45% [$t=4.63$, $p<0.01$], 0.9% [$t=5.05$, $p<0.01$], and 1.8% [$t=3.16$, $p<0.05$]. No significant PDI was observed for ΔF0s of 3.6% and more. For small values of ΔF0, the present findings agree well with those of Micheyl and Oxenham (2007), who reported significant PDI for values of ΔF0 smaller than 3.5% as long as performance in the absence of an interferer was clearly above chance (so that floor effects were avoided). For larger values of ΔF0, i.e., 7% and more, the present results contrast markedly with those of Micheyl and Oxenham (2007). They reported increased PDI for those ΔF0 values, and near-chance performance for ΔF0s of 14% and 20% in the presence of the LOW-region interferer.

To summarize, for ΔF0s smaller than 7%, the present findings replicate those of Micheyl and Oxenham (2007); when both target and interferer had resolved components, significant PDI was observed when ΔF0 was 1.8% or less (and performance for the target alone was above chance level), independent of whether the target was in the LOW and the interferer in the MID region or vice versa. In contrast, for ΔF0s larger than 3.5%, the results of the two studies differed markedly. Here, no significant PDI was observed, independent of the frequency regions occupied by target and interferer, while Micheyl and Oxenham (2007) reported the largest amount of PDI for ΔF0s of 14% and 20% and a nonmonotonic relationship between ΔF0 and the sensitivity to this difference in F0 when the target and the interferer were filtered into the MID and LOW regions, respectively.

It is possible that, in the present study, some PDI would have been observed for ΔF0s larger than 7% if performance in the absence of the interferer had been below ceiling. However, ceiling effects seem to be an unlikely explanation for the difference in results between the two studies given (i) the size of the PDI effects observed by Micheyl and Oxenham (2007) when the target was in the MID region and (ii) the fact that performance for the target alone for ΔF0s larger than 3.5% was close to ceiling in their study as well.

Note that brief testing using two subjects showed that, for ΔF0=14% the level of the interferer in the LOW region could be increased by up to 25 dB SPL without much effect on performance for F0 discrimination of the target in the MID region. Thus, slight differences in the transfer functions of the headphones used would not explain the different findings of the two studies.

The current findings rule out the suppression theory suggested by Micheyl and Oxenham (2007) and seem to favor their first explanation (both described in the Introduction) for the observed nonmonotonic relationship between ΔF0 and performance in their data in terms of analytical listening. As described in Sec. II A, the present experiment used stimuli that were slightly modified from Micheyl and Oxenham's

(2007) in order to enhance synthetic listening. It is conceivable that these modifications led to the observed monotonic relationship between ΔF0 and the sensitivity to this difference in F0 even when the target and the interferer were filtered into the MID and LOW regions, respectively. Alternatively, the difference across studies might have resulted from the choice of subjects. The second experiment tested these two alternatives by more closely replicating the conditions of Micheyl and Oxenham (2007) with four new and two of the current subjects.

III. EXPERIMENT 2: FILTER REGION FIXED

A. Rationale

The results of the first experiment replicated Micheyl and Oxenham's (2007) finding of a moderate PDI for small ΔF0s, but here performance increased monotonically with increasing ΔF0 in the presence of an interferer, irrespective of the frequency region. The objective of the second experiment was to investigate whether stimuli identical to the ones used by Micheyl and Oxenham (2007) would induce analytical listening and yield similar results to theirs.

Listeners vary naturally in their ability to hear the residue pitch of complex tones containing only a few harmonics when the fundamental component is absent. Although unlikely, the following scenario can be imagined: by chance, none of the five listeners who took part in Micheyl and Oxenham's (2007) experiment 1, where the target and the interferer were filtered into the MID and the LOW regions, respectively, listened synthetically when the interferer was present. Additionally, by chance, all of our five subjects happened to be synthetic listeners. This scenario could lead to the observed differences in the results between the two studies. To decrease the chance of this alternative explanation being applicable, four new subjects participated in experiment 2, in addition to two subjects from the first experiment. If differences between subjects' analytical listening tendency were the reason for the different findings across the two studies, having four new subjects would increase the chance of observing a nonmonotonic pattern of results for at least some of the subjects.

B. Methods

The stimuli and procedure were the same as for experiment 1 with the following exceptions. (1) The mean F0 was varied across trials over a range of ±3 semitones (approximately 41% total range). (2) The cutoff frequencies for the two filter regions were fixed at absolute frequency values of 125–625 Hz for the LOW and 1375–1875 Hz for the MID region. (3) The spectrum level of the continuous background of pink noise was set to 12 dB (re 20 μPa) at 1 kHz. For a given condition, a block of target alone trials was always run first, followed by a block of the same condition with the interferer present. These conditions replicate those of Micheyl and Oxenham (2007).

Only the larger values of ΔF0 were tested, for which the results differed across the two studies. When the target was in the MID region, the values of ΔF0 were 3.6%, 7%, 14%, and 20% of the F0 for all subjects, plus 1.8% for one subject (see below). When the target was in the LOW region, values of ΔF0 were 3.6%, 7%, and 14% of the F0 for all subjects.

The stimuli were played with a fixed sampling rate of 20 kHz and passed through an antialiasing filter (Kemo 21C30) with a cutoff frequency of 8.6 kHz (slope of 96 dB/octave). The roving of the mean F0 was achieved by generating the appropriate stimuli digitally rather than by randomizing the sampling rate. They were bandpass filtered using pairs of cascaded hardware filters (one high pass, one LOW pass, Kemo 21C30 series, each with slopes of 48 dB/octave).

Six subjects participated, one of whom was the first author. They ranged in age from 25 to 47 years, and their absolute thresholds at octave frequencies between 250 and 8000 Hz were within 15 dB of the ISO (2004) standard. Two of the subjects had also participated in experiment 1 of this study, and all except two had some musical training. Stimuli were delivered to the left ear for five subjects and to the right ear for the sixth subject.

All subjects received some practice, until performance remained stable across blocks of trials. The amount of practice needed differed across subjects (see below) and was between 4 and 10 h for the four subjects who were new to the task. In the experiment proper, at least 400 trials were run for each condition and subject.

C. Results and discussion

The solid lines in Fig. 2 show the mean d' values and the corresponding standard errors across subjects as a function of ΔF0. Open and solid symbols show results in the absence and presence of an interferer, respectively. When the target was in the LOW region and the interferer in the MID region [Fig. 2(a), top], performance was at or close to ceiling for all subjects and conditions. A repeated measures two-way analysis of variance (ANOVA) was carried out with factors presence of interferer and ΔF0, using the mean d' value for each subject and condition as input. The ANOVA revealed no significant main effect and no significant interaction. This result was expected and replicates the finding of experiment 1 in the present study and of experiment 2 in Micheyl and Oxenham (2007).

When the target was in the MID region and the interferer in the LOW region [Fig. 2(b), bottom], one of the new subjects performed generally worse than the other five subjects. Her data are plotted using dashed lines, separately from the mean data for the other subjects. While the performance of the other five subjects was at or close to ceiling for all conditions (as in experiment 1), the sixth subject generally had problems with the MID-region target, in spite of extensive training over 10 h. It is important to note that (i) this subject had some musical training and (ii) performance in the presence of the interferer for ΔF0 of 14% and 20% was not worse than for the smaller values of ΔF0. Thus, in spite of her below-ceiling performance, the pattern of results was monotonic. The results of a repeated measures two-way ANOVA, performed on the data for the target in the MID region, showed no significant main effect and no significant

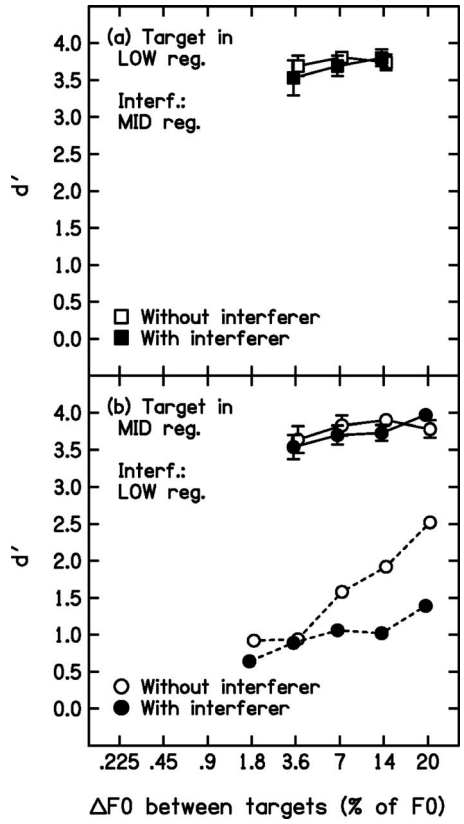


FIG. 2. Mean performance (d') in an F0-discrimination task and the associated standard errors (across subjects) obtained in experiment 2. Otherwise as Fig. 1. Panel (b) (bottom) shows mean results and corresponding standard errors across five subjects (circles connected by solid lines), and separately the results of one "different" worse-performing subject (circles connected by dashed lines).

interaction, independent of whether data from the odd performing sixth subject were included in the analysis or not.

Overall, the pattern of results for experiment 2 was very similar to that for experiment 1, and again contrasts with Micheyl and Oxenham's (2007) findings for large ΔF0s when the target and the interferer were in the MID and LOW regions, respectively. This finding indicates that the differences between the stimuli used in experiment 1, on the one hand, and in experiment 2 and by Micheyl and Oxenham (2007), on the other hand, were unlikely to be the cause of the difference between the present results and those of Micheyl and Oxenham (2007). In other words, it seems unlikely that the exact stimulus conditions chosen by Micheyl and Oxenham (2007) strongly encouraged analytical listening and thus led to the breakdown in performance for ΔF0s of 14% and 20% when the target and the interferer were in the MID and LOW regions, respectively.

It is also very unlikely that, by chance, all nine subjects in the present study were synthetic listeners *and* in Micheyl and Oxenham's (2007) study all five subjects were analytic listeners. It is not possible to calculate exactly the probability for running one experiment with five subjects to find that by chance all five of them are analytical listeners *and* then run another experiment with nine subjects to find that by chance all nine of them are synthetic listeners because the exact distribution of analytic versus synthetic listeners in the general population for the present stimuli is not known. However, to estimate the upper boundary for the probability of the occurrence of these two combined events, we assume that the probability that a randomly drawn subject will listen to the present [and Micheyl and Oxenham's (2007)] stimuli in an analytical way is 0.357 (=5/14), i.e., we assume that, for the present stimuli, this is the proportion of analytical listeners in the population, as this value will maximize the probability of the combined event occurring. With this assumed value, the upper boundary for the probability of the combined event is calculated as 0.000 109 $[=0.357^5\times(1-0.357)^9]$. In other words, the probability that chance differences between the subjects of the two studies are responsible for the observed differences in the results is less than 0.0002. Thus, the current findings seem to rule out Micheyl and Oxenham's (2007) first explanation for their nonmonotonic pattern of results in terms of analytical listening as well as their second explanation, the "suppression theory."

Some insight into a possible reason for their nonmonotonic pattern of results comes from the following observations during the initial training phase of the experiment. Of the four subjects who were new to this task, three (including the "different" worse-performing subject) needed some specific clarification about the cue they were supposed to listen to when the target was in the MID region and the interferer was in the LOW region. In the present study, listeners' training always started with large values of ΔF0 of 20% and 14% of the F0 for both the target alone and target-plus-interferer conditions. When the target was in the LOW region, no subject had difficulties in picking up the correct cue, i.e., after two or three blocks of trials, subjects' performance was at or close to ceiling for those large values of ΔF0, even in the presence of the interferer. However, when the target was in the MID region and the interferer in the LOW region, two subjects showed performance close to chance in the presence of the interferer after two blocks in spite of reaching ceiling in the target-alone condition. For one subject (one without formal musical training), it was sufficient to stress and explain again that the target sound to which attention should be directed was the one with the harsher timbre that sounded less full and smooth than the other sound. This subject's performance was at or close to ceiling for the large values of ΔF0 after five blocks (500 trials), even in the presence of the interferer. A second subject replied to the verbal descriptions with "…but there is no second tone." For this subject, the stimulus duration was increased to 1 s, and at the start of a short practice block of 11 trials, the level of the target sound was increased above that of the interferer by about 30 dB. Over the course of this practice block, the level of the target in the presence of the interferer was reduced by 3 dB after each trial until it reached its "normal" level, which was equal to that of the interferer. This led to an "Aha effect" of recognition on the part of the subject. Two more of these short practice blocks were run with the normal 400 ms duration of the stimuli and slowly reducing level before normal practice continued. After these three short special practice/cueing blocks, the subject performed at ceiling for large ΔF0s. For a third subject (the different worse performer) the verbal de-

scription and special cueing did not have much of an impact; she consistently performed more poorly than the other subjects when the target was in the MID region in both the presence and the absence of the interferer. Apart from her, generally, once a subject knew "what to listen for," when questioned they always reported hearing two tones, and this percept seemed to be irreversible.

These observations indicate that the "correct" listening cue was not always picked up spontaneously. Specifically, when the target and the interferer were in the LOW and MID regions, respectively, all subjects spontaneously listened to the correct cue. In contrast, when the target and the interferer were in the MID and LOW regions, respectively, the correct cue was spontaneously obvious for only some of the subjects, while others needed instructions or special cueing. This indicates an asymmetry in the salience of the correct cue, which might be related to the fact that the dominance region for a complex with F0 of 250 Hz is closer to the LOW than to the MID region (Plomp, 1967; Ritsma, 1967; Moore *et al.*, 1985; Gockel *et al.*, 2005b), i.e., the LOW-region complex had a somewhat more salient pitch than the MID-region complex. Alternatively, the asymmetry could be related to the fact that sounds in music (and everyday life) usually have more energy in the lower harmonics than in the higher harmonics. This could have biased subjects to pay more attention to the lower-frequency range than to the higher-frequency range of the present stimuli.

Either way, the present findings indicate that the nonmonotonic pattern of results observed by Micheyl and Oxenham (2007) was not due to a hard-wired suppression of the perception of the pitch of the complex in the MID region due to the presence of the complex in the LOW region. The present findings also show that it is unlikely that the difference between the results of Micheyl and Oxenham (2007) and the present study was due to analytical listening, caused either by their specific stimuli or by individual differences between subjects across studies. Although we did not observe nonmonotonic performance for any subject or condition, we believe that a possible explanation for the results reported by Micheyl and Oxenham (2007) may lie in our finding that when the target and interferer were in the MID and LOW regions, respectively, some subjects (two out of nine) needed special cueing during training in order to attract their attention to the correct cue. Micheyl and Oxenham (2007) did not report using any such cueing.

IV. SUMMARY AND CONCLUSIONS

F0 discrimination between two sequentially presented 400-ms harmonic target complexes was measured. Their nominal F0 was 250 Hz. The target tones were presented either alone or simultaneously with another harmonic complex, the interferer. The interferer had a constant F0 in the two intervals of the 2I-2AFC task which corresponded to the arithmetic mean of the targets' F0s. The arithmetic mean of the target F0 together with the interferer's F0 varied randomly across trials over a range of $\pm 10\%$ in experiment 1 and over a ± 3 semitone range (approximately 41% total range) in experiment 2. The target and the interferer were always bandpass filtered into separate and remote frequency regions: the LOW (125–625 Hz) and the MID region (1375–1875 Hz). In experiment 1, the filter cutoffs varied proportionally with the mean F0 across trials. In experiment 2, the filter cutoffs were fixed. Both complexes contained some resolved harmonics and thus had a salient pitch. The experimental findings were as follows.

(1) Performance in the absence and in the presence of the interferer was a monotonic function of ΔF0, the frequency difference between the two target tones. This was true regardless of whether the target was in the LOW region and the interferer in the MID region or vice versa.
(2) Some PDI (reduced performance in the presence of the remote interferer) was observed for small values of ΔF0 (1.8% or less). The pattern of PDI was the same whether the target and the interferer were filtered into the LOW and MID regions, respectively, or vice versa.
(3) In both experiments, no significant PDI was found for values of ΔF0 equal to or larger than 3.6%, regardless of whether the target was in the LOW region and the interferer in the MID region or vice versa. Specifically, in experiment 2, where stimuli were identical to those used by Micheyl and Oxenham (2007), no near-chance performance was observed when the target was in the MID region and the interferer in the LOW region for ΔF0 values of 14% and 20%.
(4) In experiment 2, for large values of ΔF0, some of the subjects who were new to the task needed some special training when the target and the interferer were in the MID and LOW regions, respectively, but not when the target and the interferer were in the LOW and MID regions, respectively.

The finding of some PDI for small values of ΔF0 (1.8% or less) between two complex tones containing resolved harmonics replicates Micheyl and Oxenham's (2007) results. The finding of a monotonic relationship between sensitivity and ΔF0 when the target was in the MID region and the interferer in the LOW region contrasts with the results of Micheyl and Oxenham (2007), even in experiment 2, which replicated their conditions closely. However, the observations during the training phase indicate that, for some subjects, there is an asymmetry in the initial salience of the correct listening cue across conditions. Although we never observed the nonmonotonic performance reported by Micheyl and Oxenham (2007), our results do clearly show that cueing subjects to listen to the correct part of a sound can have a strong effect on performance.

ACKNOWLEDGMENTS

This work was supported by EPSRC Grant No. EP/D501571/1. We thank Brian Moore, Alain de Cheveigné, Christophe Micheyl, an anonymous reviewer, and Ruth Litovsky for helpful comments on a previous version of this manuscript.

[1]In occasional cases where performance was correct in 100% of the trials,

and thus the d' value could not be determined, the performance level was adjusted downward to a value of 99.75%. This resulted in a d' value of 3.97.

Assmann, P. F., and Summerfield, A. Q. (**1990**). "Modeling the perception of concurrent vowels: Vowels with different fundamental frequencies," J. Acoust. Soc. Am. **88**, 680–697.

Assmann, P. F., and Summerfield, Q. (**1994**). "The contribution of waveform interactions to the perception of concurrent vowels," J. Acoust. Soc. Am. **95**, 471–484.

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.

Carlyon, R. P., and Shackleton, T. M. (**1994**). "Comparing the fundamental frequencies of resolved and unresolved harmonics: Evidence for two pitch mechanisms?," J. Acoust. Soc. Am. **95**, 3541–3554.

Culling, J. F., and Darwin, C. J. (**1994**). "Perceptual and computational separation of simultaneous vowels: Cues arising from LOW-frequency beating," J. Acoust. Soc. Am. **95**, 1559–1569.

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

Gockel, H., Carlyon, R. P., and Moore, B. C. J. (**2005a**). "Pitch discrimination interference: The role of pitch pulse asynchrony," J. Acoust. Soc. Am. **117**, 3860–3866.

Gockel, H., Carlyon, R. P., and Plack, C. J. (**2004**). "Across-frequency interference effects in fundamental frequency discrimination: Questioning evidence for two pitch mechanisms," J. Acoust. Soc. Am. **116**, 1092–1104.

Gockel, H., Carlyon, R. P., and Plack, C. J. (**2005b**). "Dominance region for pitch: Effects of duration and dichotic presentation," J. Acoust. Soc. Am. **117**, 1326–1336.

Gockel, H. E., Moore, B. C. J., Carlyon, R. P., and Plack, C. J. (**2007**). "Effect of duration on the frequency discrimination of individual partials in a complex tone and on the discrimination of fundamental frequency," J. Acoust. Soc. Am. **121**, 373–382.

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.

ISO 389-8 (**2004**). Acoustics—Reference zero for the calibration of audiometric equipment—Part 8: Reference equivalent threshold sound pressure levels for pure tones and circumaural earphones (International Organization for Standardization, Geneva).

Laguitton, V., Demany, L., Semal, C., and Liegeois-Chauvel, C. (**1998**). "Pitch perception: A difference between right- and left-handed listeners," Neuropsychologia **36**, 201–207.

Meddis, R., and Hewitt, M. (**1991**). "Virtual pitch and phase sensitivity studied using a computer model of the auditory periphery. I: Pitch identification," J. Acoust. Soc. Am. **89**, 2866–2882.

Meddis, R., and Hewitt, M. (**1992**). "Modeling the identification of concurrent vowels with different fundamental frequencies," J. Acoust. Soc. Am. **91**, 233–245.

Meddis, R., and O'Mard, L. (**1997**). "A unitary model of pitch perception," J. Acoust. Soc. Am. **102**, 1811–1820.

Micheyl, C., and Oxenham, A. J. (**2007**). "Across-frequency pitch discrimination interference between complex tones containing resolved harmonics," J. Acoust. Soc. Am. **121**, 1621–1631.

Moore, B. C. J. (**1982**). *An Introduction to the Psychology of Hearing*, 2nd ed. (Academic, London).

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 Peters, R. W. (**1985**). "Relative dominance of individual partials in determining the pitch of complex tones," J. Acoust. Soc. Am. **77**, 1853–1860.

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., and Ohgushi, K. (**1993**). "Audibility of partials in inharmonic complex tones," J. Acoust. Soc. Am. **93**, 452–461.

Patel, A. D., and Balaban, E. (**2001**). "Human pitch perception is reflected in the timing of stimulus-related cortical activity," Nat. Neurosci. **4**, 839–844.

Plomp, R. (**1964**). "The ear as a frequency analyzer," J. Acoust. Soc. Am. **36**, 1628–1636.

Plomp, R. (**1967**). "Pitch of complex tones," J. Acoust. Soc. Am. **41**, 1526–1533.

Plomp, R., and Mimpen, A. M. (**1968**). "The ear as a frequency analyzer II," J. Acoust. Soc. Am. **43**, 764–767.

Ritsma, R. J. (**1967**). "Frequencies dominant in the perception of the pitch of complex sounds," J. Acoust. Soc. Am. **42**, 191–198.

Scheffers, M. T. M. (**1983**). "Sifting vowels: Auditory pitch analysis and sound segregation," Ph.D. thesis, Groningen University, The Netherlands.

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

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.

Smoorenburg, G. F. (**1970**). "Pitch perception of two-frequency stimuli," J. Acoust. Soc. Am. **48**, 924–942.

Terhardt, E. (**1974**). "On the perception of periodic sound fluctuations (roughness)," Acustica **30**, 201–213.

Binaural sluggishness precludes temporal pitch processing based on envelope cues in conditions of binaural unmasking

Katrin Krumbholz[a)]
MRC Institute of Hearing Research, University Park, Nottingham NG7 2RD, United Kingdom

David A. Magezi
MRC Institute of Hearing Research, University Park, Nottingham NG7 2RD, United Kingdom and School Psychology, University of Nottingham, University Park, Nottingham NG7 2RD, United Kingdom

Rosanna C. Moore
MRC Institute of Hearing Research, University Park, Nottingham NG7 2RD, United Kingdom and School of Biomedical Sciences, University of Nottingham, Medical School, Queen's Medical Centre, Nottingham NG7 2UH, United Kingdom

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

(Received 26 October 2007; revised 24 October 2008; accepted 3 December 2008)

Binaural sluggishness refers to the binaural system's inability to follow fast changes in the interaural configuration of the incoming sound stream. Several studies have measured binaural sluggishness by measuring signal detection in conditions of binaural unmasking when the interaural configuration of the masker is changed over time. However, it has been shown that, in conditions of binaural unmasking, binaural sluggishness also affects the perception of temporal changes in the properties of the signal (i.e., its frequency or level) and not just in the interaural configuration of the masker. By measuring the temporal modulation transfer function for sinusoidally modulated noise presented in conditions of binaural unmasking, the first experiment of the current study showed that, due to binaural sluggishness, the internal representation of binaurally unmasked sounds conveys little or no information about envelope fluctuations with rates within the pitch range (i.e., above 30 Hz). The second experiment measured the masked detection threshold for musical interval recognition in binaurally unmasked harmonic tones and showed that, in conditions of binaural unmasking, pitch wanes when the harmonics become unresolved by the cochlear filters. These results suggest that binaural sluggishness precludes temporal pitch processing based on envelope cues in binaurally unmasked sounds. © *2009 Acoustical Society of America.* [DOI: 10.1121/1.3056557]

PACS number(s): 43.66.Hg, 43.66.Pn [MW] Pages: 1067–1074

I. INTRODUCTION

The human binaural system processes interaural differences in the temporal fine structure or the temporal envelope of sounds with an accuracy of only a few tens of microseconds (for review, see Durlach and Colburn, 1978; Bernstein, 2001). At the same time, the binaural system has been shown to be sluggish to respond to changes in interaural configuration over time. For instance, listeners are unable to tell the direction of a sound source moving back and forth between left and right when the rate of the movement exceeds only a few hertz (Blauert, 1968, 1970; see also Grantham and Wightman, 1978; Grantham, 1982, 1984). This binaural sluggishness is widely thought to reflect the minimum integration time of the binaural system and is commonly modeled as a moving-average filter, which integrates the instantaneous output of the binaural processor according to a temporal weighting function, referred to as the "binaural temporal window" (e.g., Culling and Colburn, 2000). The duration of the window has been estimated to be in the range of several tens to a few hundreds of milliseconds, depending on the experimental conditions (Grantham and Wightman, 1979; Kollmeier and Gilkey, 1990; Akeroyd and Summerfield, 1999a; Culling and Summerfield, 1998); this is more than an order of magnitude longer than the duration of the monaural integration window (Holube *et al.*, 1998; see also Viemeister, 1979; Plack and Moore, 1990; Wiegrebe and Krumbholz, 1999).

Several studies have measured binaural sluggishness using signal detection tasks in which the interaural temporal configuration (interaural phase or time difference) of the signal differs from that of the masker and which thus invoke a binaural release from masking of the signal (for a review on binaural unmasking, see Durlach and Colburn, 1978). Most of these studies measured the effect of changing the interaural configuration of the masker over time. However, some studies have shown that, in conditions of binaural unmasking, binaural sluggishness also affects the perception of temporal changes in the signal (i.e., changes in the signal fre-

[a)]Author to whom correspondence should be addressed. Electronic mail: katrin@ihr.mrc.ac.uk

quency or level), even if these changes are diotic and are thus not perceived as changes in the signal's spatial attributes. Hall and Grose (1992), for instance, measured the detectability of a temporal gap in a pure-tone signal presented antiphasically (interaural phase difference of 180° or π) in a diotic noise masker (N_0S_π; see Durlach and Colburn, 1978). When the signal was presented at levels below the masked threshold for the homophasic signal (N_0S_0; signal and masker presented diotically) and was thus perceived only through binaural channels, the gap detection threshold for the antiphasic signal was more than an order of magnitude larger than that for a homophasic signal with the same sensation level (SL). Similarly, Culling and Colburn (2000) found that performance in discriminating the direction of frequency change in fast sequences of pure tones was severely degraded in conditions of binaural unmasking compared to diotic masking conditions. These results are inconsistent with suggestions that binaural sluggishness observed in conditions of binaural masking release might reflect sluggishness in the binaural system's ability to adapt its processing parameters to a changing masker (Yost, 1985), because, in both studies, the interaural configuration of the masker was constant over time. It is widely believed that a binaurally unmasked signal is perceived by virtue of the decrease in interaural correlation that it produces in the diotic masker (Durlach *et al.*, 1986). Culling and Colburn (2000) proposed that temporal information conveyed by a binaurally unmasked signal is perceived by temporal fluctuations in this signal-induced interaural decorrelation and that binaural sluggishness means that only the slow rates of these fluctuations are audible.

According to this logic, pitch perception based on temporal cues should be strongly impaired if not impossible in conditions of binaural unmasking. The auditory-nerve responses to most tonal stimuli convey both temporal and spectral (place) information on pitch. The timing of action potentials in response to pure tones, for instance, is phase locked to the stimulating waveform. At the same time, pure tones produce a peak in the distribution of activity along the tonotopic array (excitation pattern). Harmonic tones consist of multiple frequency components; when the components are resolved by the cochlear filters, they produce peaks in the excitation pattern similar to pure tones and the auditory-nerve responses are phase locked to the component frequencies. However, at higher spectral frequencies or lower fundamental frequencies, the harmonic components become unresolved. In this case, the only cue to pitch are the periodic modulations in the temporal envelope of the auditory-nerve responses, which arise as a result of the interactions between harmonic components within individual cochlear filters (beating). If these purely temporal pitch cues were eliminated by binaural sluggishness, the pitch of unresolved harmonic tones would be expected to be inaudible in conditions of binaural unmasking. The current study was designed to test this prediction.

II. EXPERIMENT 1

The findings of Hall and Grose (1992) on temporal gap detection in binaurally unmasked sounds suggest that the envelope modulations in the auditory-nerve responses to spectrally unresolved harmonic tones would be inaudible in conditions of binaural unmasking. The first experiment was designed to test this hypothesis. The experiment measured the temporal modulation transfer function (TMTF) of a bandpass noise signal presented in conditions of binaural unmasking. The TMTF is a measure of the fidelity with which a system transmits sinusoidal amplitude modulations at different modulation rates. Viemeister (1979) applied the TMTF to measure the resolution of monaural temporal processing (see also Strickland and Viemeister, 1997). The TMTF is obtained by measuring the threshold for detecting sinusoidal amplitude modulation in a noise signal as a function of modulation rate. Using a noise rather than a sinusoidal carrier excludes spectral cues when the spectral sidebands produced by the modulation become resolved by the cochlear filters.

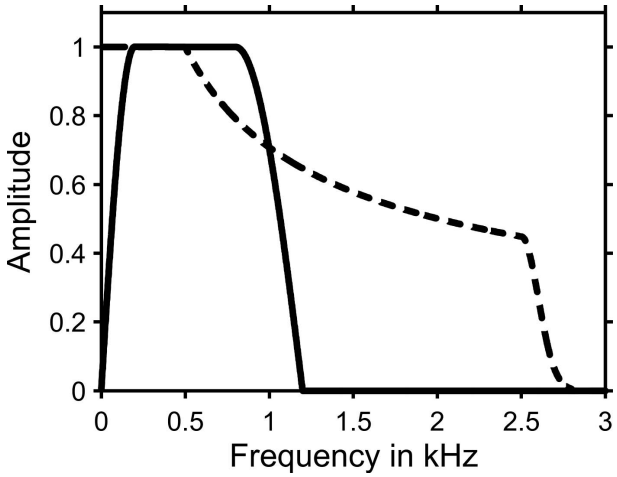


FIG. 1. Spectral envelope of the masker (dashed line) and signal (solid line) in experiments 1 and 2, plotted on an arbitrary linear scale.

In the current experiment, the noise signal that carried the amplitude modulation that was to be detected was presented either homophasically (N_0S_0) or antiphasically (N_0S_π) in a diotic noise masker. To ensure that the modulation was not detected by changes in the spatial attributes of the signal, the amplitude modulation was in phase at the two ears, even when the carrier was presented antiphasically.

A. Methods

1. Stimuli

All stimuli were generated digitally with a sampling rate of 15 kHz and a 16 bit amplitude resolution using Tucker Davis Technologies System 2 (TDT, Alachua, FL, USA) and MATLAB (The Mathworks, Natick, MA, USA). The noise signal was bandpass filtered between 0.1 and 1 kHz (3 dB down) to limit the signal passband to the spectral region where the masking level difference between the homophasic and antiphasic masking conditions would be sizeable. The edges of the signal passband were rounded according to a quarter cycle of a cosine function with a width of 0.2 kHz on the lower edge and 0.4 kHz on the upper edge (see solid line in Fig. 1). In the modulation detection threshold measurements, the modulated signal was filtered after being modu-

lated, and the modulated and unmodulated signals were set to the same overall level. The spectrum level of the noise masker was flat up to 0.5 kHz, from where it decreased by 3 dB/octave to produce a roughly constant level of excitation across frequency (Zwicker and Fastl, 1990). The masker was lowpass filtered at 2.583 kHz (3 dB down); its upper spectral edge was shaped according to the positive half of a normal distribution function with a −6 dB width of 0.235 kHz (see dashed line in Fig. 1). The stimuli were filtered in the frequency domain using the TDT AP2 signal processor. Both signal and masker had a duration of 800 ms and were gated on and off simultaneously with 25 ms squared cosine ramps. The masker had an overall level of 60 dB SPL (sound pressure level).

The stimuli were digital-to-analog converted (TDT DD1), antialiasing filtered with a 4.5 kHz cutoff and a 48 dB/octave slope (VBF 8, Kemo, Beckenham, UK), attenuated (TDT PA4), passed through a headphone buffer (TDT HB6), and presented via headphones (K240 DF, AKG, Vienna, Austria) to the participant, who was seated in a double-walled, sound-attenuated room (IAC, Winchester, UK). The headphones were calibrated with a Brüel&Kjær (Nærum, Denmark) $\frac{1}{2}$ in. microphone (type 4134), artificial ear (type 4153), and measuring amplifier (type 2610).

2. Outline and procedure

The experiment consisted of three parts. In the first part, the detection threshold of the noise signal was measured in homophasic and antiphasic masking conditions in order to allow specification of the SLs of the signal for the modulation detection measurements. The detection thresholds were measured with a two-interval, two-alternative forced-choice (2I2AFC) procedure. Each trial consisted of two 800 ms observation intervals, both of which contained the masker. Only one of the intervals (chosen randomly) contained the signal, and the task was to identify the signal interval by pressing one of two response buttons. The signal was unmodulated in the detection threshold measurements. The intervals were separated by a 500 ms silent gap, and visual feedback was provided at the end of each trial. The signal level was decreased after three consecutive correct responses and was increased after each incorrect response to track 79% correct performance (Levitt, 1971). The step size of the level increments and decrements was 5 dB up to the first reversal in level, 3 dB from there to the second reversal, and 2 dB for the rest of the ten reversals that made up each threshold run. Each threshold estimate was the average of the signal level at the last eight reversals. At least three such threshold estimates were averaged to obtain the threshold value for each of the two masking conditions. The order in which the conditions were tested was counterbalanced across the three threshold runs.

In the TMTF measurements, the antiphasic signal was presented at a fixed SL of 10 dB; the purpose of the second part of the experiment was to determine the level of the homophasic signal that would produce a similar modulation detection threshold as the 10 dB SL antiphasic signal at the lowest modulation rate tested (4 Hz). For that, the modulation detection threshold of the 10 dB SL antiphasic signal was first measured at 4 Hz using a similar 2I2AFC procedure as used for the detection threshold measurements. In this case, both intervals contained both the masker and a signal. Only one of the signals (chosen randomly) was amplitude modulated, and the task was to identify the interval containing the modulated signal. The adaptive parameter was the modulation depth, m, which was incremented and decremented by the same decibel steps as the signal level in the detection threshold measurements (5, 3, and 2 dB). All other parameters of the procedure were the same as in the detection threshold measurements. Then, the level of the homophasic signal that would yield the same modulation detection threshold as the antiphasic signal at 10 dB SL and 4 Hz was determined. The task was identical to that used in the previous test (identify the interval containing the modulated signal). In this case, however, the modulation depth, m, was fixed at the individual modulation detection threshold for the antiphasic signal at 10 dB SL and 4 Hz for each participant, and the adaptive parameter was the signal level. The step sizes of the level increments and decrements were the same as in the previous tests (5, 3, and 2 dB), as were all other parameters of the procedure.

In the third part of the experiment, the TMTF was measured for both homophasic and antiphasic masking conditions. In the homophasic condition, the signal level was set to the value determined in the previous test for each participant; the antiphasic signal was presented at 10 dB SL. The task and procedure were the same as those used in the modulation detection threshold measurements in the second part of the experiment. The adaptive parameter was the modulation depth, m. The modulation rate was fixed throughout each threshold run. At least three threshold estimates for each condition were collected in a counterbalanced order.

3. Participants

Four participants (2 females, 2 males) were tested: author KK, two colleagues with prior experience in psychoacoustic testing, and one student who was paid at an hourly rate. The participants were between 19 and 31 years of age and had no reported history of hearing impairment or neurological disease.

B. Results and interim discussion

Both the detection and modulation detection thresholds measured in experiment 1 were very consistent across participants, and so, only the average thresholds are shown in the figures. The detection thresholds revealed a robust binaural masking level difference (BMLD) of, on average, 13.2 dB [$t(3)=12.424$, $p=0.001$; compare open bars in Fig. 2].

At 10 dB SL (corresponding to an absolute level of, on average, 48.8 dB SPL; see dashed horizontal line in Fig. 2 and Sec. II A 2), the antiphasic signal yielded an average modulation detection threshold of $-20\log_{10}(m)=13.8$ dB at the 4 Hz modulation rate. Note that, at the level of the antiphasic signal (48.8 dB SPL), the homophasic signal would have been below the detection threshold (compare the open bars and dashed horizontal line in Fig. 2). The signal level

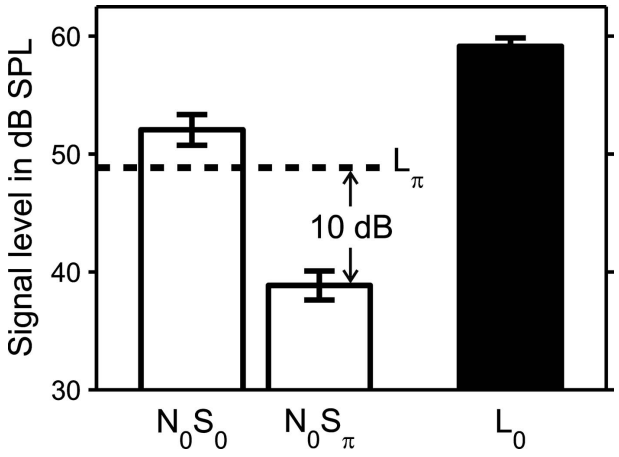


FIG. 2. Average detection thresholds of four participants (open bars) for the homophasic (N_0S_0) and antiphasic (N_0S_π) bandpass noise signals in experiment 1 in dB SPL. The masker had an overall level of about 60 dB SPL. The horizontal dashed line (L_π) shows the average presentation level of the antiphasic signal in the TMTF measurements, corresponding to a SL of 10 dB; L_π is not represented by a bar since it is a derived, rather than a measured quantity. The black bar (L_0) shows the level of the homophasic signal required to yield the same modulation detection threshold as the 10 dB SL antiphasic signal at the 4 Hz modulation rate [$-20\log_{10}(m)$ = 13.8 dB, on average, where m is the modulation index], which corresponded to an average SL of 7.1 dB. Error bars show the standard error of the mean.

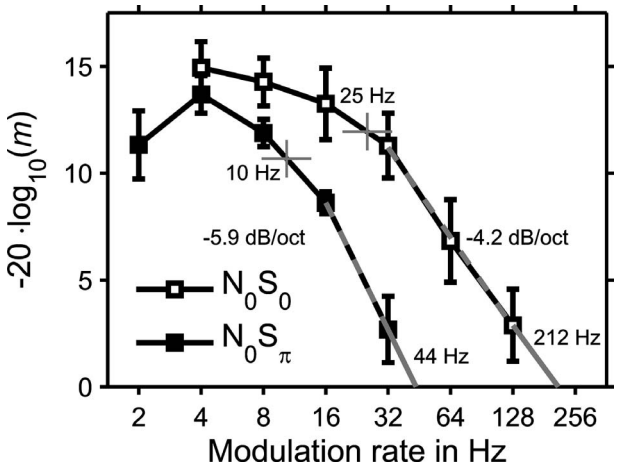


FIG. 3. Average TMTFs of four participants for a homophasically (N_0S_0, open symbols) and antiphasically (N_0S_π, filled symbols) presented bandpass noise signal. The modulation detection thresholds are expressed as $-20\log_{10}(m)$, where m is the modulation index, and plotted as a function of the modulation rate in Hz. Error bars show the standard error of the mean. The gray crosshairs show the 3 dB down points (measured from the modulation threshold at 4 Hz by linear interpolation) of the TMTFs. The dashed gray lines show the slopes of the functions, determined by linear regression of the modulation thresholds at 16 and 32 Hz for the antiphasic condition and at 32, 64, and 128 Hz for the homophasic condition. The short solid gray lines at the bottom of the graph show the extrapolated axis intercepts of the functions and represent the modulation rates where the modulation becomes inaudible.

needed to produce the same modulation detection threshold at 4 Hz in the homophasic condition amounted to an average of 59.1 dB SPL (filled bar in Fig. 2), corresponding to a SL of 7.1 dB. Thus, for the 4 Hz modulation to be detectable at the same modulation depth of, on average, 13.8 dB, the signal had to be presented at an average of 7.1 dB SL in the homophasic and 10 dB SL in the antiphasic masking condition.

Above 4 Hz, modulation detection performance for the antiphasic signal was considerably worse than for the homophasic signal even though the antiphasic signal was presented at a higher SL (10 dB compared to an average of 7.1 dB). For the antiphasic signal (filled symbols in Fig. 3), the decline of the TMTF towards higher modulation rates was steeper and started at a lower modulation rate than for the homophasic signal (open symbols). The TMTF for the antiphasic signal crossed the 3 dB down point at a modulation rate of about 10 Hz (measured from the modulation threshold at 4 Hz by linear interpolation between neighboring data points), compared to 25 Hz for the homophasic condition (see a gray crosshairs in Fig. 3). The slope of the TMTF (assessed through linear regression of the modulation thresholds at 16 and 32 Hz for N_0S_π and at 32, 64, and 128 Hz for N_0S_0; see gray dashed lines in Fig. 3) amounted to 5.9 dB/octave for the antiphasic condition, compared to 4.2 dB/octave for the homophasic condition. For the antiphasic signal, the modulation detection task became very difficult at a 32 Hz modulation rate, and at 64 Hz, it was impossible to measure the modulation threshold at all. For the homophasic signal, the modulation threshold was measurable up to 128 Hz. Linear extrapolation of the TMTFs suggests that modulation became undetectable at 44 Hz in the antiphasic condition, compared to 212 Hz in the homophasic condition (see gray solid lines in Fig. 3).

These differences were confirmed by statistical analysis. A two-way repeated-measures analysis of variance (ANOVA) of the individual modulation detection thresholds with factors masking condition (homophasic and antiphasic) and modulation rate (4, 8, 16, and 32 Hz) revealed significant main effects of masking condition [$F(1,3)=31.218$, $p=0.011$] and modulation rate [$F(3,9)=31.727$, $p<0.001$], as well as a significant interaction between masking condition and modulation rate [$F(3,9)=11.014$, $p=0.002$]. While the homophasic modulation detection threshold at 4 Hz was slightly (about 1.3 dB) smaller than the antiphasic one, this difference was statistically insignificant [$t(3)=1.839$, $p=0.163$]. The interaction between masking condition and modulation rate shows that the differences between the TMTFs for the homophasic and antiphasic conditions were not merely due to a difference in sensitivity, but reflect a difference in the time constants limiting the perception of amplitude modulation in these conditions.

In the homophasic condition, the TMTF would be assumed to have been at least partly limited by peripheral factors. In the current experiment, the signal was filtered to the low-frequency region, where the cochlear-filter bandwidths are relatively narrow. Passing a noise carrier through a narrowband filter produces random envelope fluctuations at the filter output, which mask the sinusoidal amplitude modulation that is to be detected (Dau *et al.*, 1999; Stellmack *et al.*, 2005). Moreover, the narrow cochlear filters attenuate the spectral sidebands of the modulation and thus reduce the effective modulation depth at the filter output. In the antiphasic condition, these peripheral limitations appear to have been overridden by the effect of binaural sluggishness.

The results from experiment 1 confirm the conjecture that the internal representation of binaurally unmasked sounds conveys little or no information about temporal envelope fluctuations with rates above 30 Hz, which would be necessary for pitch perception (Krumbholz *et al.*, 2000; Pressnitzer *et al.*, 2001), suggesting that spectrally unresolved pitch would be inaudible in conditions of binaural unmasking. This assumption was tested in the second experiment.

III. EXPERIMENT 2

The purpose of experiment 2 was to measure the pitch salience of binaurally unmasked harmonic tones as a function of the degree of spectral resolution of harmonics by the cochlear filters. For that, the signal level required to perform a musical interval recognition (MIR) task was measured for both homophasic and antiphasic harmonic tones; the fundamental frequency (F0) of the harmonic tones was varied between 128 and 48 Hz to manipulate spectral resolution. As F0 is decreased, the frequency separation between the harmonic components decreases, and so, more harmonics fall within the auditory filter at a given frequency. As spectral resolution decreases, the processing of pitch relies increasingly on temporal envelope information. Experiment 1 suggests that pitch-related temporal envelope information is inaudible in conditions of binaural unmasking. This means that the MIR task should become increasingly difficult towards lower F0s. The MIR thresholds were compared to the detection thresholds of the signals, because the binaural advantage in signal detection should be independent of F0. A MIR rather than a pitch difference detection task was used to avoid participants using perceptual cues other than pitch that may have been related to F0 in a systematic way. For instance, listeners can discriminate the repetition rate of periodic stimuli even when the rate is too low to elicit pitch (e.g., Krumbholz *et al.*, 2000); in that case, discrimination is based on the perception of roughness, flutter, or pulsation. Pilot testing showed that for harmonic tones presented in quiet, the MIR task used in the current study yields a similar estimate of the lower limit of pitch as a melody recognition task (Pressnitzer *et al.*, 2001).

A. Methods

1. Stimuli

The stimuli were generated and presented in the same way as in experiment 1 (see Sec. II A 1). The noise masker had the same spectral characteristics (see dashed line in Fig. 1), presentation level, and duration as in experiment 1. The harmonic-tone signals were presented in sine phase, bandpass filtered in the same way as the noise signal in experiment 1, and presented with the same duration. As in experiment 1, the signals were presented either antiphasically or homophasically in the diotic noise masker.

2. Procedure

The detection thresholds of the harmonic-tone signals were measured with the same procedure as used for measuring the detection threshold of the bandpass noise signal in experiment 1 (see Sec. II A 2). The MIR threshold was measured with a two-interval, three-alternative forced-choice procedure. As in the detection threshold measurements, a 3-down 1-up rule was used. Each trial consisted of two 800 ms observation intervals, separated by a 500 ms silent gap. Both intervals contained both the masker and a harmonic-tone signal. The F0 of the harmonic tones were separated by one of three musical intervals, a full tone, a musical third, or a musical fifth (200, 400, or 700 cents), chosen randomly. The first note was always the lower note. The task was to indicate which musical interval was presented by pressing one of three response buttons. In order to avoid participants basing their judgments on the absolute pitch or some other perceptual attribute related to the absolute value of the F0 of the upper note, the F0 of the lower note was randomized over a range of ± 2 semitones (400 cents). Five different nominal values of the F0 of the lower note were tested (48, 56, 64, 96, and 128 Hz). The adaptive parameter was the signal level. The signal level was incremented and decremented by 5 and 3 dB up to its first reversal and from the first to the second reversal, respectively; after the second reversal, the size of the level increments and decrements was reduced to 2 dB for the rest of the ten reversals that made up each threshold run. Each threshold estimate was the average of the signal levels at the last eight reversals. At least three such threshold estimates were averaged to obtain the threshold value for each condition. The order in which conditions were tested was counterbalanced across the three threshold runs.

3. Participants

Three participants (2 females, 1 male) were tested; one was author KK, and the other two were students who were paid at an hourly rate. The participants were between 20 and 32 years of age and had no reported history of hearing impairment or neurological disease. All three participants had regularly played a musical instrument or sung in a choir from an early age and needed no training in MIR.

B. Results and interim discussion

In both the homophasic and antiphasic masking conditions, the detection threshold for the harmonic-tone signals increased slightly with decreasing F0 [see open symbols in Fig. 4(A)]. A two-way repeated-measures ANOVA of the individual detection thresholds with factors masking condition (homophasic and antiphasic) and F0 (48, 56, 64, and 96 Hz) indicated that this effect was significant [main effect of F0: $F(4,8)=7.899$, $p=0.007$], as was the difference between the homophasic and antiphasic masking conditions [main effect of masking condition: $F(1,2)=77.217$, $p=0.013$]. However, the interaction between masking condition and F0 was not significant [$F(4,8)=0.504$, $p=0.735$], indicating that the BMLD for the harmonic tones was largely independent of F0 [difference between open circles and squares in Fig. 4(A); open symbols in Fig. 4(B)]; on average, the BMLD for detection amounted to 11.9 dB.

In contrast, a two-way repeated-measures ANOVA of the MIR thresholds, again with factors masking condition

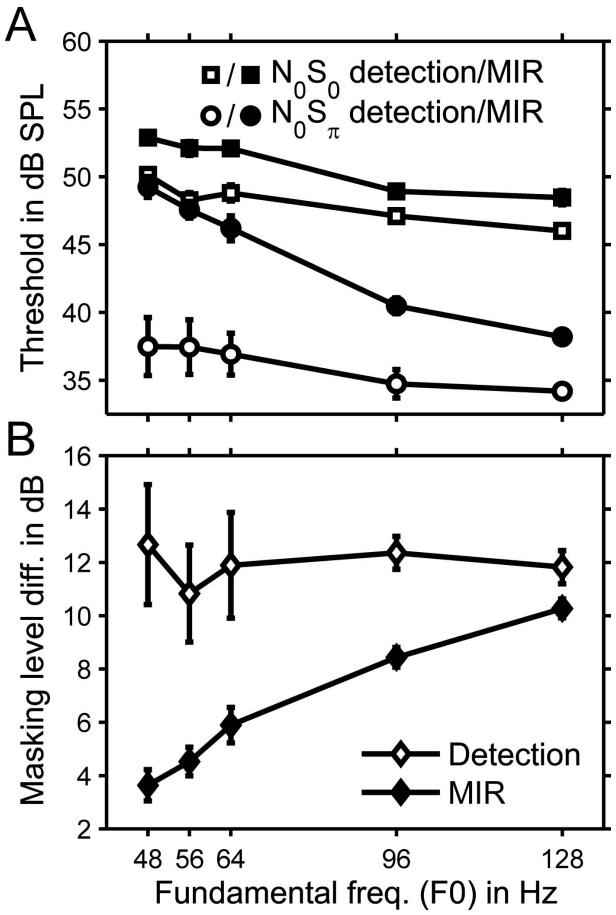


FIG. 4. (A) Average detection (open symbols) and MIR thresholds (filled symbols) of three participants in dB SPL, plotted as a function of the F0 in Hz. In the case of the MIR thresholds, F0 refers to the nominal F0 of the lower note in the intervals. The masker had an overall level of about 60 dB SPL. The signals were homophasic (N_0S_0, squares) and antiphasic (N_0S_π, circles) harmonic tones. (B) BMLD in dB for the detection (open symbols) and MIR tasks (filled symbols). Error bars show the standard error of the mean.

and F0, exhibited a highly significant interaction between these two factors [$F(4,8)=54.489$; $p<0.001$]. In terms of the absolute level, the MIR threshold for the homophasic condition increased slightly with decreasing F0 [filled squares in Fig. 4(A)]. However, when expressed in terms of the SL (difference between filled and open squares), the MIR threshold for the homophasic tones was essentially independent of F0, amounting to an average of only 2.2 dB SL [as revealed by a one-way repeated-measures ANOVA of the homophasic MIR thresholds expressed in dB SL with factor F0: $F(4,8)=1.457$, $p=0.301$]. In contrast, the MIR threshold for the antiphasic condition increased strongly with decreasing F0, irrespective of whether it was expressed in terms of absolute level (filled circles) or SL (difference between filled and open circles); at 128 Hz, the MIR threshold amounted to an average of 4 dB SL, compared to 11.8 dB SL at 48 Hz. The statistical significance of this effect was confirmed with a one-way repeated-measures ANOVA of the antiphasic MIR thresholds expressed in dB SL with factor F0 [$F(4,8)=7.496$ m; $p=0.008$]. The increase in MIR threshold with decreasing F0 in the antiphasic condition meant that the BMLD for MIR decreased from about 10.3 dB at 128 Hz to only 3.6 dB at 48 Hz [difference between filled circles and squares in Fig. 4(A); filled diamonds in Fig. 4(B)].

The increase in MIR threshold with decreasing F0 in the antiphasic masking condition had been expected based on the finding of experiment 1 that temporal envelope information is not preserved in conditions of binaural unmasking (Sec. II B) and the fact that pitch becomes increasingly reliant on temporal envelope information at low F0s due to a decrease in spectral resolution. The MIR threshold function for the antiphasic masking condition in Fig. 4(A) (filled circles) suggests that temporal envelope-based pitch cues started to become relevant at F0s below about 100 Hz; at this point, the homophasic signals started to elicit a sensation of roughness. According to the results by Shackleton and Carlyon (1994) on the effect of component phase on pitch perception, harmonic sounds are resolved when fewer than two harmonics fall within the 10 dB bandwidth of the cochlear filters and unresolved when there are more than 3.25 harmonics per filter bandwidth. Based on this definition, the limit of spectral resolution would be expected to range between F0=73 and 119 Hz at the upper edge of the signal passband (at 1 kHz) in the current experiment. Thus, for F0s above about 100 Hz, the harmonic tones would be assumed to have been resolved over most of the signal passband. As the F0 of the signal was decreased, however, an increasing portion of its passband would have become unresolved, thus reducing the resolved portion. Under the assumption that only the resolved portion contributed to the pitch percept in the antiphasic masking condition, this explains the progressive increase in MIR threshold for F0s below about 100 Hz. Thus, the MIR threshold for the antiphasic condition would be assumed to reflect the signal level at which the resolved portion of the signal passband becomes sufficiently audible for performing the task.

At the highest F0 of 128 Hz, harmonics would be assumed to have been resolved over the entire signal passband, and the MIR threshold for the antiphasic signal was almost as small as that for the homophasic signal [4 versus 2.5 dB SL; $t(2)=2.106$, $p=0.17$]. This is consistent with the finding by Akeroyd *et al.* (2001) that dichotic pitches are "true" pitches in the sense that they are capable of conveying musical intervals and melody.

IV. GENERAL DISCUSSION

The current study confirms previous findings showing that in conditions of binaural masking release, binaural sluggishness not only smears temporal changes in the interaural properties of the masker but also affects the perception of the temporal properties of the signal (Hall and Grose, 1992; Culling and Colburn, 2000). By comparing the TMTF of a binaurally unmasked (N_0S_π) bandpass noise signal with that of the same signal presented in diotic masking conditions (N_0S_0), experiment 1 of the current study showed that the internal representation of binaurally unmasked sounds conveys little or no information about envelope fluctuations with rates above the lower limit of melodic pitch, which is a little above 30 Hz (Krumbholz *et al.*, 2000; Pressnitzer *et al.*, 2001). This suggests that the pitch of spectrally unresolved

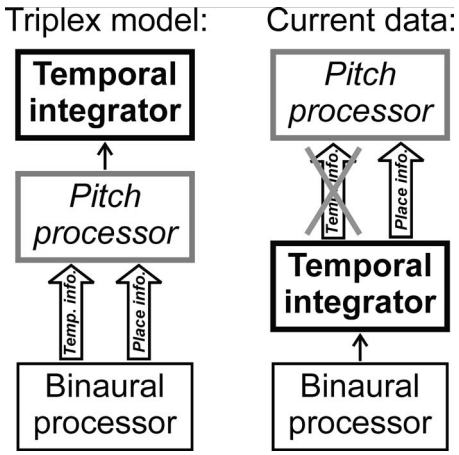


FIG. 5. Schematic representation of possible processing hierarchies. The left column shows the triplex model, which assumes that the temporal integration stage follows the pitch processor, so the pitch-related temporal information is fully preserved at the input to the pitch processor. In the triplex model, pitch processing is assumed to be based on temporal cues. In contrast, the current data suggest that the pitch processor follows the temporal integration stage of the binaural processor (right column). Under certain additional assumptions (see text), this could mean that binaural channels mediate only place and no temporal information on pitch.

harmonic tones should not be audible in conditions of binaural unmasking because the periodic envelope fluctuations produced by harmonic beating within cochlear filters would be eliminated by binaural sluggishness. This prediction was confirmed in experiment 2, which showed that MIR with binaurally unmasked harmonic tones decreases dramatically as the fundamental frequency (F0) of the harmonic tones decreases, and the unresolved portion of the signal passband increases. The small remaining BMLDs for MIR that were observed even at the lowest F0s tested (e.g., 3.6 dB at 48 Hz) were probably based on the residual spectral pitch information that was still available in the lower part of the signal passband in these conditions. According to Shackleton and Carlyon's (1994) definition of spectral resolvability, about 35% of the passband would still have been resolvable even for the lowest F0 tested (48 Hz).

Binaural sluggishness has been modeled as a temporal integrator (Holube *et al.*, 1998; Culling and Summerfield, 1998; Akeroyd and Summerfield, 1999a). In the triplex model proposed by Licklider (1959) (see also Patterson and Akeroyd, 1995), the integrator is thought to be preceded by the pitch processor, which, in turn, is preceded by the binaural processor (see left column in Fig. 5). The triplex model is a combined model of binaural and monaural temporal processing, which was inspired by a dichotic pitch phenomenon discovered by Cramer and Huggins (1958). Like Akeroyd and Summerfield's (1999b) temporal account of dichotic pitch perception, the triplex model assumes that pitch-related temporal information is fully preserved by the binaural processing stage and that pitch perception is based on temporal cues even when the stimulus is perceived through binaural channels (i.e., in conditions of binaural unmasking). The current data suggest that this assumption is at least partly wrong. They indicate that the input to the pitch processor from binaural channels carries little or no temporal envelope-based cues to pitch, thus suggesting that the binaural temporal integrator should be assumed to precede, rather than follow, the pitch processor (see right column in Fig. 5).

The current study used spectrally unresolved harmonic tones and amplitude-modulated noise because these stimuli convey no spectral (or place) pitch information (Burns and Viemeister, 1976, 1981). The pitches of these stimuli are based on periodic fluctuations in the amplitude or envelope of the cochlear filter responses. In contrast, in low-frequency pure tones and spectrally resolved harmonic tones, temporal pitch information is mediated by the fine structure, rather than the envelope, of the filter responses. Most temporal pitch models assume that fine-structure and envelope-based pitch cues are processed by a common mechanism (Licklider, 1951; Meddis and O'Mard, 1997; Bernstein and Oxenham, 2005). With this and the additional assumption that binaural sluggishness affects pitch-related fine-structure cues in a similar way as it has here been shown to affect envelope cues, the current results suggest that all forms of pitch perception in conditions of binaural unmasking, including dichotic pitch phenomena, such as Huggins pitch (Cramer and Huggins, 1958) or multiple-phase-shift pitch (Bilsen, 1976; for review, see Culling *et al.*, 1998a, 1998b; Culling, 2000), would have to be based on purely spectral (place) cues. However, these two assumptions may not necessarily be warranted. In particular, fine-structure and envelope information may be analyzed by different mechanisms, which may be located at different levels in the processing hierarchy. For instance, the processing of fine-structure information may be based on the differences in the phase of the basilar-membrane response between adjacent places along the membrane (Loeb *et al.*, 1983; Shamma, 1985; Carlyon and Shamma, 2003) and would be assumed to occur below or at the level of the inferior colliculus (IC), because the accuracy of phase locking is known to decrease dramatically beyond the brainstem (Rouiller *et al.*, 1979). Temporal envelope processing and binaural sluggishness, on the other hand, may occur at much higher levels. Physiological data suggest that pitch-related envelope information may be preserved up to levels as high as the auditory cortex (Lu *et al.*, 2001) and that binaural sluggishness occurs above the level of the IC (Joris *et al.*, 2006). Moreover, it may be the case that binaural sluggishness does not affect temporal fine-structure information in the same way as it affects envelope information. Some, albeit indirect, support for this hypothesis comes from recent psychophysical and physiological results suggesting that binaural sluggishness leaves some aspects of the temporal information in binaural stimuli unaffected; Siveke *et al.* (2008) showed that binaural sluggishness affects the perception of periodic changes in the interaural correlation but not in the interaural time difference of a binaural stimulus.

Thus, whether or not binaural sluggishness affects pitch cues based on temporal fine structure in the same way as it affects envelope-based pitch cues remains an open and challenging question. If fine-structure cues do turn out to be affected by binaural sluggishness, binaural unmasking would constitute a unique case for investigating pitch perception in the absence of temporal cues.

ACKNOWLEDGMENTS

This work was funded by the Medical Research Council (UK).

Akeroyd, M. A., Moore, B. C. J., and Moore, G. A. (**2001**). "Melody recognition using three types of dichotic-pitch stimulus," J. Acoust. Soc. Am. **110**, 1498–1504.

Akeroyd, M. A., and Summerfield, A. Q. (**1999a**). "A binaural analog of gap detection," J. Acoust. Soc. Am. **105**, 2807–2820.

Akeroyd, M. A., and Summerfield, A. Q. (**1999b**). "A fully-temporal account of the perception of dichotic pitches," Br. J. Audiol. **33**, 106–107.

Bernstein, L. R. (**2001**). "Auditory processing of interaural timing information: New insights," J. Neurosci. Res. **66**, 1035–1046.

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.

Bilsen, F. A. (**1976**). "Pronounced binaural pitch phenomenon," J. Acoust. Soc. Am. **59**, 467–468.

Blauert, J. (**1968**). "Ein Beitrag zur Trägheit des Richtungshörens in der Horizontalebene (On the lag of sound localization in the horizontal plane)," Acustica **20**, 200–206.

Blauert, J. (**1970**). "Zur Trägheit des Richtungshörens bei Laufzeit—und Intensitätsstereophonie (About the inertia of auditory localization in running time and intensity stereophony)," Acustica **23**, 287–293.

Burns, E. M., and Viemeister, N. F. (**1976**). "Nonspectral pitch," J. Acoust. Soc. Am. **60**, 863–869.

Burns, E. M., and Viemeister, N. F. (**1981**). "Played-again SAM: Further observations on the pitch of amplitude-modulated noise," J. Acoust. Soc. Am. **70**, 1655–1660.

Carlyon, R. P., and Shamma, S. (**2003**). "An account of monaural phase sensitivity," J. Acoust. Soc. Am. **114**, 333–348.

Cramer, E. M., and Huggins, W. H. (**1958**). "Creation of pitch through binaural interaction," J. Acoust. Soc. Am. **30**, 413–417.

Culling, J. F. (**2000**). "Dichotic pitches as illusions of binaural unmasking. III. The existence region of the Fourcin pitch," J. Acoust. Soc. Am. **107**, 2201–2208.

Culling, J. F., and Colburn, H. S. (**2000**). "Binaural sluggishness in the perception of tone sequences and speech in noise," J. Acoust. Soc. Am. **107**, 517–27.

Culling, J. F., Marshall, D. H., and Summerfield, A. Q. (**1998a**). "Dichotic pitches as illusions of binaural unmasking. II. The Fourcin pitch and the dichotic repetition pitch," J. Acoust. Soc. Am. **103**, 3527–3539.

Culling, J. F., and Summerfield, Q. (**1998**). "Measurements of the binaural temporal window using a detection task," J. Acoust. Soc. Am. **103**, 3540–3553.

Culling, J. F., Summerfield, A. Q., and Marshall, D. H. (**1998b**). "Dichotic pitches as illusions of binaural unmasking. I. Huggins' pitch and the 'binaural edge pitch'," J. Acoust. Soc. Am. **103**, 3509–3526.

Dau, T., Verhey, J., and Kohlrausch, A. (**1999**). "Intrinsic envelope fluctuations and modulation-detection thresholds for narrow-band noise carriers," J. Acoust. Soc. Am. **106**, 2752–2760.

Durlach, N. I., and Colburn, H. S. (**1978**). "Binaural phenomena," in *The Handbook of Perception*, edited by E. C. Carterette, and M. P Friedman (Academic, New York).

Durlach, N. I., Gabriel, K. J., Colburn, H. S., and Trahiotis, C. (**1986**). "Interaural correlation discrimination: II. Relation to binaural unmasking," J. Acoust. Soc. Am. **79**, 1548–1557.

Grantham, D. W. (**1982**). "Detectability of time-varying interaural correlation in narrow-band noise stimuli," J. Acoust. Soc. Am. **72**, 1178–1184.

Grantham, D. W. (**1984**). "Discrimination of dynamic interaural intensity differences," J. Acoust. Soc. Am. **76**, 71–76.

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.

Hall, J. W., and Grose, J. H. (**1992**). "Masking release for gap detection," Philos. Trans. R. Soc. London, Ser. B **336**, 331–337.

Holube, I., Kinkel, M., and Kollmeier, B. (**1998**). "Binaural and monaural auditory filter bandwidths and time constants in probe tone detection experiments," J. Acoust. Soc. Am. **104**, 2412–2425.

Joris, P. X., van de Sande, B., Recio-Spinoso, A., and van der Heijden, M. (**2006**). "Auditory midbrain and nerve responses to sinusoidal variations in interaural correlation," J. Neurosci. **26**, 279–289.

Kollmeier, B., and Gilkey, R. H. (**1990**). "Binaural forward and backward masking: Evidence for sluggishness in binaural detection," J. Acoust. Soc. Am. **87**, 1709–1719.

Krumbholz, K., Patterson, R. D., and Pressnitzer, D. (**2000**). "The lower limit of pitch as determined by rate discrimination," J. Acoust. Soc. Am. **108**, 1170–1180.

Levitt, H. (**1971**). "Transformed up-down Methods in Psychoacoustics," J. Acoust. Soc. Am. **49**, 466–477.

Licklider, J. C. R. (**1951**). "A duplex theory of pitch perception," Experientia **7**, 128–134.

Licklider, J. C. R. (**1959**). "Three auditory theories," in *Psychology: A Study of a Science*, edited by S. Koch (McGraw-Hill, New York).

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.

Lu, T., Liang., L., and Wang, X. (**2001**). "Temporal and rate representations of time-varying signals in the auditory cortex of awake primates," Nat. Neurosci. **4**, 1131–1138.

Meddis, R., and O'Mard, L. (**1997**). "A unitary model of pitch perception," J. Acoust. Soc. Am. **102**, 1811–1820.

Patterson, R. D., and Akeroyd, M. A. (**1995**). "Time-interval patterns and sound quality," in Advances in *Hearing Research: Proceedings of the Tenth International Symposium on Hearing*, edited by G. A. Manley, G. M. Klump C. Köppl, H. Fastl, and H. Oeckinghaus (World Scientific, Singapore), pp. 545–556.

Plack, C. J., and Moore, B. C. (**1990**). "Temporal window shape as a function of frequency and level," J. Acoust. Soc. Am. **87**, 2178–2187.

Pressnitzer, D., Patterson, R. D., and Krumbholz, K. (**2001**). "The lower limit of melodic pitch," J. Acoust. Soc. Am. **109**, 2074–2084.

Rouiller, E., de Ribaupierre, Y., and de Ribaupierre, F. (**1979**). "Phase-locked responses to low frequency tones in the medial geniculate body," Hear. Res. **1**, 213–226.

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. I: The representation of speech sounds in the responses of the auditory nerve," J. Acoust. Soc. Am. **78**, 1612–1621.

Siveke, I., Ewert, S. D., Grothe, B., and Wiegrebe, L. (**2008**). "Psychophysical and physiological evidence for fast binaural processing," J. Neurosci. **28**, 2043–2052.

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.

Strickland, E. A., and Viemeister, N. F. (**1997**). "The effects of frequency region and bandwidth on the temporal modulation transfer function," J. Acoust. Soc. Am. **102**, 1799–1810.

Viemeister, N. F. (**1979**). "Temporal modulation transfer functions based upon modulation thresholds," J. Acoust. Soc. Am. **66**, 1364–1380.

Wiegrebe, L., and Krumbholz, K. (**1999**). "Temporal resolution and temporal masking properties of transient stimuli: Data and an auditory model," J. Acoust. Soc. Am. **105**, 2746–2756.

Yost, W. A. (**1985**). "Prior stimulation and the masking-level difference," J. Acoust. Soc. Am. **78**, 901–907.

Zwicker, E., and Fastl, H. (**1990**). *Psychoacoustics: Facts and Models* (Springer-Verlag, Berlin).

Estimation of the center frequency of the highest modulation filter

Brian C. J. Moore,[a] Christian Füllgrabe, and Aleksander Sek[b]
Department of Experimental Psychology, University of Cambridge, Downing Street, Cambridge CB2 3EB, England

(Received 21 May 2008; revised 5 September 2008; accepted 4 December 2008)

For high-frequency sinusoidal carriers, the threshold for detecting sinusoidal amplitude modulation increases when the signal modulation frequency increases above about 120 Hz. Using the concept of a modulation filter bank, this effect might be explained by (1) a decreasing sensitivity or greater internal noise for modulation filters with center frequencies above 120 Hz; and (2) a limited span of center frequencies of the modulation filters, the top filter being tuned to about 120 Hz. The second possibility was tested by measuring modulation masking in forward masking using an 8 kHz sinusoidal carrier. The signal modulation frequency was 80, 120, or 180 Hz and the masker modulation frequencies covered a range above and below each signal frequency. Four highly trained listeners were tested. For the 80-Hz signal, the signal threshold was usually maximal when the masker frequency equaled the signal frequency. For the 180-Hz signal, the signal threshold was maximal when the masker frequency was below the signal frequency. For the 120-Hz signal, two listeners showed the former pattern, and two showed the latter pattern. The results support the idea that the highest modulation filter has a center frequency in the range 100–120 Hz.
 [DOI: 10.1121/1.3056562]

PACS number(s): 43.66.Mk, 43.66.Dc, 43.66.Ba [MW] Pages: 1075–1081

I. INTRODUCTION

The perception of amplitude modulation (AM) has been proposed to depend on an array of modulation filters (Kay, 1982; Dau *et al.*, 1997a, 1997b). It is assumed that the envelopes of the outputs of the peripheral auditory filters are fed to a second array of overlapping bandpass filters tuned to different envelope modulation frequencies. This set of filters is usually called a modulation filter bank (MFB), and the modulation filters are assumed to be implemented at a higher level in the auditory system than the auditory nerve (Møller, 1972; Langner and Schreiner, 1988; Schreiner and Langner, 1988; Joris and Yin, 1992). The concept of the MFB implies that the auditory system performs a spectral analysis of the envelope at the output of each auditory filter, although the putative modulation filters appear to be rather broad, with Q values (center frequency divided by bandwidth) in the range 0.35–2 (Ewert and Dau, 2000; Ewert *et al.*, 2002; Sek and Moore, 2002, 2003; Verhey *et al.*, 2003; Wojtczak and Viemeister, 2005).

It is known that the sensitivity to AM, as measured by the temporal modulation transfer function (TMTF), decreases with increasing modulation frequency. For high-frequency sinusoidal carriers, the TMTF rolls off above about 120 Hz (Kohlrausch *et al.*, 2000; Moore and Glasberg, 2001; Füllgrabe and Lorenzi, 2003), while for broadband noise carriers the roll off occurs at a somewhat lower frequency (Viemeister, 1979). The discrepancy has been explained in terms of the masking effect of the inherent random fluctuations in a noise carrier (Dau *et al.*, 1997b). The decreasing sensitivity to AM with increasing modulation frequency for high-frequency sinusoidal carriers cannot be explained in terms of fluctuations in the carrier; it must reflect an intrinsic limitation of processing in the auditory system. Within the framework of the MFB concept, Dau (1996) described two ways in which this limitation might arise: (1) modulation filters tuned to high modulation frequencies (above about 120 Hz) might be less sensitive than modulation filters tuned to low modulation frequencies, either because of reduced gain or because of greater internal noise in the filters tuned to higher modulation frequencies; and (2) the center frequencies of the modulation filters may span a limited range, with a highest center frequency in the range 100–150 Hz. In the latter case, the roll off in the TMTF for high modulation frequencies would reflect the characteristic of the high-frequency side of the highest modulation filter. Dau (1996) showed that models based on either of these assumptions could account for the general form of the TMTF for high-frequency sinusoidal carriers.

Physiological data do not give a clear indication of the center frequency of the highest modulation filter, partly because of uncertainty as to the location of the putative MFB within the auditory system, and partly because of differences across species, which make extrapolation to humans difficult. Single neurons in the auditory system sometimes show low-pass response patterns as a function of the input modulation frequency, while others shows bandpass patterns (Joris *et al.*, 2004). The bandpass pattern is more common in the inferior colliculus than in the cochlear nucleus (Joris *et al.*, 2004). In a review, Palmer (1995) pointed out that, for those neurons showing a bandpass response pattern, the highest value of

[a]Author to whom correspondence should be addressed. Electronic mail: bcjm@cam.ac.uk
[b]Also at Institute of Acoustics, Adam Mickiewicz University, 85 Umultowska, 61-614 Poznan, Poland.

the best modulation frequency (BMF) in the cochlear nucleus of the cat ranged from 240 Hz for onset units to 700 Hz for primarylike units. In the inferior colliculus, the highest BMFs were mostly below 200 Hz in the rat and guinea pig, while in the cat BMFs were mostly below 100 Hz, although BMFs of 300–1000 Hz also occurred.

The idea that the highest modulation filter in humans is centered at a modulation frequency in the range 100–150 Hz is analogous to the idea that the auditory filters have a limited range of center frequencies. Sensitivity to very low audio frequencies (below about 50 Hz) might depend on the characteristics of the low-frequency side of the lowest auditory filter (Moore *et al.*, 1997), while sensitivity to very high audio frequencies (above about 15 000 Hz) might depend on the characteristics of the high-frequency side of the highest auditory filter (Ashihara *et al.*, 2006). For people with a high-frequency dead region (a region of the cochlea where there are no functioning inner hair cells and/or neurons), the highest effective auditory filter might be centered at a frequency well below 15 000 Hz (Moore, 2004). This idea has been tested by the measurement of psychophysical tuning curves (PTCs). The tip frequency of a PTC is the masker frequency at which the level of the masker required for threshold is lowest, i.e., the most effective masker frequency. Normally, the tip frequency lies very close to the signal frequency. However, if the signal frequency falls in a dead region, or falls above the center frequency of the highest auditory filter, the tip frequency of the PTC may be shifted (Thornton and Abbas, 1980; Buus *et al.*, 1986; Moore *et al.*, 2000; Moore and Alcántara, 2001; Yasin and Plack, 2005). This suggests that the signal is being detected using an auditory filter which is not centered at the signal frequency, and is sometimes referred to as "off-frequency listening" (Johnson-Davies and Patterson, 1979; O'Loughlin and Moore, 1981). It is usually assumed that the tip frequency marks the boundary of the dead region or the center frequency of the highest auditory filter. The data of Buus *et al.* (1986) and Yasin and Plack (2005) suggest that the highest auditory filter is centered somewhat above 15 000 Hz, although the exact value probably varies across individuals and may depend on age.

Similar logic can be applied to test the idea that the modulation filters have center frequencies that span a limited range, and, if so, to estimate the center frequency of the "top" modulation filter. A relevant experiment was performed by Ewert and Dau (2000). They measured masked-threshold patterns (MTPs) for signal modulation frequencies of 4, 16, 64, and 256 Hz using as a (simultaneous) modulation masker a 0.5-octave wide band of noise centered at various frequencies above and below the signal frequency. The carrier was a bandpass filtered Gaussian noise or low-noise noise (Pumplin, 1985). For the three lowest signal frequencies, the MTPs peaked at the signal frequency. However, for the 256 Hz signal frequency, the MTPs showed a peak at a frequency well below 256 Hz. Following Kohlrausch *et al.* (2000), Ewert and Dau (2000) modeled the results by assuming that the modulation filters covered a broad range of center frequencies, but they included an additional first-order lowpass filter following the modulation filters, with a cutoff frequency of 150 Hz. This filter effectively implements the assumption that the modulation filters become less sensitive (have less gain) for modulation frequencies above 150 Hz. The model accounted well for the general form of the data, but it did not accurately predict the downward-shifted tips of the MTPs for the 256-Hz signal frequency. The discrepancy suggests that it may be more appropriate to model the results by assuming that the modulation filters have only a limited range of center frequencies, with the highest filter centered between 100 and 150 Hz.

We attempted to address this issue more definitively by exploiting the phenomenon of forward masking in the AM domain (Wojtczak and Viemeister, 2005). Forward masking was used in preference to simultaneous masking for three reasons: (1) to avoid temporal interactions between the envelopes of the masker and the signal (Strickland and Viemeister, 1996; Lorenzi *et al.*, 1999), which might provide cues for the detection of the signal that depend on factors other than frequency selectivity in the modulation domain; (2) to avoid distortion products in the modulation domain (Moore *et al.*, 1999; Verhey *et al.*, 2003; Sek and Moore, 2004; Füllgrabe *et al.*, 2005), which again might provide cues for the detection of the signal, which are not closely related to frequency selectivity in the modulation domain; and (3) to allow a greater range of modulation depths of the signal. The signal modulation depth, m, can be as great as 1 in forward masking, but has to be restricted to a smaller value in simultaneous masking to avoid overmodulation.

The carrier was a high-frequency (8000 Hz) sinusoid, chosen so as to satisfy two requirements: (1) The spectral sidebands produced by the AM would not be resolved even for the highest modulation frequency used, which was 360 Hz (Zwicker, 1956; Sek and Moore, 1994; Kohlrausch *et al.*, 2000; Moore and Glasberg, 2001); (2) the carrier had no inherent amplitude fluctuations, which might mask the signal modulation. For high modulation frequencies, the range of modulation depths between the threshold for detecting the modulation and "full" (100%) modulation is rather small. This can make it difficult to measure PTCs in the modulation domain, especially when forward masking is used, since it may not be possible to make the masker modulation depth sufficient to mask the signal. To overcome this problem, we fixed the masker modulation depth and measured the threshold for detecting the signal modulation as a function of the *masker* modulation frequency. This was done for three signal modulation frequencies: 80, 120, and 180 Hz. These were chosen to span the range of center frequencies that have been postulated for the top modulation filter. We reasoned that if the signal modulation frequency fell below the center frequency of the top modulation filter, then the signal threshold should be maximal for a masker frequency that was *equal* to the signal frequency, as found by Wojtczak and Viemeister (2005) for modulation frequencies up to 80 Hz. However, if the signal modulation frequency fell above the center frequency of the top modulation filter, then the signal threshold should be maximal for a masker frequency that was *below* the signal frequency.

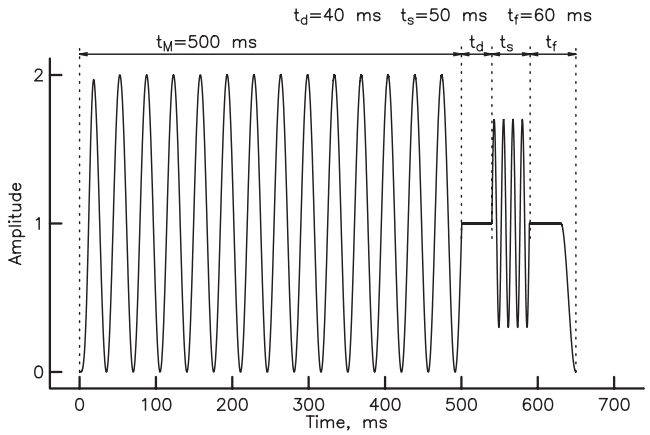


FIG. 1. Illustration of the envelope of the stimuli for an interval in which both masker and signal modulations are present.

II. METHOD

A. Listeners

Four listeners were tested L1, L2, and L3 were male and L4 was female. All listeners had audiometric thresholds $\leq$15 dB hearing loss (HL) at octave frequencies between 0.25 and 8 kHz. Their ages ranged from 21 to 37 years (mean age=26 years; standard deviation, SD=7 years). Prior to data collection, all listeners received 12 h of practice during which each condition was presented between two and four times. All listeners were university students and were paid for their services.

B. Stimuli

Stimuli were digitally generated using a PC-controlled Tucker-Davis Technologies (TDT) system with a 16-bit digital-to-analog (DA) converter (TDT DD1, 50-kHz sampling rate). Following DA conversion, the signals were attenuated (TDT PA4), passed through a headphone buffer (TDT HB6), and delivered to a double-walled sound attenuating booth. Stimuli were passed through a final manual attenuator (Hatfield 2125) and presented via the left earpiece of a Sennheiser HD580 headphone. The sensation level (SL) of the unmodulated carrier was set to 60 dB, based on individually measured absolute thresholds (see below for details).

The stimuli were chosen to be similar to those of Wojtczak and Viemeister (2005), experiment 2, except that they used a broadband noise carrier while we used a sinusoidal carrier. In each trial, the 8000-Hz carrier was presented in two bursts with 20-ms raised-cosine ramps at the start and end. The bursts were separated by a 415-ms silent interval. Figure 1 illustrates the envelope of a single stimulus containing a masker and a signal. The masker modulation was applied as soon as the carrier was turned on, with a modulation depth, m, equal to 1 (100% modulation). The masker modulation started at a phase of 270°, i.e., at its most negative value, so that the carrier amplitude at the onset was zero. The masker modulation continued for approximately 500 ms and was turned off at a positive-going zero crossing (0° phase). The exact duration of the masker modulation was determined as follows. The masker period in milliseconds, P_m, was divided into 500. The result was rounded to the nearest integer, denoted N. The duration of the modulation masker was then set to $(N \times P_m) + 0.25P_m$.

Following the end of the masker modulation, the carrier was unmodulated for 40 ms (indicated by t_d in the figure), and then the signal modulation (with adjustable m) was applied for 50 ms (indicated by t_s in the figure), starting and ending at positive-going zero crossings. The carrier continued for 60 ms after the end of the signal modulation (indicated by t_f in the figure), with an unmodulated portion of 40 ms followed by a 20-ms ramp. In a nonsignal interval, the signal modulation depth was set to zero, but the modulation pattern was otherwise identical. To measure the "absolute" threshold for detecting the signal modulation, the masker modulation depth was set to zero, but the stimuli were otherwise the same.

The transition from masker modulation to no modulation, signal modulation, and no modulation was chosen to occur at zero crossings in the modulation waveforms, to reduce spectral splatter in the modulation domain (Sek and Moore, 2002). While some splatter would have occurred, the side lobes produced by the splatter were at least 20 dB lower in level than the main lobe at the primary modulator frequencies. Given the broad tuning of the modulation filters, it seems unlikely that the splatter had a material influence on the results.

Three signal modulation frequencies were used: 80, 120, and 180 Hz. These frequencies were such that the 50-ms signal duration contained an integer number of signal cycles. For the signal frequency of 80 Hz, the masker frequencies were 20, 28, 40, 57, 80, 113, 160, 226, and 320 Hz. For the signal frequency of 120 Hz, the masker frequencies were 30, 42, 60, 85, 120, 170, 240, and 339 Hz. Finally, for the signal frequency of 180 Hz, the masker frequencies were 45, 64, 90, 127, 180, 255, and 360 Hz.

C. Procedure

A two-interval, two-alternative forced-choice procedure was used. In each trial, two bursts of the carrier were presented. Observation intervals were marked by lights on the response box. Both bursts contained the masker modulation, but only one burst, selected randomly, contained the signal modulation. The listener was required to indicate the interval containing the signal modulation by pressing one of two buttons (labeled "1" and "2") on a response box. Visual feedback was provided after each trial. The next trial was initiated 500 ms after the listener pressed the button. A three-down, one-up rule was used to estimate the value of the signal modulation depth necessary for 79.4%-correct detection. The signal modulation depth at the start of a run was chosen to be about 8 dB (in terms of $20 \log_{10} m$) greater than the estimated threshold (as determined during training runs), so to make the signal easy to detect. When this was not possible, the starting modulation depth was set to 0 dB ($m=1$). The initial step size was 4 dB. The step size was reduced to 2 dB after two reversals and to 1 dB after two more reversals. A run was terminated after 12 reversals, and the mean value of $20 \log_{10} m$ at the last eight reversal points was

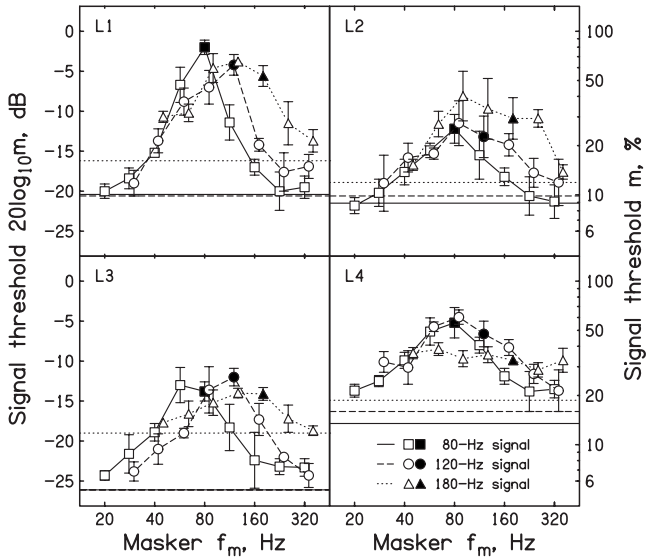


FIG. 2. Mean results for the four individual listeners. The signal modulation depth required for threshold, expressed as $20 \log_{10} m$, is plotted as a function of the masker modulation frequency, f_m. The symbols indicate the signal frequency (squares: 80 Hz, circles: 120 Hz, and triangles: 180 Hz). The filled symbols indicate conditions where the masker frequency was equal to the signal frequency. The horizontal lines indicate signal thresholds in the absence of a modulation masker, i.e., absolute thresholds for modulation detection. The error bars show standard deviations across repeated runs.

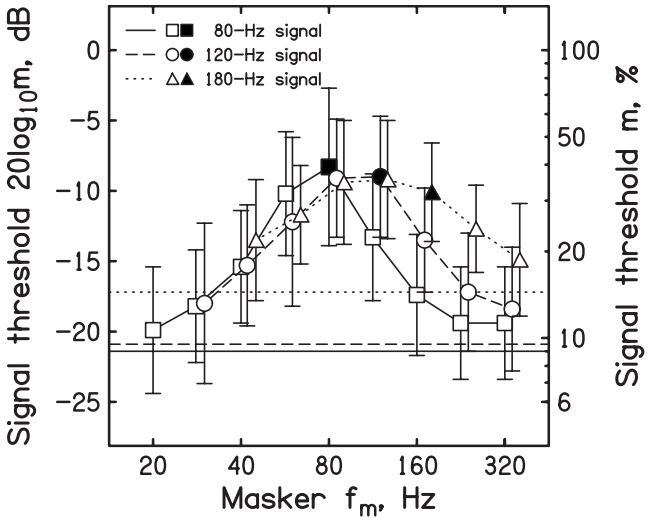


FIG. 3. As Fig. 2, but showing averages of the mean results across listeners, with corresponding standard deviations. The relatively large error bars are mainly a consequence of differences in overall thresholds across listeners.

taken as the threshold estimate. When the SD of the values of $20 \log_{10} m$ at the last eight reversal points was ≥3 dB, the results for that run were discarded and an extra run was performed. For each listener and condition, at least four valid threshold estimates were obtained. An additional estimate was obtained when one of the estimates was more than 3 dB away from the mean threshold, or when the SD across estimates for a given condition was ≥3 dB. For each signal modulation frequency, the threshold for detecting the signal in the absence of a masker was also estimated, using the same method.

Absolute thresholds for detecting the unmodulated carrier were estimated using a similar procedure (except that the carrier was present in only one of the two observation intervals in a trial). These absolute thresholds were used to set the carrier level to 60 dB SL for each listener.

III. RESULTS

The pattern of results obtained during training was generally similar to the pattern obtained in the experiment proper, except that performance did tend to improve during training, so the masked thresholds were lower for the experiment proper than during training. Figure 2 shows the individual results for the experiment proper. Figure 3 shows the mean results across listeners. The signal modulation depth at masked threshold, expressed as $20 \log_{10} m$, is plotted as a function of the masker frequency. Each symbol denotes a different signal frequency. The filled symbols indicate conditions where the masker frequency was equal to the signal frequency. The horizontal lines indicate signal thresholds in the absence of a modulation masker, i.e., absolute thresholds for modulation detection. As expected (Kohlrausch *et al.*, 2000), the absolute signal threshold was consistently higher for the 180-Hz signal frequency than for the two lower signal frequencies. For listeners L2 and L4, the absolute threshold was slightly higher for the 120-Hz signal than for the 80-Hz signal, while for listeners L1 and L3, absolute thresholds were similar for these two signal frequencies.

Consistent with the results of Wojtczak and Viemeister (2005), who used signal frequencies up to 80 Hz, the masking patterns showed tuning in the modulation domain. For the 80-Hz signal frequency (squares), the signal threshold was highest when the masker frequency equaled the signal frequency, for all listeners except L3; for the latter, the signal threshold was slightly higher for the 57-Hz masker than for the 80-Hz masker. For the mean data, the signal threshold was clearly highest for the 80-Hz masker frequency. A similar pattern of results was found by Wojtczak and Viemeister (2005) for a signal modulation frequency of 80 Hz. The masked thresholds found here were also similar to those reported by Wojtczak and Viemeister (2005). However, our absolute thresholds for modulation detection were slightly lower (better) than obtained by Wojtczak and Viemeister (2005), probably because they used a noise carrier while we used a sinusoidal carrier.

For the 120-Hz signal frequency (circles), the signal threshold was highest when the masker frequency equaled the signal frequency for L1 and L3, but was highest when the masker frequency was below the signal frequency for L2 and L4. For the mean data, the signal threshold was almost the same for the masker frequencies of 85 and 120 Hz. For the 180-Hz signal frequency (triangles), the signal threshold was highest when the masker frequency was below the signal frequency for all four listeners. The signal threshold was highest for masker frequencies of 127, 90, 127, and 64 Hz for L1, L2, L3, and L4, respectively (although the masking pattern for L4 showed a rather broad maximum). For the mean data, the signal threshold was highest (and almost the same) for the masker frequencies of 90 and 127 Hz.

To quantify the sharpness of the masking patterns, and to quantify the shifts in the peaks of the patterns relative to

TABLE I. Results of fitting second-order Butterworth filters to the individual and mean data. The columns show, from left to right, the listener, the signal frequency, the rms error in decibels, the center frequency of the best-fitting filter, the ratio (signal frequency/center frequency), and the Q value of the best-fitting filter.

Listener	Signal frequency (Hz)	rms error (dB)	Center frequency (Hz)	Ratio	Q
L1	80	1.58	74	1.08	2.74
	120	1.88	96	1.25	1.75
	180	1.07	117	1.54	1.09
L2	80	1.21	78	1.03	0.90
	120	0.95	98	1.22	0.77
	180	1.37	121	1.49	0.92
L3	80	1.67	71	1.13	1.11
	120	0.90	104	1.15	1.41
	180	0.57	122	1.48	0.55
L4	80	1.35	74	1.08	0.84
	120	1.03	86	1.39	0.80
	180	1.44	127	1.42	0.40
Mean	80	1.48	74	1.08	1.23
	120	0.68	95	1.26	0.97
	180	0.24	117	1.54	0.62

the signal frequency, we fitted the data with second-order Butterworth bandpass filters, as was done by Wojtczak and Viemeister (2005). Such filters are symmetrical on a logarithmic frequency scale. Wojtczak and Viemeister (2005) constrained the center frequency of each fitted filter to be equal to the signal frequency, and they adjusted the ratio between the upper and lower cutoff frequencies to minimize the root-mean-square (rms) difference between the data and the fitted filter. In contrast, we chose the upper and lower cutoff frequencies for each filter as independent free parameters. The outcomes for the individual data and for the mean data across listeners are shown in Table I. The mean data and the filters that gave the best fit to the mean data are shown in Fig. 4. In this figure, the data have been normalized so that the highest measured signal threshold falls at 0 dB, and the tip of each fitted filter also falls at 0 dB.

The rms errors (column 3 of Table I) were generally somewhat smaller than found by Wojtczak and Viemeister (2005). The center frequencies of the fitted filters (column 4 of Table I) were slightly below the signal frequency for all signal frequencies, but the discrepancy was much more marked for the 180-Hz signal frequency than for the 80-Hz signal frequency. For the 180-Hz signal frequency, the center frequency of the fitted filter ranged from 117 to 127 Hz across listeners. The ratio (signal frequency)/(center frequency), shown in the fifth column of Table I, increased markedly with increasing signal frequency. A one-way within-subjects analysis of variance on the ratios showed a significant effect of signal frequency; $F(2,9)=34.3$, $p<0.001$.

When the center frequency of each fitted filter was constrained to equal the signal frequency, the rms error of the fit to the mean data increased slightly (from 1.48 to 1.69 dB)

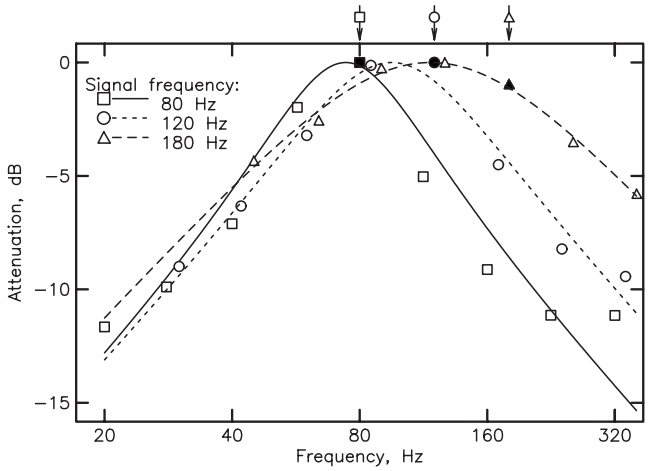


FIG. 4. Mean data across listeners, normalized so that the highest measured threshold was set to 0 dB. The filled symbols indicate conditions where the masker frequency was equal to the signal frequency. The signal frequencies are also indicated by arrows at the top of the figure. The best-fitting second-order Butterworth filter for each signal frequency is also shown.

for the signal frequency of 80 Hz, but, based on a variance-ratio test, the increase was not significant; $F(8,8)=1.3$, $p>0.05$. For the 120-Hz signal frequency, the increase (from 0.68 to 2.73 dB) was larger and was significant; $F(7,7)=16.1$, $p<0.001$. For the 180-Hz signal frequency, the increase (from 0.24 to 3.98 dB) was larger still and was significant; $F(6,6)=275.0$, $p<0.001$. Thus, for the signal frequencies of 120 and 180 Hz, the best fit to the data was clearly obtained using filters that were centered below the respective signal frequencies.

The Q values of the fitted filters (center frequency divided by bandwidth at the 3-dB down point, sixth column of Table I) ranged from 0.4 to 2.74. For the mean data, the Q values for the two lower signal frequencies were close to 1. These Q values are broadly consistent with previous estimates of the sharpness of the modulation filters (Ewert and Dau, 2000; Ewert *et al.*, 2002; Sek and Moore, 2002, 2003; Verhey *et al.*, 2003), but they are larger than the Q values reported by Wojtczak and Viemeister (2005). However, Wojtczak and Viemeister (2005) acknowledged that their fitted filters were "clearly too broad near the peak."

Overall, these results are consistent with the idea that the modulation filters have center frequencies that span a limited range, and that the top modulation filter has a center frequency below 180 Hz. The center frequency of the top filter appears to be about 100–120 Hz and may vary somewhat across listeners. If this interpretation is correct, then the slope of the TMTF for modulation frequencies above about 120 Hz should be determined by the slope of the high-frequency side of the top modulation filter. This latter slope can be estimated from the high-frequency sides of the curves in Figs. 2 and 3, for signal frequencies of 120 and 180 Hz. For the mean data, the slope is about −5 dB/oct (measured between 170 and 339 Hz for the 120-Hz signal frequency, and between 180 and 360 Hz for the 180-Hz signal frequency). This slope is very similar to the slopes of the TMTFs for 8- and 10-kHz sinusoidal carriers, as measured over the same range of modulation frequencies by Kohl-

rausch *et al.* (2000). Thus, this aspect of the data is also consistent with the idea that the modulation filters have center frequencies that span a limited range, and that the top modulation filter has a center frequency of about 100–120 Hz.

The data are not consistent with the idea that the modulation filters have center frequencies that span a broad range (including frequencies above 150 Hz), but are followed by a lowpass filter with a cutoff frequency of 150 Hz, as proposed by Ewert and Dau (2000). If that were the case, one would expect the signal threshold always to be maximal when the masker frequency equaled the signal frequency, but this did not happen for the highest signal frequency.

Some caution is needed in interpreting the data, as it is possible that the results were influenced by nonsensory factors. The highly modulated maskers may have led to attention being distracted from the signals, which had a barely detectable modulation depth. The magnitude of this effect would depend on the salience of the masker modulation, and this would tend to decrease as the masker modulation frequency increased above about 120 Hz. This could account for some of the asymmetry in the measured masking patterns for the signal modulation frequencies of 120 and 180 Hz. However, the temporal pattern of the stimuli was chosen so as to minimize any possible confusion of the masker and signal; there was a 40-ms unmodulated portion of the carrier between the masker and the signal. Also, our subjects were highly trained, so it seems doubtful that they would focus their attention on the masker rather than on the signal. Overall, while we cannot rule out a role of nonsensory factors, we feel that the distinct shifts in the peaks of the masking patterns found for the highest signal frequency are most readily explained using the concept that there are no modulation filters with center frequencies above about 120 Hz.

IV. CONCLUSIONS

Consistent with earlier results for signal frequencies up to 80 Hz (Wojtczak and Viemeister, 2005), modulation masking patterns obtained in forward masking showed tuning in the modulation domain for signal frequencies of 80, 120, and 180 Hz. The Q values, estimated using second-order Butterworth filters fitted to the data, were typically about 1.

For the 80-Hz signal frequency, the signal threshold was maximal when the masker frequency equaled the signal frequency for three of the four highly trained listeners. For the 180-Hz signal, the signal threshold was maximal when the masker frequency was below the signal frequency, for all listeners. For the 120-Hz signal, two listeners showed the former pattern, and two showed the latter pattern.

The center frequencies of second-order Butterworth filters fitted to the data were close to the signal frequency for the signal frequency of 80 Hz, but were significantly below the signal frequency for the signal frequencies of 120 and 180 Hz.

The results support the idea that the modulation filters span a limited range of center frequencies, and that the highest modulation filter has a center frequency of about 100–120 Hz, depending somewhat on the listener.

ACKNOWLEDGMENTS

This work was supported by an MRC grant (B.C.J.M.), an EU Marie Curie Fellowship (C.F.), and a grant from Deafness Research UK (B.C.J.M. and A.S.). We thank Torsten Dau, Stan Sheft, and Magdalena Wojtczak for helpful comments on an earlier version of this paper.

Ashihara, K., Kurakata, K., Mizunami, T., and Matsushita, K. (**2006**). "Hearing threshold for pure tones above 20 kHz," Acoust. Sci. & Tech. **27**, 12–19.

Buus, S., Florentine, M., and Mason, C. R. (**1986**). "Tuning curves at high frequencies and their relation to the absolute threshold curve," in *Auditory Frequency Selectivity*, edited by B. C. J. Moore and R. D. Patterson (Plenum, New York).

Dau, T. (**1996**). "Modeling auditory processing of amplitude modulation," Ph.D. thesis, University of Oldenburg, Germany.

Dau, T., Kollmeier, B., and Kohlrausch, A. (**1997a**). "Modeling auditory processing of amplitude modulation. I. Detection and masking with narrowband carriers," J. Acoust. Soc. Am. **102**, 2892–2905.

Dau, T., Kollmeier, B., and Kohlrausch, A. (**1997b**). "Modeling auditory processing of amplitude modulation. II. Spectral and temporal integration," J. Acoust. Soc. Am. **102**, 2906–2919.

Ewert, S. D., and Dau, T. (**2000**). "Characterizing frequency selectivity for envelope fluctuations," J. Acoust. Soc. Am. **108**, 1181–1196.

Ewert, S. D., Verhey, J. L., and Dau, T. (**2002**). "Spectro-temporal processing in the envelope-frequency domain," J. Acoust. Soc. Am. **112**, 2921–2931.

Füllgrabe, C., and Lorenzi, C. (**2003**). "The role of envelope beat cues in the detection and discrimination of second-order amplitude modulation," J. Acoust. Soc. Am. **113**, 49–52.

Füllgrabe, C., Moore, B. C. J., Demany, L., Ewert, S. D., Sheft, S., and Lorenzi, C. (**2005**). "Modulation masking produced by second-order modulators," J. Acoust. Soc. Am. **117**, 2158–2168.

Johnson-Davies, D., and Patterson, R. D. (**1979**). "Psychophysical tuning curves: Restricting the listening band to the signal region," J. Acoust. Soc. Am. **65**, 765–770.

Joris, P. X., and Yin, T. C. (**1992**). "Responses to amplitude-modulated tones in the auditory nerve of the cat," J. Acoust. Soc. Am. **91**, 215–232.

Joris, P. X., Schreiner, C. E., and Rees, A. (**2004**). "Neural processing of amplitude-modulated sounds," Physiol. Rev. **84**, 541–577.

Kay, R. H. (**1982**). "Hearing of modulation in sounds," Physiol. Rev. **62**, 894–975.

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.

Langner, G., and Schreiner, C. E. (**1988**). "Periodicity coding in the inferior colliculus of the cat. I. Neuronal mechanisms," J. Neurophysiol. **60**, 1799–1822.

Lorenzi, C., Berthommier, F., and Demany, L. (**1999**). "Discrimination of amplitude-modulation phase spectrum," J. Acoust. Soc. Am. **105**, 2987–2990.

Møller, A. R. (**1972**). "Coding of amplitude and frequency modulated sounds in the cochlear nucleus of the rat," Acta Physiol. Scand. **86**, 223–238.

Moore, B. C. J. (**2004**). "Dead regions in the cochlea: Conceptual foundations, diagnosis and clinical applications," Ear Hear. **25**, 98–116.

Moore, B. C. J., and Alcántara, J. I. (**2001**). "The use of psychophysical tuning curves to explore dead regions in the cochlea," Ear Hear. **22**, 268–278.

Moore, B. C. J., and Glasberg, B. R. (**2001**). "Temporal modulation transfer functions obtained using sinusoidal carriers with normally hearing and hearing-impaired listeners," J. Acoust. Soc. Am. **110**, 1067–1073.

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., 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, B. C. J., Sek, A., and Glasberg, B. R. (**1999**). "Modulation masking produced by beating modulators," J. Acoust. Soc. Am. **106**, 908–918.
O'Loughlin, B. J., and Moore, B. C. J. (**1981**). "Off-frequency listening: Effects on psychoacoustical tuning curves obtained in simultaneous and forward masking," J. Acoust. Soc. Am. **69**, 1119–1125.
Palmer, A. R. (**1995**). "Neural signal processing," in *Hearing*, edited by B. C. J. Moore (Academic, San Diego, CA).
Pumplin, J. (**1985**). "Low-noise noise," J. Acoust. Soc. Am. **78**, 100–104.
Schreiner, C. E., and Langner, G. (**1988**). "Coding of temporal patterns in the central auditory system," in *Auditory Function: Neurobiological Bases of Hearing*, edited by G. Edelman, W. Gall, and W. Cowan (Wiley, New York).
Sek, A., and Moore, B. C. J. (**1994**). "The critical modulation frequency and its relationship to auditory filtering at low frequencies," J. Acoust. Soc. Am. **95**, 2606–2615.
Sek, A., and Moore, B. C. J. (**2002**). "Mechanisms of modulation gap detection," J. Acoust. Soc. Am. **111**, 2783–2792.
Sek, A., and Moore, B. C. J. (**2003**). "Testing the concept of a modulation filter bank: The audibility of component modulation and detection of phase change in three-component modulators," J. Acoust. Soc. Am. **113**, 2801–2811.
Sek, A., and Moore, B. C. J. (**2004**). "Estimation of the level and phase of the simple distortion tone in the modulation domain," J. Acoust. Soc. Am. **116**, 3031–3037.
Strickland, E. A., and Viemeister, N. F. (**1996**). "Cues for discrimination of envelopes," J. Acoust. Soc. Am. **99**, 3638–3646.
Thornton, A. R., and Abbas, P. J. (**1980**). "Low-frequency hearing loss: Perception of filtered speech, psychophysical tuning curves, and masking," J. Acoust. Soc. Am. **67**, 638–643.
Verhey, J. L., Ewert, S. D., and Dau, T. (**2003**). "Modulation masking produced by complex tone modulators," J. Acoust. Soc. Am. **114**, 2135–2146.
Viemeister, N. F. (**1979**). "Temporal modulation transfer functions based on modulation thresholds," J. Acoust. Soc. Am. **66**, 1364–1380.
Wojtczak, M., and Viemeister, N. F. (**2005**). "Forward masking of amplitude modulation: Basic characteristics," J. Acoust. Soc. Am. **118**, 3198–3210.
Yasin, I., and Plack, C. J. (**2005**). "Psychophysical tuning curves at very high frequencies," J. Acoust. Soc. Am. **118**, 2498–2506.
Zwicker, E. (**1956**). "Die elementaren grundlagen zur bestimmung der informationskapazität des gehörs (The foundations for determining the information capacity of the auditory system)," Acustica **6**, 356–381.

Continuous versus discrete frequency changes: Different detection mechanisms?

Laurent Demany[a)]
Laboratoire Mouvement, Adaptation, Cognition (UMR CNRS 5227), Université de Bordeaux, BP 63, 146 Rue Leo Saignat, F-33076 Bordeaux, France

Robert P. Carlyon
MRC Cognition and Brain Sciences Unit, 15 Chaucer Road, Cambridge CB2 7EF, United Kingdom

Catherine Semal
Laboratoire Mouvement, Adaptation, Cognition (UMR CNRS 5227), Université de Bordeaux, BP 63, 146 Rue Leo Saignat, F-33076 Bordeaux, France

(Received 23 May 2008; revised 25 September 2008; accepted 16 November 2008)

Sek and Moore [J. Acoust. Soc. Am. **106**, 351–359 (1999)] and Lyzenga *et al.* [J. Acoust. Soc. Am. **116**, 491–501 (2004)] found that the just-noticeable frequency difference between two pure tones relatively close in time is smaller when these tones are smoothly connected by a frequency glide than when they are separated by a silent interval. This "glide effect" was interpreted as evidence that frequency glides can be detected by a specific auditory mechanism, not involved in the detection of discrete, time-delayed frequency changes. Lyzenga *et al.* argued in addition that the glide-detection mechanism provides little information on the direction of frequency changes near their detection threshold. The first experiment reported here confirms the existence of the glide effect, but also shows that it disappears when the glide is not connected smoothly to the neighboring steady tones. A second experiment demonstrates that the direction of a 750 ms frequency glide can be perceptually identified as soon as the glide is detectable. These results, and some other observations, lead to a new interpretation of the glide effect, and to the conclusion that continuous frequency changes may be detected in the same manner as discrete frequency changes.
 [DOI: 10.1121/1.3050271]

PACS number(s): 43.66.Mk, 43.66.Fe [MW] Pages: 1082–1090

I. INTRODUCTION

How do listeners detect continuous frequency changes? This is an important question because changes of that kind abound in speech and other meaningful sounds.

It is clear that, in some cases, continuous frequency changes may be detected by means of "static" spectral cues. Imposing a frequency modulation (FM) on a pure tone widens its power spectrum (Hartmann, 1997). This spectral-width cue will be usable, for instance, if the task is to detect a sinusoidal FM with a rate of several tens of hertz. A similar cue may be used if the task is to detect unidirectional frequency glides in tone bursts with a very short duration. However, since the spectral analysis of sounds by the auditory system is performed using filters with short impulse responses (Moore, 2004, Chaps. 1 and 5), the detection of *slow* FM in stimuli lasting several hundreds of milliseconds is presumably not based on static spectral cues. How is FM detected in such conditions?

The simplest hypothesis, often called "the snapshot hypothesis," is that FM is detected by means of comparisons between frequency samples taken at different times, as if the stimulus actually consisted of successive tone bursts. Hartmann and Klein (1980) proposed a mathematical model of FM detection based on this assumption. They showed that the model correctly predicted, among other things, differences between the psychometric functions obtained for the detection of sinusoidal FM and for the discrimination between two successive steady tone bursts differing in frequency. In the same vein, Demany and Semal (1989) showed that the frequency dependence of thresholds in the sinusoidal-FM detection task can be accounted for on the basis of the snapshot hypothesis, at least up to 4 kHz.

An alternative hypothesis, on which we focus here, is that continuous frequency changes can be detected by a specific mechanism, not involved in the detection of frequency differences between temporally separate steady tones. This "dynamic mechanism" (in the words of Dooley and Moore, 1988) would encode FM as a primary feature of sounds. Such a view is consistent with the fact that, in the auditory cortex of mammals, many neurons respond in a strong and selective manner to frequency glides (e.g., Whitfield and Evans, 1965; Zhang *et al.* 2003).

About three decades ago, psychophysical support for the dynamic-mechanism hypothesis was looked for in experiments that aimed at demonstrating selective adaptation effects in the FM domain. The idea was that the dynamic-mechanism hypothesis would be supported if it appeared that the detection of a given FM was impaired by repeated previous presentations of the same FM with a larger modulation depth, while being less affected by previous presentations of

a)Author to whom correspondence should be addressed. Electronic mail: laurent.demany@u-bordeaux2.fr

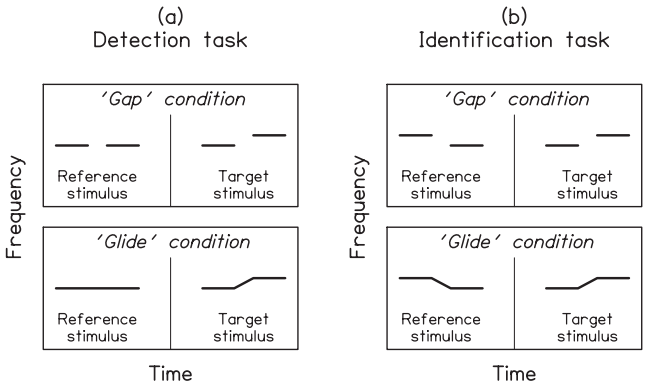


FIG. 1. Schematic spectrogram of stimuli used by Sek and Moore (1999) and Lyzenga *et al.* (2004). In the detection task, listeners had to discriminate target stimuli including a frequency change from reference stimuli with a steady frequency. In the identification task [not used by Sek and Moore (1999)], listeners had to discriminate target stimuli including a frequency rise from reference stimuli including a frequency fall. Both tasks were performed in a gap condition, where each stimulus consisted of two separate tone bursts, and in a glide condition where each stimulus had a continuous waveform.

stimuli including another type of FM, or amplitude modulation (AM) instead of FM. Several research groups did report selective adaptation effects of this type (for a review, see Kay, 1982). However, the data are now considered unconvincing, in part due to methodological problems (Wakefield and Viemeister, 1984), and also because the reported effects seem to disappear in trained listeners (Moody *et al.* 1984).

More recently, a quite different argument has been put forward in support of the dynamic-mechanism hypothesis. In two studies (Sek and Moore, 1999; Lyzenga *et al.* 2004), thresholds for the detection of frequency changes were measured using stimuli schematized in Fig. 1(a). On each trial, the listener was presented with two successive stimuli. One of them, the target stimulus, included a frequency change; the other (reference) stimulus did not. The task was to indicate if the target was the first or the second stimulus. In one condition, termed the "gap" condition, each stimulus consisted of two successive pure tones, with a silent gap in between. In a second condition, termed the "glide" condition, these two pure tones were no longer separated by a gap but smoothly connected to each other; so, the reference stimulus became a single pure tone and the target stimulus included a frequency glide smoothly connecting two frequency plateaux. For various durations of the gap or glide (5–200 ms) and of the frequency plateaux, it was found that the just-detectable frequency change was smaller in the glide condition than in the gap condition. To account for this finding, the authors argued that, in the glide condition, change detection was based on a combination of two cues: (1) a cue derived from a comparison between frequency samples taken at the beginning and the end of the stimulus, as in the gap condition; and (2) a cue provided by the glide itself. It could be reasonably assumed that the latter cue was not a static spectral cue. The authors thus interpreted this cue as the output of a change-detection mechanism responding exclusively to continuous or instantaneous changes.

In the two conditions described above, the listeners' task was merely to *detect* frequency changes. However, Lyzenga *et al.* (2004) also used the gap and glide conditions depicted in Fig. 1(b). On each trial, this time, both of the presented stimuli included a frequency change; the two changes had the same magnitude but opposite directions, and the task was to indicate if the upward change took place in the first or the second stimulus. Therefore, the listeners now had to *identify* the direction of frequency changes, which again varied in magnitude across trials. It appeared that, contrary to the detection thresholds, the identification thresholds were not significantly different in the gap and glide conditions. This led Lyzenga *et al.* (2004) to suggest that the dynamic change-detection mechanism provides little or no information about the direction of a frequency glide.

Here, we report new experiments that are closely related to those of Sek and Moore (1999) and Lyzenga *et al.* (2004). Experiment 1 tested the idea that the advantage of the glide condition over the gap condition in the detection task does not stem from the existence of a mechanism detecting the glide, but stems instead from the sole fact that the glide smoothly connects the two frequency plateaux. It was thought that this smooth connection could facilitate the detection of a difference between the plateaux, because the memorization of the first plateau might be improved by the transformation of a succession of two "auditory objects" into a single auditory object. Experiment 2 assessed the ability of listeners to identify the direction of both continuous and discrete frequency changes at their respective detection thresholds. According to the conclusions of Lyzenga *et al.* (2004), it should be difficult to identify the direction of a just-detectable frequency glide. In contrast, the snapshot hypothesis predicted that direction identification should not be more difficult for a glide than for a discrete change taking place following a gap.

II. EXPERIMENT 1

A. Method

1. Task and conditions

On each trial, two successive stimuli separated by a 700 ms interval were presented to the listener. One of them (the target) included a downward frequency change whereas the other (reference) included no frequency change. The listener had to indicate if the target was the first or the second stimulus, these two possibilities being a priori equiprobable. In order to force the listeners to base their judgments on *within-stimulus* frequency changes (see in this respect Sek and Moore, 1999), the center frequency of the stimuli was roved within trials. For each stimulus, this center frequency was selected randomly between 400 and 2400 Hz, the probability distribution being rectangular on a log-frequency scale. Four conditions, illustrated in Fig. 2, were run.

In condition 1, each stimulus consisted of two successive 250 ms sinusoidal tones, which had steady frequencies (differing from each other in the target stimulus) and were separated by a 250 ms silent gap. The tones had a nominal sound pressure level (SPL) of 65 dB and random initial phases. They were gated on and off with 10 ms cosinusoidal amplitude ramps.

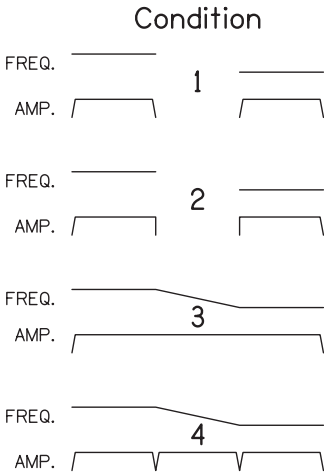


FIG. 2. Schematic representation of the target stimuli used in the four conditions of experiment 1. In condition 1, each stimulus consisted of two successive sine tones, gated on and off with 10 ms ramps and separated by a silent gap; the second tone was lower in frequency than the first one. Condition 2 was identical except that the amplitude ramps limiting the central gap were much more abrupt. In condition 3, the central section of the stimuli was a falling frequency glide; the stimuli had a continuous waveform and their amplitude envelope was flat. In condition 4, frequency varied exactly as in condition 3, but the amplitude envelope was no longer flat because the three successive sections of the stimuli were gated on and off with 10 ms amplitude ramps. In all conditions, the amplitude ramps were cosinusoidal rather than linear.

Condition 2 was identical to condition 1 except for one point: In this second condition, the 250 ms gap located in the center of the stimuli was bounded by amplitude ramps of only 0.1 ms instead of 10 ms.

In condition 3, the gap was replaced, for the target stimulus, with a 250 ms frequency glide connecting 250 ms frequency plateaux without any change in amplitude or discontinuity in the stimulus waveform. This glide was linear on a logarithmic frequency scale. The reference stimulus was simply a 750 ms pure tone.

In the fourth and last condition, each stimulus consisted of three consecutive 250 ms tones, gated on and off with 10 ms cosinusoidal amplitude ramps. The onset ramps of the second and third tones started immediately after the end of the preceding offset ramps, so that the stimuli again had a total duration of 750 ms. The three tones making up the reference stimulus had identical steady frequencies. In the target stimulus, the first and third tones had steady but different frequencies, and the middle tone was a frequency glide starting with the frequency of the first tone and ending with the frequency of the third tone; as in condition 3, this glide was linear on a logarithmic frequency scale.

In each condition, change-detection thresholds were measured as described in Sec. II.A.2. Given the findings of Sek and Moore (1999) and Lyzenga *et al.* (2004), it was expected that thresholds would be lower in condition 3 than in conditions 1 and 2. With respect to condition 4, two opposite predictions could be made. If the advantage of condition 3 over conditions 1 and 2 stemmed from the detectability of a frequency change during the glides, then a reasonable prediction was that performance in condition 4 would be more similar to performance in condition 3 than to performance in conditions 1 and 2. Alternatively, it could be supposed that the glides of condition 3 were advantageous not because they provided information but merely because they made the target stimuli continuous. Under this assumption, a logical prediction was that performance in condition 4 would be more similar to performance in condition 1 than to performance in condition 3. If any effect of discontinuity were due to the spectral splatter associated with the transition between sound and silence, rather than to the introduction of a silent gap, then performance in condition 2 might be somewhat worse than in condition 1 since the gap located in the center of the stimuli for these conditions had sharper bounds in condition 2 than in condition 1. The very sharp transitions of condition 2 produced more spectral splatter than the smoother transitions of condition 1.

2. Procedure and listeners

Thresholds were measured in separate blocks of trials for the four conditions, using an adaptive procedure tracking the 75% correct point on the psychometric function (Kaernbach, 1991). In each block of trials, the frequency change to be detected (C) initially had a magnitude of 60 cents (1 cent=1/100 semitone=1/1200 octave). C was decreased following each correct response, and increased following each incorrect response. A block ended after the 14th reversal in the variation of C. Up to the fourth reversal, C was multiplied by 2.25 when it was increased, and divided by the cube root of the same factor when it was decreased. After the fourth reversal, C was either multiplied by 1.5 or divided by the cube root of this factor. The threshold measured in a block of trials was defined as the geometric mean of all the C values used from the fifth reversal onwards.

Listeners were tested individually in a triple-walled sound-attenuating booth (Gisol, Bordeaux). They wore headphones (Sennheiser HD265), through which the stimuli were delivered binaurally. The stimuli were generated via 24 bit digital-to-analog converters (RME) at a sampling rate of 44.1 kHz. Responses were given by means of mouse-clicks on two virtual buttons on a monitor screen, and were immediately followed by visual feedback. Response times were not limited. Within a block of trials, there was a pause of about 700 ms between each response and the onset of the next stimulus. Each experimental session consisted of two or three sequences of four blocks, within which each condition was used once, in a random position (1, 2, 3, or 4). Overall, the collected data consisted of 16 threshold measurements per condition and listener.

Five listeners were tested: four students who were in their twenties (L1, L2, L3, L4) and the first author (L5, 53). All listeners had normal audiograms up to 4 kHz and previous experience in similar tasks. They were given a few practice sessions before the experiment proper.

B. Results

Figure 3 displays the geometric means of the 16 threshold estimates made for each listener and condition (open symbols), as well as the grand geometric mean for each condition (filled circles). The geometric standard errors of the data points corresponding to the open symbols have an av-

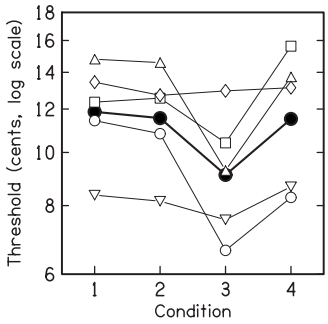


FIG. 3. Results of experiment 1. The open symbols represent the thresholds of the five listeners (L1: circles; L2: upward-pointing triangles; L3: downward-pointing triangles; L4: diamonds; L5: squares) for each condition. The filled circles represent the geometric means of these individual thresholds.

erage value of 8.8% (range: 4.3%–14.2%). It can be seen that, globally, performance was similar in conditions 1, 2, and 4, but better in condition 3. However, the effect of condition on performance was much larger for some listeners than for others. For each of the four planned pairwise comparisons (conditions 1 versus 2, 1 versus 3, 1 versus 4, and 3 versus 4), a two-way analysis of variance (ANOVA) (condition × listener) was performed on the logarithms of the threshold estimates. These ANOVAs confirmed that performance in condition 3 differed significantly from performance in condition 1 [$F(1,150)=26.1$, $P<0.001$] and condition 4 [$F(1,150)=24.5$, $P<0.001$], while performance in condition 1 did not differ from performance in conditions 2 and 4 ($F<1$ in each case). However, a significant interaction between the condition and listener factors was found in the comparisons between conditions 1 and 3 [$F(4,150)=3.9$, $P=0.005$], 1 and 4 [$F(4,150)=3.1$, $P=0.02$], and 3 and 4 [$F(4,150)=2.5$, $P=0.04$].

In order to check that the thresholds obtained in condition 4 were significantly more similar to those obtained in condition 1 than to those obtained in condition 3, we performed an additional ANOVA on the differences observed, within each sequence of four threshold measurements, between conditions 1 and 4, and between conditions 4 and 3; the processed data were again the logarithms of the threshold estimates. This two-way ANOVA [type of difference (1–4 versus 4–3) × listener] did reveal a significant main effect of type of difference [$F(1,150)=7.0$, $P=0.009$]; the interaction of the two factors was also found to be significant [$F(4,150)=2.9$, $P=0.025$].

C. Discussion

The fact that thresholds were significantly lower in condition 3 than in condition 1 confirms the main finding of Sek and Moore (1999) and Lyzenga *et al.* (2004). But on the other hand, their interpretation of this finding is seriously challenged by the fact that the mean threshold measured in our condition 4 was much more similar to the threshold obtained in condition 1 (or 2) than to the threshold obtained in condition 3. This observation indicates that the advantage of condition 3 over condition 1 may stem merely from the temporal continuity of the stimuli used in condition 3 rather than from the existence of "glide detectors."

Actually, support for this hypothesis is provided by some of the results of Lyzenga *et al.* (2004). Their experiment on change detection included, in addition to the gap and glide conditions depicted in Fig. 1(a), a "noise" condition in which the central portion of the stimuli was neither a gap nor a frequency glide but a noise burst. This noise burst was presented at a relatively high level, and therefore elicited a continuity illusion. In the target stimuli, the listeners could hear illusory glides smoothly connecting the two frequency plateaux. Surprisingly, thresholds in this noise condition were significantly lower than thresholds in the gap condition. To account for that, the authors supposed that the dynamic mechanism detecting the real glides of the glide condition was also able to detect illusory glides. However, there is a somewhat circular aspect to this reasoning. When a frequency change was just-detectable in the noise condition, the same frequency change was not detectable in the gap condition. Hence, one would have to assume that the auditory system introduced an illusory glide between two tones whose frequencies it could not otherwise discriminate, and then used the glide to detect the frequency difference, rather like Baron Munchausen pulling himself out of a bog by his own hair. It seems more parsimonious to hypothesize that the noise condition was advantageous merely because in that condition the stimuli were perceived as continuous.

Returning to the results of the present experiment, one should note that the advantage of condition 3 over the other conditions might be ascribed to the use of spectral cues by the listeners. Although the target stimuli of condition 3 had a perfectly continuous waveform, some spectral splatter was produced in these stimuli when the initial frequency plateau suddenly became a frequency glide, and when the glide suddenly became a new plateau. This spectral splatter may have provided a cue since it was absent in the reference stimuli. In order to minimize the influence of that cue in their glide condition, Lyzenga *et al.* (2004) presented their stimuli in a background of pink noise. We did not do so. However, it will be seen that the results of our second experiment discredit the idea that the advantage of condition 3 stemmed from the use of spectral cues in this condition.

III. EXPERIMENT 2

A. Rationale

According to Lyzenga *et al.* (2004), the auditory system contains a mechanism specific for the detection of dynamic acoustic changes such as frequency glides, but this dynamic mechanism is not sensitive to the direction of a glide, or at least does not facilitate the identification of its direction. Such a view implies that listeners should be unable to identify the direction of a glide at its detection threshold. By contrast, as pointed out in the Introduction, a prediction of the snapshot hypothesis is that direction identification at the detection threshold should not be systematically more difficult for continuous changes than for discrete, time-delayed changes. In experiment 2, using both continuous frequency

changes and discrete ones, we compared the magnitude that a given type of change must have in order to be just-detectable to the magnitude that the same type of change must have in order to be reliably identified as an upward change or a downward change. One and the same psychophysical paradigm was employed to measure the detection thresholds and the identification thresholds. This paradigm had been previously employed by Semal and Demany (2006) in a study investigating exclusively the perception of discrete changes.

B. Method

On each trial, as in experiment 1, the listener was presented with two successive stimuli separated by a 700 ms interval: a target stimulus containing a frequency change and a reference stimulus in which frequency did not change. Again, the target stimulus was either the first or the second stimulus, equiprobably. This time, however, the direction of the frequency change was no longer fixed: Frequency could go up or down, equiprobably. In separate blocks of trials, which did not differ from each other with respect to the stimulus characteristics, the listeners performed two different tasks: a detection (*D*) task and an identification (*I*) task. In the *D* task, one had to indicate if the target was the first or the second stimulus. In the *I* task, one had to indicate if frequency changed upwards or downwards. As in experiment 1, the center frequency of the stimuli was roved within trials, and could take any value between 400 and 2400 Hz. In each block of trials, again, a fixed type of frequency change was produced, and the magnitude of the frequency change (in cents) was varied across trials in order to estimate a threshold defined as the magnitude of change for which the probability of a correct response was 0.75.

Five types of frequency change were used, yielding five experimental conditions. In condition 1, the stimuli were the same as those used in condition 1 of experiment 1, except for the randomization of change direction. In this condition, therefore, the *D* and *I* tasks were performed on stimuli including a central gap and the data provided information on the perception of discrete frequency changes. The frequency changes produced in the other four conditions were continuous. In condition 2, the stimuli were the same as those used in condition 3 of experiment 1, except for the randomization of change direction; each target stimulus thus consisted of a 250 ms frequency glide smoothly connecting 250 ms frequency plateaux, in a linear manner on a log-frequency scale. In the three remaining conditions, the target stimuli were "pure" frequency glides, in which frequency was constantly varying, linearly on a log-frequency scale. These glides, and the corresponding reference stimuli with a steady frequency, had a duration of 750 ms in condition 3, 250 ms in condition 4, and 50 ms in condition 5. In all conditions, the stimuli were gated on and off with 10 ms cosinusoidal amplitude ramps and had a nominal SPL of 65 dB.

The adaptive procedure used to measure thresholds was exactly the same as that employed in experiment 1, except for the initial magnitude of the frequency change presented to the listeners; this initial magnitude was 100 cents in the first four conditions, and 300 cents in condition 5 (because thresholds were markedly higher in the latter condition). In each experimental session, ten threshold measurements were made, one for each combination of condition and task (*D* or *I*). These ten measurements were made in a random order. Overall, the collected data consisted of 16 threshold measurements per condition, task, and listener.

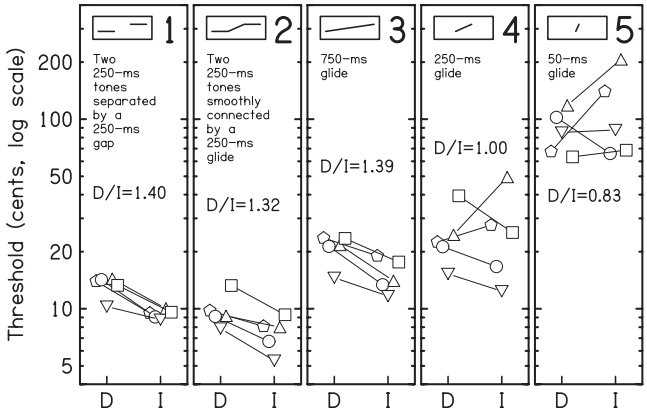


FIG. 4. Results of experiment 2. Detection (*D*) and identification (*I*) thresholds of the five listeners in the five conditions. For each condition, the quotient *D*/*I* of the geometric mean of the *D* thresholds and the geometric mean of the *I* thresholds is also indicated. The identical symbols represent the same listener in Figs. 3 and 4.

We shall report here the data provided by five listeners. Four of them were identified as L1, L2, L3, and L5 in experiment 1; the fifth listener (L6), who did not participate in experiment 1, was an audiometrically normal student in her twenties. Four additional listeners were tested, but their results will not be considered below because they had difficulties in the *I* task when the frequency changes were discrete (first condition).[1] Semal and Demany (2006) previously pointed out that, for some audiometrically normal listeners, it is difficult to identify the direction of very small but nonetheless well-detected frequency changes between two successive pure tones separated by a gap. Such listeners appeared to be inefficient *detectors* of frequency changes in Semal and Demany's study (2006).

C. Results and discussion

Figure 4 displays the geometric mean of the 16 threshold estimates made for each condition, task, and listener; in this figure and in Fig. 3, identical symbols represent the same listener. The geometric standard errors of the data points have an average value of 9.6% (range: 4.6%–18.3%). For each of the five conditions, we also indicate in Fig. 4 the statistic *D*/*I* obtained when the geometric mean of all the *D* thresholds is divided by the geometric mean of all the *I* thresholds.

In the first condition, involving discrete changes, each listener had a higher threshold in the *D* task than in the *I* task; *D*/*I* was equal to 1.40. An identical condition had been used by Semal and Demany (2006), and they obtained very similar results from their best three subjects (those who had the lowest thresholds in the two tasks). As explained in detail by Micheyl *et al.* (2008) (see also Semal and Demany, 2006), the standard version of signal detection theory—i.e., the

constant-variance Gaussian model—predicted in this condition (as well as the other four conditions used here) a D/I ratio of 1.56 for an ideal listener identifying always correctly the direction of a perceived frequency change. The empirically obtained D/I, 1.40, is close to this theoretical value. For an ideal listener identifying always correctly the direction of a perceived change, the "high-threshold" theory (Green and Swets, 1974, Chap. 5) predicted a D/I ratio of 1, i.e., a ratio *lower* than 1.40. Therefore, it is reasonable to conclude that the five listeners who provided the data analyzed here were able to identify the direction of a discrete frequency change as soon as this change was detectable.

Consider now the results obtained in condition 2, where the frequency changes were continuous. It can be noted first that each listener had a lower D threshold in this condition than in condition 1. This confirms the main finding of Sek and Moore (1999) and Lyzenga *et al.* (2004), as well as observations that we made in experiment 1. However, a more important result is that each listener also had a lower I threshold in condition 2 than in condition 1; the corresponding effect is statistically significant [$t(4)=3.66$; $P=0.02$]. From condition 1 to condition 2, D/I decreased, but only very slightly, as indicated in Fig. 4; this global trend was observed in only three of the five listeners. The fact that the I thresholds were lower in condition 2 than in condition 1 is not consistent with the idea that the advantage provided by the glides in condition 2 stems from the existence of a dynamic change-detection mechanism detecting glides without providing information on their direction. Another idea discredited by this finding is that, in the D task, the advantage of condition 2 over condition 1 could be merely due to the detection of spectral splatter in the transitions between the glides and the frequency plateaux flanking them: This spectral splatter provided no useful cue in the I task since it did not depend on the direction of the frequency changes (the ascending stimuli being temporal inversions of the descending stimuli).

In condition 3, the stimuli had the same duration as in conditions 1 and 2 (750 ms) and the targets had a constantly gliding frequency. As shown in Fig. 4, the D and I thresholds were consistently worse than in condition 1 (and 2), but D/I had almost exactly the same value as in condition 1. Both of these findings support the snapshot hypothesis. It was predicted by the snapshot hypothesis that thresholds would be worse than in condition 1 because taking a frequency sample of finite duration must result in some averaging of consecutive instantaneous frequencies; this averaging reduced the "effective" frequency span of the target stimuli in condition 3, but not condition 1. It is possible that some dynamic change-detection mechanism was sensitive to the frequency glides of condition 3 when they were above their detection threshold. However, our results suggest that at threshold, these glides were detected by means of snapshot comparisons.

In conditions 4 and 5, where the target stimuli consisted again of pure frequency glides but had shorter durations (250 and 50 ms, respectively), the thresholds increased, especially from condition 4 to condition 5. This trend is consistent with observations by Lyzenga *et al.* (2004). The I thresholds increased to a larger extent than the D thresholds, so that D/I decreased. This effect was strongly listener dependent, but it is clear that, overall, the shortening of stimulus duration reduced the listeners' ability to identify the direction of the frequency changes at their detection threshold. From the point of view of the snapshot hypothesis, the rise of the D and I thresholds makes sense: Detecting a frequency change within a short stimulus may be difficult because, in this case, the listener cannot sample frequency using a temporal window which is both short enough to allow for a comparison between nonoverlapping samples and long enough for accurate frequency measurements (Moore, 1973). In contrast, the reason why change detection should be more difficult for short stimulus durations is not obvious under the dynamic-mechanism hypothesis: Shortening a frequency glide spanning a fixed frequency distance increases, of course, the speed of frequency change; one might expect this increase in speed to have a favorable, rather than deleterious, effect for a dynamic change-detection mechanism.

We shall consider in Sec. IV how the decrease in D/I from condition 3 to condition 5 can be accounted for. The most important finding of the present experiment is that in conditions 2 and 3, as well as in condition 1, D/I was such that the listeners could be assumed to identify correctly the direction of a change as soon as they detected it. In conditions 2 and 3, our results are at odds with the idea that continuous (as opposed to discrete) frequency changes are optimally detected by an auditory mechanism providing no information about change direction. This idea had been put forth by Lyzenga *et al.* (2004) to account for their own data. In their study, however, D and I thresholds were not measured within the same experiment and compared to each other. Instead, Lyzenga *et al.* (2004) made only within-task (D or I) comparisons, as described in our Introduction and illustrated in Fig. 1. Consider again this figure. On each trial run in the I task, frequency changed by some amount a in the target stimulus, and $-a$ in the reference stimulus; this resulted in a difference of $2a$ between the two stimuli. In the D task, on the other hand, the corresponding difference was smaller by a factor of 2. So, for an ideal listener, at least in the gap condition, the threshold ratio D/I should have been equal to 2. An examination of the thresholds plotted in Figs. 2–5 of Lyzenga *et al.*, 2004 reveals that in the gap condition, D/I actually had a mean value of only 0.9. It thus seems that even in the gap condition, at least some of the listeners tested by Lyzenga *et al.* (2004) failed to identify correctly the direction of changes that they nonetheless reliably detected. Listeners showing this difficulty are not very uncommon (Semal and Demany, 2006). Indeed, as pointed out above, a sample of this population was tested in the present experiment. We did not report in detail the corresponding data here because they do not provide clear information concerning the main issue.

IV. EXPERIMENT 3

In experiment 2, D/I decreased from condition 3 to condition 5. One can make sense of this observation in the framework of the snapshot hypothesis by assuming that, in

TABLE I. Performance of the four subjects of experiment 3 in that experiment and in conditions 1 and 5 of experiment 2. These four listeners (L2, L3, L5, and L6) are, respectively, identified by upward-pointing triangles, downward-pointing triangles, squares, and pentagons in Fig. 4. The last column of the table displays the geometric means of the individual values.

	L2	L3	L5	L6	Mean
Experiment 3					
D threshold (cents)	93.1	60.6	71.1	67.8	72.2
D/*I*	0.35	1.19	1.41	0.65	0.79
Experiment 2, condition 5					
D threshold (cents)	118.2	85.1	63.3	67.5	81.0
D/*I*	0.57	0.97	0.92	0.48	0.70
Experiment 2, condition 1					
D threshold (cents)	14.4	10.3	13.3	14.0	12.9
D/*I*	1.43	1.17	1.39	1.47	1.36

the short target stimuli of conditions 4 and 5, the listeners sometimes misjudged the temporal order of frequency samples that they had correctly differentiated. We tested that idea in a small final experiment where the *D* and *I* tasks were performed on stimuli consisting of two successive 25 ms tones, gated on and off by means of 10 ms cosine ramps, with no silent interval in between. These two tones had steady frequencies, differing from each other in the target stimuli. The offset ramp of the first tone and the onset ramp of the second were sufficient to produce a well-audible gap between the two tones. The stimuli had the same overall duration (50 ms) as those used in condition 5 of experiment 2, but were structurally more similar to those of condition 1. Four listeners served as subjects. These were L2, L3, L5, and L6, the four listeners for whom, in condition 5 of experiment 2, *D*/*I* had been smaller than 1. L1 (identified by circles in Figs. 3 and 4) was not recruited again because in her case *D*/*I* had been similar in all conditions. Using the same procedure as before, 16 threshold measurements were made for each listener and task.

Table I displays the geometric means of these 16 measurements for the *D* task, and the associated values of *D*/*I*. We also indicate in this table the corresponding statistics for conditions 1 and 5 of experiment 2. It can be seen that for each listener, the *D* threshold measured in the present experiment was not very different from the *D* threshold measured previously in condition 5. With respect to *D*/*I*, however, there were substantial individual differences. For two listeners (L2 and L6), the new *D*/*I* was smaller than 1 and much closer to the value previously found in condition 5 than to the value found in condition 1. For the other two listeners (L3 and L5), the opposite was true.

The small *D*/*I* ratio observed here for L2 and L6 presumably originates from time-order confusions. For these two listeners, therefore, it may well be that the small *D*/*I* ratio obtained in condition 5 of experiment 2 also originated from time-order confusions. For L3 and especially L5, on the other hand, this is unlikely; their behavior here suggests that, in condition 5 of experiment 2, they did not detect the glides as specified by the snapshot hypothesis but used a different cue.[2] It is conceivable that the cue in question was the output of a dynamic change-detection mechanism which does not provide optimal information about change direction. Such a mechanism would be more apt to detect frequency glides in short stimuli than in long ones because the slope of a just-detectable glide is larger when the stimulus is short, as shown by experiment 2 and several previous studies (Sergeant and Harris, 1962; Nabelek and Hirsh, 1969; Madden and Fire, 1997; Lyzenga *et al.*, 2004). However, an alternative possibility is the existence of a static spectral cue. At any given time, the instantaneous excitation pattern produced by a gliding tone in a bank of auditory filters should be somewhat less sharp than if the tone had a steady frequency. This relative spreading of the "instantaneous internal spectrum" is presumably detectable *per se* when the slope of the glide exceeds a certain threshold.

V. GENERAL DISCUSSION

The present study confirmed the existence of an intriguing auditory phenomenon initially described by Sek and Moore (1999) and subsequently investigated by Lyzenga *et al.* (2004): A small frequency difference between two successive tones (long enough to evoke a maximally salient pitch) is easier to detect when the tones are smoothly connected by a frequency glide than when they are separated by a silent gap of the same duration. Our study also shows, however, that the origin of this glide effect may not be the one suggested up to now. It was previously thought that the glide improves the detection of a frequency change because a change can be heard during the glide itself. We suggest instead that the glide is advantageous just because it connects the two tones in a continuous manner, and thus transforms a succession of two auditory objects into a single auditory object. In support of this alternative hypothesis, we found in experiment 1 that the glide is no longer advantageous when its two ends are disconnected from the neighboring tones. Another observation supporting the same hypothesis was reported by Lyzenga *et al.* (2004): They found that when the glide is replaced with a noise burst producing a continuity illusion, change detection is still better than when the tones are separated by a silent gap.

The glide effect of Sek and Moore (1999) clearly deserves to be elucidated because, if its initial interpretation

were correct, this effect would be the strongest piece of psychophysical evidence for a dynamic change-detection mechanism in the auditory system. Our alternative interpretation implies that such a mechanism may, in fact, not exist. However, we did not demonstrate that continuous frequency changes are *always* encoded by means of snapshot comparisons or on the basis of static spectral cues. It is possible that the dynamic mechanism exists but detects a frequency glide only when the speed of frequency change exceeds some value, below which snapshot comparisons could be more efficient. Cusack and Carlyon (2003) showed that detecting a tone containing sinusoidal FM among several steady tones is easier than the reverse (detecting a steady tone in a background of FM tones). They took this as evidence that the auditory system encodes FM as a primary sound feature, or in other words that the dynamic mechanism exists. More recently, Carlyon *et al.* (2004) found that although listeners heard the FM associated with a sinusoidally modulated tone continue when a portion of that tone was replaced with noise, they could not tell whether or not the tone resumed at the same FM phase as that which it would have had if it had really remained on. Carlyon *et al.* (2004) suggested from this observation that the auditory system is able to encode sinusoidal FM in a way that discards information on the phase of the FM, while preserving information such as the carrier frequency and the presence, depth, and rate of FM. In both of these previous studies, the depth and rate of the FM used were such that the FM was well above its normal detection threshold. Correlatively, instantaneous frequency changed at a relatively high speed. Therefore, the conclusion drawn from the two studies is not inconsistent with the idea that no dynamic mechanism is involved in the detection of FM when the speed of frequency change is low. Alternative explanations are that explicit encoding of dynamic changes applies only to periodic FM, or that it is more important in tasks, like the ones employed by Cusack and Carlyon (2003), where the cognitive load is more demanding than the detection or discrimination of frequency changes applied to isolated tones.

Another possibility is that the dynamic mechanism does not detect efficiently isolated glides, such as those used here in conditions 3–5 of experiment 2, or condition 4 of experiment 1, but works better when the stimulus is a frequency glide immediately preceded and/or followed by a frequency plateau *without any discontinuity*. Lyzenga *et al.* (2004) attempted to account for the *D* thresholds obtained in their glide condition [Fig. 1(a), lower panel] by assuming that in this condition the listeners combined two cues: a cue derived from a comparison between the two frequency plateaux and a cue provided during the glide itself by a dynamic change-detection mechanism. The first cue was assessed from the *D* thresholds measured in the gap condition, and the second from the *D* thresholds measured in a condition where the target stimulus consisted of nothing more than a frequency glide.[3] Lyzenga *et al.* (2004) found that performance in their glide condition was too good to be predicted in a simple manner from performance in the other two conditions. They assumed, therefore, that the two change-detection mechanisms providing the cues allegedly combined in the glide condition operated in a synergistic way. This apparent synergy was taken as evidence that frequency glides are more efficiently detected when they are preceded and/or followed by a frequency plateau. However, using the same equations as Lyzenga *et al.* (2004) we failed to find evidence for a synergistic effect in the results of the second experiment reported here: The *D* thresholds in condition 2 could be predicted from the *D* thresholds in conditions 1 and 4 by assuming simply that the internal noise limiting performance in condition 2 had two partially independent sources, respectively limiting performance in condition 1 and in condition 4. Under our hypothesis about the origin of the glide effect, the success of this prediction is fortuitous and listeners were, in fact, not using in condition 2 the information used in condition 4. It is warranted to assume that listeners extracted no information from the glides in condition 2 since they clearly extracted no information from the glides in condition 4 of experiment 1.

Although the results reported here are compatible with the idea that the auditory system contains a dynamic change-detection mechanism detecting efficiently frequency glides *when they are smoothly connected to frequency plateaux*, we think that our alternative interpretation of Sek and Moore's (1999) glide effect is more parsimonious. Admittedly, this alternative interpretation based on the snapshot hypothesis is incomplete: It remains to be explained why detecting a small difference between two frequency samples should be easier when they belong to the same auditory object than when this is not the case. But the snapshot hypothesis itself is obviously very reasonable. An important point is that snapshot comparisons are not necessarily less "automatic" than the dynamic change-detection mechanism suggested by Sek and Moore (1999) and Lyzenga *et al.* (2004). Indeed, Demany and Ramos (2005, 2007) (see also Demany *et al.*, 2008) showed that a frequency difference between two pure tones separated by a substantial time interval can be consciously perceived as an upward or downward pitch change even when the first of these tones has not been consciously perceived. This phenomenon strongly suggests that the auditory system contains automatic frequency-shift detectors. It may well be that these detectors respond not only to discrete, time-delayed frequency shifts, but also to continuous frequency changes.

ACKNOWLEDGMENTS

The authors are grateful to Marie Dejos and Maialen Erviti for their precious collaboration. They also thank Johannes Lyzenga and Brian Moore for beneficial discussions.

[1]We knew in advance that, for these four listeners, in the first condition, the *I* threshold would be markedly higher than the *D* threshold. This group was nevertheless tested in order to see if, for some listeners, it can be easier to identify the direction of a just-detectable frequency change when the change is continuous than when it is discrete. We did not observe significant trends in that direction. The four listeners' performance was markedly poorer in the *I* task than in the *D* task for all conditions, not only the first one. In this group, the ratios of the *D* and *I* thresholds were strongly correlated across conditions. In all conditions, the *D* thresholds were also systematically higher than those measured in the five listeners who provided the data analyzed below.

[2]For L3 and L5, D/I was larger than 1 in condition 4 of experiment 2,

where the target stimuli consisted of 250 ms frequency glides. Thus, the "different cue" hypothetically used by L3 and L5 was presumably used only for the 50 ms glides of condition 5.

[3]Lyzenga *et al.* (2004) used the classical model of signal detection theory to compute their predictions. Sek and Moore (1999) had previously analyzed their own data in a similar way, but their predictions were problematic because some of the relevant data were actually missing.

Carlyon, R. P., Micheyl, C., Deeks, J. M., and Moore, B. C. J. (**2004**). "Auditory processing of real and illusory changes in FM phase," J. Acoust. Soc. Am. **116**, 3629–3639.

Cusack, R., and Carlyon, R. P. (**2003**). "Perceptual asymmetries in audition," J. Exp. Psychol. Hum. Percept. Perform. **26**, 713–725.

Demany, L., Pressnitzer, D., and Semal, C. (**2008**). "On the binding of successive tones: Implicit versus explicit pitch comparisons," J. Acoust. Soc. Am. **123**, 3049.

Demany, L., and Ramos, C. (**2005**). "On the binding of successive sounds: Perceiving shifts in nonperceived pitches," J. Acoust. Soc. Am. **117**, 833–841.

Demany, L., and Ramos, C. (**2007**). "A paradoxical aspect of auditory change detection," in *Hearing—From Sensory Processing to Perception*, edited by B. Kollmeier, G. Klump, V. Hohmann, U. Langemann, M. Mauermann, S. Uppenkamp, and J. Verhey (Springer, Heidelberg), pp. 313–321.

Demany, L., and Semal, C. (**1989**). "Detection thresholds for sinusoidal frequency modulation," J. Acoust. Soc. Am. **85**, 1295–1301.

Dooley, G. J., and Moore, B. C. J. (**1988**). "Detection of linear frequency glides as a function of frequency and duration," J. Acoust. Soc. Am. **84**, 2045–2057.

Green, D. M., and Swets, J. A. (**1974**). *Signal Detection Theory and Psychophysics* (Krieger, New York).

Hartmann, W. M. (**1997**). *Signals, Sound, and Sensation* (AIP, Woodbury, New York).

Hartmann, W. M., and Klein, M. A. (**1980**). "Theory of frequency modulation detection for low modulation frequencies," J. Acoust. Soc. Am. **67**, 935–946.

Kaernbach, C. (**1991**). "Simple adaptive testing with the weighted up-down method," Percept. Psychophys. **49**, 227–229.

Kay, R. H. (**1982**). "Hearing of modulation in sounds," Physiol. Rev. **62**, 894–975.

Lyzenga, J., Carlyon, R. P., and Moore, B. C. J. (**2004**). "The effects of real and illusory glides on pure-tone frequency discrimination," J. Acoust. Soc. Am. **116**, 491–501.

Madden, J. P., and Fire, K. M. (**1997**). "Detection and discrimination of frequency glides as a function of direction, duration, frequency span, and center frequency," J. Acoust. Soc. Am. **102**, 2920–2924.

Micheyl, C., Kaernbach, C., and Demany, L. (**2008**). "An evaluation of psychophysical models of auditory change perception," Psychol. Rev., **115**, 1069–1083.

Moody, D. B., Cole, D., Davidson, L. M., and Stebbins, W. C. (**1984**). "Evidence for a reappraisal of the psychophysical selective adaptation paradigm," J. Acoust. Soc. Am. **76**, 1076–1079.

Moore, B. C. J. (**1973**). "Frequency difference limens for short-duration tones," J. Acoust. Soc. Am. **54**, 610–619.

Moore, B. C. J. (**2004**). *An Introduction to the Psychology of Hearing*, 5th ed. (Elsevier, Amsterdam).

Nabelek, I. V., and Hirsh, I. J. (**1969**). "On the discrimination of frequency transitions," J. Acoust. Soc. Am. **45**, 1510–1519.

Sek, A., and Moore, B. C. J. (**1999**). "Discrimination of frequency steps linked by glides of various durations," J. Acoust. Soc. Am. **106**, 351–359.

Semal, C., and Demany, L. (**2006**). "Individual differences in the sensitivity to pitch direction," J. Acoust. Soc. Am. **120**, 3907–3915.

Sergeant, R. L., and Harris, J. D. (**1962**). "Sensitivity to unidirectional frequency modulation," J. Acoust. Soc. Am. **34**, 1625–1628.

Wakefield, G. H., and Viemeister, N. F. (**1984**). "Selective adaptation to linear frequency-modulated sweeps: Evidence for direction-specific FM channels?," J. Acoust. Soc. Am. **75**, 1588–1592.

Whitfield, I. C., and Evans, E. F. (**1965**). "Responses of auditory cortical neurons to stimuli of changing frequency," J. Neurophysiol. **28**, 655–672.

Zhang, L. I., Tan, A. Y. Y., Schreiner, C. E., and Merzenich, M. M. (**2003**). "Topography and synaptic shaping of direction selectivity in primary auditory cortex," Nature (London) **424**, 201–205.

Characteristics of phonation onset in a two-layer vocal fold model

Zhaoyan Zhang[a)]
School of Medicine, University of California, Los Angeles, 31-24 Rehabilitation Center, 1000 Veteran Avenue, Los Angeles, California 90095-1794

(Received 10 June 2008; revised 14 October 2008; accepted 21 November 2008)

Characteristics of phonation onset were investigated in a two-layer body-cover continuum model of the vocal folds as a function of the biomechanical and geometric properties of the vocal folds. The analysis showed that an increase in either the body or cover stiffness generally increased the phonation threshold pressure and phonation onset frequency, although the effectiveness of varying body or cover stiffness as a pitch control mechanism varied depending on the body-cover stiffness ratio. Increasing body-cover stiffness ratio reduced the vibration amplitude of the body layer, and the vocal fold motion was gradually restricted to the medial surface, resulting in more effective flow modulation and higher sound production efficiency. The fluid-structure interaction induced synchronization of more than one group of eigenmodes so that two or more eigenmodes may be simultaneously destabilized toward phonation onset. At certain conditions, a slight change in vocal fold stiffness or geometry may cause phonation onset to occur as eigenmode synchronization due to a different pair of eigenmodes, leading to sudden changes in phonation onset frequency, vocal fold vibration pattern, and sound production efficiency. Although observed in a linear stability analysis, a similar mechanism may also play a role in register changes at finite-amplitude oscillations.
 [DOI: 10.1121/1.3050285]

PACS number(s): 43.70.Bk, 43.70.Gr [AL] Pages: 1091–1102

I. INTRODUCTION

The vocal folds are complex layered structures consisting of a muscular layer at the base, multiple layers of lamina propria in the middle, and an outmost epithelium layer. The geometry and mechanical properties of these different layers can be altered by laryngeal adjustments or due to vocal pathologies. Such changes often lead to qualitatively distinct vocal fold vibration and voice quality (Titze, 1994; Colton and Casper, 1996). An ultimate goal of voice production research is to understand and predict the acoustic consequences of such changes in geometry and biomechanical properties of the vocal folds. The influence on phonation onset frequency, phonation threshold pressure, and vocal fold vibration characteristics is of particular interest.

The influence of biomechanical properties of the vocal folds on phonation has been the focus of many previous works. Due to their simplicity, lumped-mass models were often used in these studies. For example, based on a body-cover representation of the vocal fold, Titze *et al.* (1988) used a string model to study pitch control mechanisms in human phonation. They showed that although the contraction of the cricothyroid (CT) muscle always causes the phonation frequency to increase, the contraction of the thyroarytenoid (TA) muscle may decrease or increase the phonation frequency depending on the effective depth of the vocal fold in vibration. Using a three-mass body-cover model, Story and Titze (1995) showed that a variety of vocal fold vibration patterns can be produced using different combinations of body and cover stiffnesses. Increasing body stiffness generally leads to lower body layer amplitudes and higher pitches. In a recent study, Tokuda *et al.* (2007) used a three-mass cover-only model to study chest-falsetto register transition (a sudden qualitative change in vocal fold vibration pattern, see, e.g., Titze, 1994; Svec *et al.*, 1999). They showed that such an abrupt transition occurs as a spontaneous shift in dominance between different eigenmodes of the vocal folds. In their study, such shift was induced by variation in the stiffness of the middle mass, which caused the vocal fold vibration to switch from being dominated by the first eigenmode of the vocal folds to being dominated by the third eigenmode of the vocal folds.

Although lumped-mass models provide valuable insights into the physics of phonation, their use in practical applications is limited. Because of the lack of direct correspondence to realistic, directly measurable properties of the vocal folds, model parameters of lumped-mass models need to be estimated. This is often difficult, if not impossible, as some model parameters are dynamic variables of the coupled fluid-structure system, which cannot be determined *a priori*. For example, Titze *et al.* (1988) showed that the effective depth of vibration is an important factor in the determination of phonation frequency. However, evaluation of the effective depth of vibration requires information on the vibration field within the vocal fold structure, which is generally unknown and highly depends on the specific biomechanical properties of the vocal folds. It is still unknown how much control the human has over the effective depth of vibration (Titze *et al.*, 1988). Similarly, the coupling stiffness between the upper and lower masses in the two-mass model and its variants (e.g., Ishizaka and Flanagan, 1972; Story and Titze, 1995) determines the phase difference between the two masses and

[a)]Electronic mail: zyzhang@ucla.edu

therefore the phonation threshold pressure (Titze, 1988). However, it remains unknown how the body and cover stiffnesses would affect this coupling stiffness. Estimation of these important model parameters is extremely difficult and often requires knowledge of the vocal fold vibration field that is to be solved for. This difficulty makes the lumped-mass models less appealing for use in practical applications in which the influence of the realistic vocal fold geometry and changes in local biomechanical properties needs to be evaluated. Furthermore, lumped-mass models may have oversimplified the underlying physics of phonation. For example, the superior-inferior component of the vocal fold motion, which has been shown to have a pronounced effect on phonation (Ishizaka and Flanagan, 1977; Titze and Talkin, 1979; Zhang *et al.*, 2006a, 2006b), was often neglected in lumped-mass models of phonation.

For practical applications, a better description of the underlying physics and a realistic representation of the vocal fold geometry are required. Based on continuum mechanics, continuum models allow vocal fold dynamics to be calculated from directly measurable parameters such as biomechanical properties (or preferably the degree of laryngeal muscle contraction) and realistic geometric parameters of the vocal folds. Recent studies (Zhang *et al.*, 2006a; Zhang, 2008) show that geometric details of the vocal folds may have a large impact on vocal fold vibration. A better understanding of how realistic vocal fold geometry would affect the vocal fold eigenmodes and the eigenmode synchronization process may also provide new insights into mechanisms of register change. Furthermore, continuum models allow a natural representation of the biomechanics of the layered structure of the vocal folds. Such realistic representation of the vocal fold biomechanics is particularly critical for the modeling of vocal pathologies, for which changes in biomechanical properties are often localized (e.g., local stiffening due to vocal fold scarring).

There has been an increasing amount of work on continuum modeling of phonation (e.g., Titze and Talkin, 1979; Alipour *et al.*, 2000; De Oliveira Rosa *et al.*, 2003; Thomson *et al.*, 2005; Tao and Jiang, 2006). However, the influence of biomechanical properties and geometry of the vocal folds on phonation has not been systematically investigated. Such investigation would require a scan of the dynamic behavior of the coupled fluid-structure system over a large range of the parameter space, which is generally computationally expensive when continuum models are used. Such an approach was used in Titze and Talkin (1979) but at the cost of a reduced spatial resolution. Using a linear stability analysis approach, the influence of geometric and biomechanical properties of the vocal fold on phonation onset can be investigated at a less demanding computational cost, as shown in a recent study (Zhang *et al.*, 2007). In that paper, mechanisms of phonation onset were investigated in an isotropic two-dimensional continuum vocal fold model. They showed that the primary mechanism of phonation onset is due to the mode-coupling effect of the flow-induced stiffness, which causes two vocal fold eigenmodes to synchronize to produce an unstable eigenmode: Synchronization of two eigenmodes at the same frequency allows the flow pressure induced by one eigenmode to interact with the tissue velocity of the other eigenmode. Synchronization of two eigenmodes at the same frequency but different phases establishes a flow pressure field that is at least partially in-phase with the vocal fold velocity, allowing the cross-mode interaction to establish a net energy flow from airflow into the vocal fold tissue.

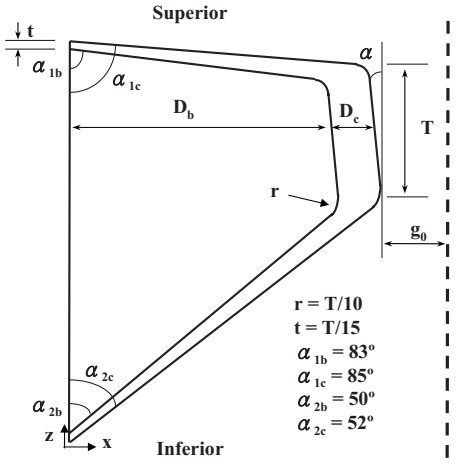


FIG. 1. The two-dimensional vocal fold model and the glottal channel. The coupled vocal fold-flow system was assumed to be symmetric about the glottal channel centerline, and only the left half of the system was considered in this study. T is the thicknesses of the medial surface of the vocal fold in the flow direction; D_b and D_c are the depths of the vocal fold body and cover layers at the center of the medial surface, respectively; g_0 is the minimum prephonatory glottal half-width of the glottal channel at rest. The divergence angle α is the angle formed by the medial surface of the vocal fold with the z-axis. Other parameters include the thickness of the cover layer at the base of the vocal fold, t, the rounding fillet radius, r, and the entrance and exit angles of the body and cover layers. The dashed line indicates the glottal channel centerline.

In this study, the influence of changes in geometry and biomechanical properties of the vocal fold on phonation onset is investigated in a continuum model of the vocal folds. The focus is to investigate the influence of the stiffness of vocal fold body and cover and vocal fold geometry on phonation threshold pressure, phonation onset frequency, vocal fold vibration, and sound production efficiency. A two-layer body-cover model of the vocal fold, as suggested by Hirano (1974), is used. The body layer consists of the muscle fibers, and the deep layer of the lamina propria. The cover layer consists of the epithelium, and the superficial and intermediate layers of the lamina propria, which have no contractile properties and are generally more pliable than the body layer (Hirano, 1974). We will show that increasing vocal fold body stiffness gradually reduces the vibration amplitude of the body layer and restricts vocal fold motion to the medial surface, which leads to increased sound production efficiency. At certain conditions, a slight change in the vocal fold stiffness or vocal fold geometry may cause phonation onset to occur at a different eigenmode, leading to a sudden change in phonation onset frequency and vocal fold vibration characteristics.

II. MODEL DESCRIPTION

Figure 1 shows a sketch of the two-dimensional continuum vocal fold model. For simplicity, a left-right symmetry in the flow field and vocal fold vibration about the glottal

channel centerline was imposed, and only half the system was considered. The vocal fold consisted of two linear plane-strain elastic layers, including a body layer and a cover layer, each with distinct Young's modulus E, Poisson's ratio ν, density ρ, and structure loss factor σ. The major geometric parameters of the vocal fold model include the thickness of the medial surface T, the depths of the body (D_b) and cover (D_c) layers at the center of the medial surface, the divergence angle of the medial surface α, and the minimum glottal half-width at rest g_0. The vocal folds were coupled to a one-dimensional potential flow driven by a given constant flow rate Q at the glottal entrance. The flow was assumed to separate from the glottal walls at a location downstream of the minimum glottal constriction whose width was 1.2 times the minimum glottal width (Lous *et al.*, 1998). A zero pressure recovery was assumed downstream of the flow separation point. As no vocal tract was considered, the flow pressure at the flow separation point was set to zero. Note that the flow separation model of this study gives only a rough estimation of the actual flow separation location (Decker and Thomson, 2007). The sensitivity of phonation onset to variations of the flow separation location was investigated in a separate study (Zhang, 2008).

A nondimensional formulation of system equations was used. The vocal fold thickness $\bar{T}$, the cover layer density $\bar{\rho}_c$, and the wave velocity of the vocal fold cover layer $\sqrt{\bar{E}_c/\bar{\rho}_c}$ were used for the reference scales of length, density, and velocity, respectively. The nondimensional variables are defined as follows:

$$
\begin{aligned}
&T = 1, \quad E_c = 1, \quad \rho_c = 1, \\
&D_b = \bar{D}_b/\bar{T}, \quad D_c = \bar{D}_c/\bar{T}, \quad g_0 = \bar{g}_0/\bar{T}, \\
&E_b = \bar{E}_b/\bar{E}_c, \quad \rho_b = \bar{\rho}_b/\bar{\rho}_c, \quad \rho_f = \bar{\rho}_f/\bar{\rho}_c, \\
&P_s = \bar{P}_s/\bar{E}_c, \quad U_j = \bar{U}_j/\sqrt{\bar{E}_c/\bar{\rho}_c}, \quad f = \bar{f}/(\sqrt{\bar{E}_c/\bar{\rho}_c}/\bar{T}),
\end{aligned}
\tag{1}
$$

where ρ_f is the density of air, U_j is the mean jet velocity, P_s is the mean subglottal pressure, and f is the phonation frequency. The subcripts b and c denote the properties of the body and cover layers, respectively. Symbols without overbars denote nondimensional variables. The physical values can be recovered by multiplying the nondimensional values by the corresponding reference scales. Note that the nondimensional body stiffness corresponds to the body-cover stiffness ratio in a dimensional format. Therefore, although the results below are discussed as a function of the nondimensional body stiffness, they could also be interpreted as a function of the body-cover stiffness ratio.

For a given flow rate Q at the glottal inlet, the analysis includes two steps. In the first step, a steady-state problem of the coupled fluid-structure system is solved to obtain the deformed vocal fold geometry and the mean flow pressure distribution along the glottal channel. In the second step, a linear stability analysis as in Zhang *et al.* (2007) is performed on the deformed state of the coupled system, and the eigenvalues and the eigenvectors of the coupled system are calculated. To differentiate from the natural modes of the vocal fold structure, these eigenmodes are referred to as fluid–structure interaction (FSI) eigenmodes. The deformed state is linearly stable if all FSI eigenvalues of the coupled system have negative growth rates. Otherwise the coupled system is linearly unstable for the given flow rate. This two-step procedure is repeated for a range of subglottal flow rates, and the phonation threshold pressure would then be the corresponding subglottal pressure at which the growth rate of one FSI eigenvalue first becomes positive.

The steady-state problem of the coupled system was solved as follows. For a given subglottal flow rate Q and a zero pressure at the point of flow separation, the mean flow pressure $P(z)$ along the vocal fold surface at locations upstream of the flow separation point was given by

$$
P(z) = \frac{1}{2}\rho_f \frac{Q^2}{H_j^2}\left(1 - \frac{H_j^2}{H^2}\right), \tag{2}
$$

where H is the glottal width along the flow direction and H_j is the glottal width at the point of flow separation. The flow pressure downstream of the flow separation point was set to zero. Note that, as mentioned above, H_j is a function of the minimum glottal half-width g and the vocal fold geometry, which are again functions of the flow pressure distribution and therefore H_j. An iterative approach was used to solve the steady-state problem. The flow pressure distribution was calculated for an initial value of H_j. The structural response under this flow pressure was then solved using the commercial finite-element-modeling package COMSOL. This deformed vocal fold surface geometry was then used to update H_j. This procedure was repeated until a satisfactory convergence in the vocal fold deformation, the flow separation point, and the minimum glottal width was reached.

The procedure of the linear stability analysis was described in detail in Zhang *et al.* (2007) and is only briefly summarized here. Readers are referred to the original paper for a detailed derivation of system equations. Governing equations for vocal fold dynamics were derived from Lagrange's equations:

$$
(M - Q_2)\ddot{q} + (C - Q_1)\dot{q} + (K - Q_0)q = 0, \tag{3}
$$

where q is the generalized coordinate vector. The three matrices M, C, and K represent the mass, damping, and stiffness matrices of the vocal fold structure, respectively. A proportional structural damping was assumed for the vocal fold material so that the structural mass and damping matrices were related by $C = \sigma\omega M$, where σ is the constant structural loss factor and ω is the angular frequency. The term $Q_2\ddot{q} + Q_1\dot{q} + Q_0 q$ in Eq. (3) is the generalized force associated with the fluctuating flow pressure along the vocal fold surface as induced by vocal fold vibration. The fluctuating flow pressure was obtained as follows. First, Bernoulli's equation and continuity equation of airflow were linearized (Zhang *et al.*, 2007) around the mean deformed state of the coupled airflow–vocal fold system, which was obtained from solving the corresponding steady-state problem. The boundary conditions for the linearized flow equations included a zero fluctuating flow velocity at the glottal entrance and a zero fluctuating pressure at the flow separation location. The

fluctuating pressure was then calculated by solving the linearized flow equations under these boundary conditions. The fluctuating flow pressure consists of three components, including a flow-induced stiffness term (proportional to vocal fold displacement and represented by matrix Q_0), a flow-induced damping term (proportional to vocal fold velocity and represented by matrix Q_1), and a flow-induced mass term (proportional to vocal fold acceleration and represented by matrix Q_2). All three matrices are functions of the jet velocity U_j, which was calculated in the steady-state problem using the imposed subglottal flow rate and the deformed vocal fold geometry. Assuming $q_0=q_0e^{st}$, Eq. (3) was solved for the FSI eigenvalues s and FSI eigenvectors q_0 for a given flow rate Q.

In Zhang *et al.* (2007), Eq. (3) was solved using the Ritz method, in which polynomial functions were used as basis functions to approximate the vocal fold displacement field. In the present study, the natural eigenmodes of the vocal fold structure were used as basis functions. The vocal fold displacement field $w=[w_x,w_z]$ was approximated as

$$w_x \approx \sum_{k=1}^{N} q_k W_{x,k}, \quad w_z \approx \sum_{k=1}^{N} q_k W_{z,k}, \tag{4}$$

where $[W_{x,k},W_{z,k}]$ was the displacement field associated with the kth natural eigenmode of the vocal fold structure, q_k is the kth generalized coordinate, and N is the number of natural modes used in the approximation. In this study, $N=10$ was used. As these natural modes contain information of the vocal fold dynamics, generally only a few natural modes are necessary to obtain solutions of reasonable accuracy, which greatly reduces computational costs. The use of natural modes as basis functions also allows vocal folds of arbitrary geometry to be analyzed with no appreciable increase in computation time. In this study, the natural modes were normalized so that the total kinetic energy of the vocal fold structure was equal to 1. The generalized force matrices (Q_0,Q_1,Q_2) were numerically evaluated using the normalized basis functions.

Note that a positive minimum glottal half-width was assumed, and only the behavior of the system around the phonation onset was investigated in this study. Vocal fold collision and its influence on phonation onset were therefore not considered.

III. RESULTS

This section is organized as follows. In Sec. III A, the influence of the body-cover stiffness ratio on phonation onset is first discussed for a straight glottal channel. The influence on phonation threshold pressure, phonation onset frequency, vocal fold vibration, and sound production efficiency is discussed. Results for convergent and divergent glottal channels are discussed in Sec. III B. Competition of coexisting instabilities for dominance at phonation onset was observed for convergent glottal channels. This is discussed in Sec. III C. Due to this competition between coexisting instabilities, a slight change in either the vocal fold stiffness or the glottal divergence angle may lead to a sudden change in phonation onset frequency, vocal fold vibration, and sound production efficiency. The influence of other model parameters is then briefly discussed in Sec. III D. Finally, the implications of the results of this study on pitch control are discussed in Sec. III E.

For the results presented below, unless otherwise stated (e.g., in Sec. III D), the following values of the model parameters were used:

$$D_b=2, \quad D_c=0.333, \quad g_0=0.03, \quad \rho_b=1, \quad \rho_f=0.001\ 17, \quad \sigma=0.4. \tag{5}$$

For a medial-surface thickness of 3 mm, Eq. (5) gives a vocal fold body depth of 6 mm, a cover depth of 1 mm, and a 0.09 mm minimum glottal half-width at rest, which roughly correspond to the nominal vocal fold geometry in Titze and Talkin (1979).

A. Influence of body-cover stiffness ratio: General trends

There have been little data available on the body-cover stiffness ratio during phonation in the literature. Although the stiffness of the nonactivated body layer may be comparable to that of the cover layer, the body-cover stiffness ratio may vary in a wide range under different activities of the TA and CT muscles (Hirano, 1974). The body-cover stiffness ratio is also expected to vary in an even larger range in pathological phonation. A large range of body-cover stiffness ratio has been used in previous modeling studies. For example, in Berry *et al.* (1994), the body-cover stiffness ratio was varied from 2 to about 13. In Titze and Talkin (1979) a ratio of 10:4:2 was used for the passive stiffnesses of the muscle, ligament, and mucosa. In the three-mass body-cover model of Story and Titze (1995), the stiffness ratio of the body and cover masses was varied in the approximate range of 4–200. In this study, similar to Titze and Talkin (1979) and Story and Titze (1995), the body-cover stiffness ratio (or body stiffness E_b) was varied in a wide range (from 1 to 100) to encompass the possible physiological range that may occur in normal and pathological phonation. In normal phonation, an increasing body-cover stiffness ratio may be realized by increased contraction of the TA muscle or reduced CT muscle contraction or a combination of both.

Figure 2 shows the phonation threshold pressure, P_{th}, the phonation onset frequency, F_0, and the prephonatory minimum glottal half-width, g, as functions of the body stiffness E_b. For a straight glottal channel $\alpha=0$ (circles in Fig. 2), both the phonation threshold pressure and the phonation onset frequency increased with increasing body stiffness E_b. This is consistent with the predictions in Zhang *et al.* (2007), which stated that both the phonation threshold pressure and the phonation onset frequency increase with the natural frequency of the vocal fold structure. In this case, the increase in the natural frequency of the vocal fold structure was caused by an increase in the body stiffness. This also explains why both P_{th} and F_0 gradually approached a plateau at large values of body stiffness. For large values of body stiffness, the natural frequencies of the two-layer vocal fold structure were primarily determined by the cover stiffness

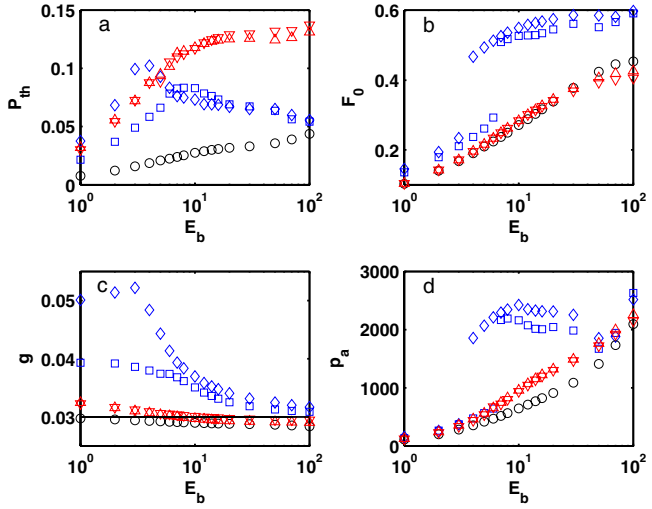


FIG. 2. (Color online) (a) Phonation threshold pressure P_{th}, (b) phonation onset frequency F_0, (c) prephonatory minimum glottal half-width g, and (d) amplitude of radiated acoustic pressure p_a as functions of body-cover stiffness ratio E_b/E_c for five different glottal channel divergence angles: ◇: −10, □: −5, ○: 0, ▽: 5, and △: 10. Also shown in (c) is the minimum glottal half-width at rest (solid line). Model parameters are given in Eq. (5).

and only increased slightly with the body stiffness. Figure 2(c) shows that the prephonatory minimum glottal half-width decreased with increasing body stiffness. A comparison between the initial and deformed vocal fold geometries shows that at small body stiffnesses, the vocal fold not only moved upwards (superiorly) but also bulged out slightly toward the glottal midline to close the glottis. As the body stiffened, the upward movement was gradually restricted and the vocal fold was forced to move more medially, leading to an even smaller prephonatory minimum glottal opening. This bulging effect was highly dependent on the vocal fold geometry (Sec. III B) and was consistently small in this case for a straight glottal channel.

The vocal fold vibration pattern at onset was calculated by substituting the corresponding eigenvector of Eq. (3) to Eq. (4). When properly scaled, the vocal fold vibration field was superimposed onto the deformed vocal fold geometry, and the corresponding vocal fold motion during one cycle can be visualized. Figures 3 and 4 show such vocal fold motion during one oscillation cycle for $E_b=1$ and $E_b=100$, respectively. The case of $E_b=1$ roughly corresponds to the first and fourth cases investigated by Story and Titze (1995), which Hirano (1974) claimed to occur when either the TA muscle is not active and the CT muscle contracts powerfully, or both the TA and CT contractions are weak. The case of $E_b=100$ roughly corresponds to the second Hirano condition and the second case in Story and Titze (1995), which may occur when the TA muscle contracts much more powerfully than the CT muscle (Hirano, 1974). Comparing Figs. 3 and 4 shows that a major difference between these two cases is that for $E_b=1$, both the body and the cover were equally involved in the vibration, whereas the body barely moved for the case of $E_b=100$. This is consistent with the description of Hirano (1974) and the observations in Story and Titze (1995) that the body layer was gradually less involved in vibration with increasing body-cover stiffness ratio. For a stiff body (Fig.

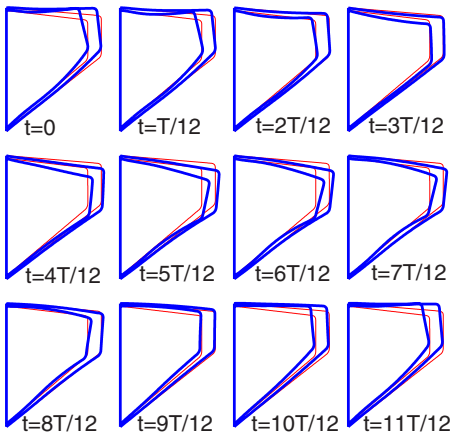


FIG. 3. (Color online) The vocal fold geometry during one oscillation cycle. $E_b/E_c=1$, $\alpha=0$ and other model parameters are given in Eq. (5). The first frame in time is shown in the leftmost plot in the first row, and the last frame is shown in the rightmost plot in the last row. The thin lines correspond to the mean deformed vocal fold geometry at onset as obtained from solving the steady-state problem.

4), the motion was restricted to the medial surface of the cover layer. The wavelength of the vibration along the medial surface was also much smaller in Fig. 4 than in Fig. 3, leading to a more wavelike motion in the case of $E_b=100$.

Figure 5 shows the vocal fold motion along the vocal fold surface as a function of superior-inferior location and time. With increasing body stiffness, regions of large-amplitude motion in both the medial-lateral and superior-inferior components were gradually reduced to the superior portion of the medial surface. In the case of $E_b=1$, the entire medial surface (spans from $z=1.8$ to 2.8 approximately) almost vibrated at the same phase. This is similar to the first and fourth cases of Story and Titze (1995) in which the upper and lower masses moved approximately in-phase. In Fig. 5(a), there is another region of in-phase motion along the inferior surface. However, the vibration amplitude in the medial-lateral direction was much weaker compared to the in-phase region along the medial surface. For the case of

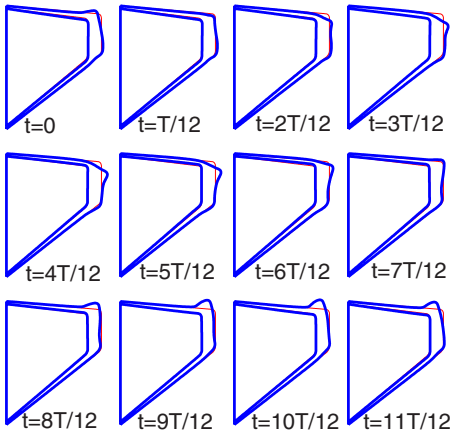


FIG. 4. (Color online) The vocal fold geometry during one oscillation cycle. $E_b/E_c=100$, $\alpha=0$, and other model parameters are given in Eq. (5). The first frame in time is shown in the leftmost plot in the first row, and the last frame is shown in the rightmost plot in the last row. The thin lines correspond to the mean deformed vocal fold geometry at onset as obtained from solving the steady-state problem.

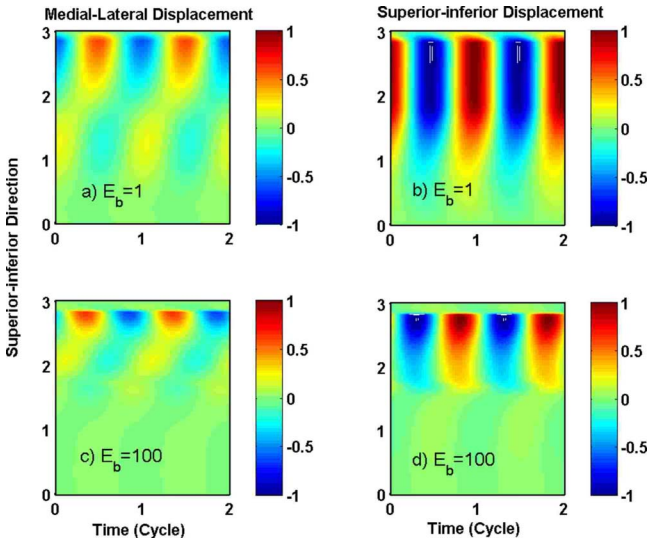


FIG. 5. (Color online) The spatiotemporal plot of the medial-lateral (left) and superior-inferior (right) components of the vocal fold surface displacement for $E_b/E_c=1$ (top) and $E_b/E_c=100$ (bottom). $\alpha=0$ and other model parameters are given in Eq. (5). For each case, the two components were normalized by the maximum value of the two components along the surface.

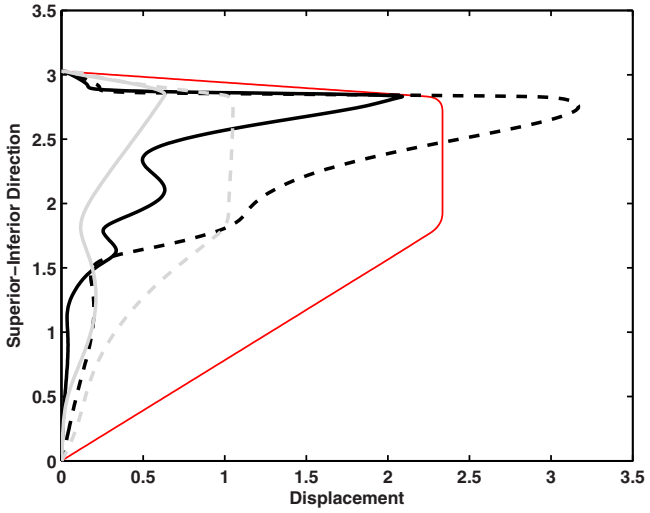


FIG. 6. (Color online) The amplitudes of the medial-lateral (thick solid lines) and superior-inferior components (dashed lines) of the vocal fold surface displacement along the flow direction for $E_b/E_c=1$ (gray lines) and $E_b/E_c=100$ (dark lines). $\alpha=0$ and other model parameters are given in Eq. (5). For each case, the FSI eigenmode was normalized so that the vibrational energy was 1. The thin solid line denotes the vocal fold surface.

$E_b=100$, the in-phase region along the medial surface was reduced in size and restricted to the superior part of the medial surface, whereas no in-phase region was observed along the inferior surface. Consequently, a phase difference can be clearly observed between the superior and inferior portions of the medial surface, with the inferior portion of the medial surface vibrating at a much smaller amplitude. This feature was not observed in any case studied by Story and Titze (1995), in which the amplitude of the lower cover mass was at least comparable to that of the upper cover mass.

Note that in both cases of Fig. 5, the vocal fold exhibited considerable superior-inferior motion. Although it was often neglected in lumped-mass models, this vertical motion was also observed in other continuum model simulations (e.g., Titze and Talkin, 1979; Berry *et al.*, 1994) and experiments (Dollinger *et al.*, 2005; Zhang *et al.*, 2006a).

The gradual confinement of large-amplitude motion to the medial surface may allow a better flow modulation and may therefore benefit sound production. Figure 6 shows the amplitudes of the medial-lateral and superior-inferior components of the vocal fold surface displacement associated with the normalized FSI eigenmode at onset for the two cases $E_b=1$ (gray lines) and $E_b=100$ (dark thick lines). For comparison between cases of different body stiffnesses, the FSI eigenvector of the coupled system for each case was normalized so that the average kinetic energy of the entire vocal fold structure over one cycle was 1. Consistent with Figs. 4 and 5, the vocal fold vibration amplitude was significantly reduced along the inferior and superior surfaces in the case of $E_b=100$ as compared to the case of $E_b=1$. For the same vibrational energy of the vocal fold structure (due to eigenvector normalization), the restriction of the vibration energy to the cover layer and the medial surface led to a much larger vibration amplitude at the superior portion of the medial surface in the case of $E_b=100$ than in the case of $E_b=1$. Note that vocal fold motion in this region (the superior portion of the medial surface) is the most effective in terms of flow modulation and affecting the intraglottal pressure distribution and is therefore the most efficient in terms of sound production.

To quantify the effect of different vocal fold vibration patterns on voice production, a sound production efficiency was defined as the amplitude of the acoustic pressure p_a radiated into an infinitely long uniform vocal tract due to vocal fold vibration associated with the corresponding *normalized* FSI eigenvector. The acoustic pressure p_a was calculated as (Zhang *et al.*, 2002)

$$p_a = -\frac{1}{2H_{\text{in}}}\int_S p n_z \cdot dS - \frac{1}{2H_{\text{in}}}\int_S \rho_f c\dot{w} \cdot n dS, \tag{6}$$

where p is the fluctuating flow pressure along the vocal fold surface S induced by the vocal fold motion w, c is the speed of sound, and n_z is the vertical component (in the flow direction) of the normal vector of the vocal fold surface pointing into the vocal fold. Note that Eq. (6) was obtained by assuming that the dimensions of the vocal folds were much smaller than the acoustic wavelength of interest so that there was no time delay between contributions to the far field sound from vocal fold motion at different spatial locations along the vocal fold surface. The terms in the right hand side of Eq. (6) include contributions of both the dipole source (due to the fluctuating transglottal pressure) and the monopole source (due to instantaneous volume change of the vocal folds) (Zhang *et al.*, 2002). The flow pressure p was calculated using the normalized FSI eigenvector of the coupled system, as described in Zhang *et al.* (2007). The acoustic pressure p_a thus calculated represented the acoustic pressure produced by vocal fold vibration of unit kinetic energy and thus quantified the voice production efficiency of the corresponding vocal fold vibration pattern. The amplitude of the calculated p_a was shown as a function of the body stiffness in Fig. 2(d). As expected, the sound production efficiency increased with

increasing body stiffness, as the vocal fold motion was gradually restricted to the cover layer and the medial surface. For all cases of this study, the contribution of the monopole source to the total acoustic pressure was negligible compared to that of the dipole source.

Although no results were shown for other values of E_b, as the body stiffness increased, the vocal fold vibration pattern evolved continuously from a vibration pattern similar to that in the case of $E_b=1$ toward one similar to the case of $E_b=100$. With increasing body stiffness, the vocal fold motion was gradually restricted to the cover layer, leading to increased sound production efficiency. As the region of in-phase large-amplitude motion was gradually reduced in size and moved to the superior portion of the medial surface, the phase difference between the superior and inferior potions of the medial surface increased with increasing body stiffness, gradually leading to a wavelike motion along the vocal fold surface.

B. Effects of glottal channel geometry

Figure 2 also shows the results obtained for nonstraight glottal channel geometries. Other model parameters were the same as given in Eq. (5). For divergent glottal channels (triangle symbols in Fig. 2) and convergent glottal channels with small body stiffnesses (diamond and square symbols in Fig. 2), similar observations as discussed in Sec. III A can be made on the variation in phonation threshold pressure, phonation onset frequency, and radiated acoustic pressure with body stiffness. In general, the phonation threshold pressures for both divergent and convergent channels were higher than those for the straight glottal channel. This difference could be due to different intraglottal pressure distributions associated with different glottal channels. Another possible explanation is that for a given minimum glottal width, the average glottal width was larger for divergent and convergent channels than the straight glottal channel, which weakened the fluid-structure coupling strength (Sec. III D). The sound production efficiency was slightly higher for divergent glottal channels and convergent channels with small body stiffnesses than for the straight glottal channel.

At small values of the body stiffness, the divergent glottis was slightly pushed open by the glottal flow. For convergent glottal channels, the glottis-opening effect was even more pronounced at small body stiffnesses, and the prephonatory glottal opening was consistently larger than the glottal opening at rest for all values of body stiffness investigated. This is probably due to the relatively high intraglottal pressure associated with convergent glottal channels. For both geometries with increasing body stiffness, the prephonatory glottal opening gradually decreased. For divergent geometries and large enough body stiffnesses ($E_b>10$), the prephonatory glottal opening was actually smaller than the glottal opening at rest. This is consistent with the observation for the straight glottis, as discussed in Sec. III A: stiffening the body restricted the vertical motion and caused the vocal fold to bulge out more in the medial direction, therefore reducing the prephonatory minimum glottal opening. This is also consistent with clinical observations that vocal folds with increased cover stiffness (e.g., for scarred vocal folds) are often blown apart by the airflow even if the glottis was closed completely at rest. In this case, surgical medialization would not improve much the voice, and other measures are required to reduce the cover stiffness (Isshiki, 1998).

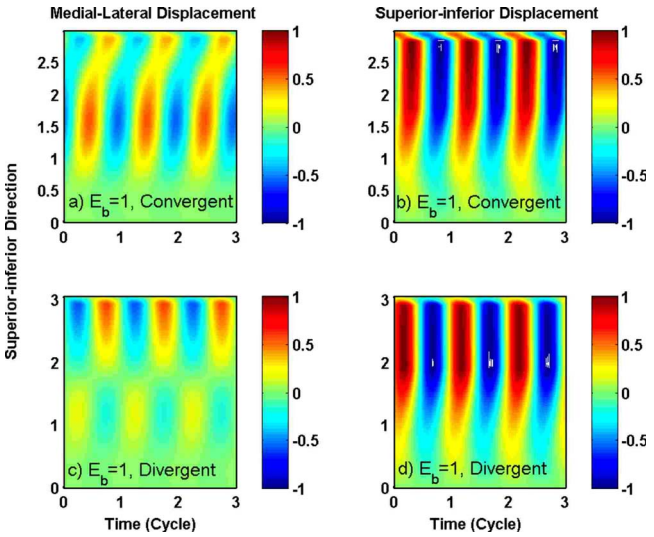


FIG. 7. (Color online) The spatiotemporal plot of the medial-lateral (left) and superior-inferior (right) components of the vocal fold surface displacement for $\alpha=-5$ (top) and $\alpha=5$ (bottom). $E_b/E_c=1$ and other model parameters are given in Eq. (5). For each case, the two components were normalized by the maximum value of the two components along the surface.

Figure 7 shows the vocal fold vibration pattern for a convergent ($\alpha=-5$) and a divergent ($\alpha=5$) glottis for $E_b=1$. The vocal fold vibration for a convergent glottis [Figs. 7(a) and 7(b)] exhibited two regions of in-phase medial-lateral vibration with comparable amplitudes, which is in contrast to only one region of dominant medial-lateral motion for a straight glottis [Fig. 5(a)]. In Fig. 7(a), one region is located in the small area around the superior edge of the medial surface, while the other spanned a much larger area in the superior-inferior direction. The vocal fold vibration within each region was almost in-phase, but the two regions vibrated slightly out-of-phase with each other. Note that the existence of two regions of in-phase motion along the medial surface is reminiscent of two lumped masses vibrating slightly out-of-phase, as described by the two-mass model of Ishizaka and Flanagan (1972).

The vocal fold vibration for a divergent glottis [Figs. 7(c) and 7(d)] was qualitatively similar to that of the straight glottis, with dominant medial-lateral vibration along the medial surface and a much weaker motion along the inferior surface. With increasing body stiffness, the region of dominant motion was gradually reduced to the superior portion of the medial surface, as in the case of a straight glottis [Figs. 5(c) and 5(d)].

C. Competition of coexisting instabilities

For convergent glottal channels with large body stiffnesses, the phonation onset pattern as a function of body stiffness was quite different from that for straight and divergent glottal channels. For small body stiffnesses, the phonation threshold pressure and the phonation onset frequency

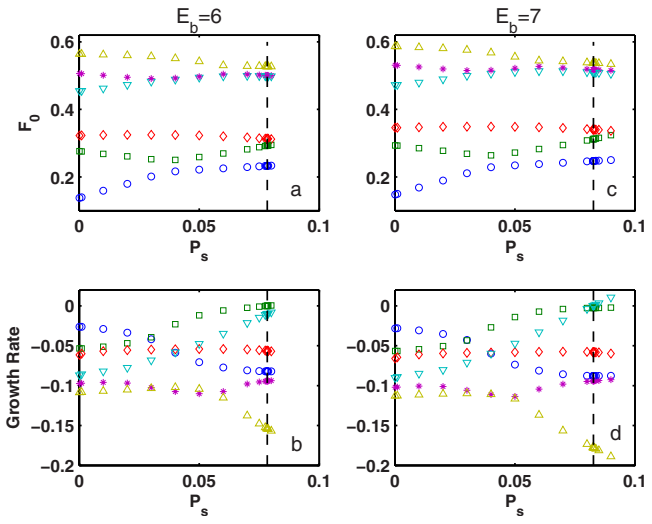


FIG. 8. (Color online) The frequencies (top) and growth rates (bottom) of the first six eigenvalues (○: first; □: second; ◇: third; ▽: fourth; ∗: fifth; △: sixth) of the coupled fluid-structure system as a function of the subglottal pressure for $E_b/E_c=6$ (left) and $E_b/E_c=7$ (right). $\alpha=-5$ and other model parameters are given in Eq. (5). The vertical line indicates the point of onset. Despite a slight change in the body-cover stiffness ratio E_b/E_c, onset occurred as a different eigenmode was destabilized.

gradually increased with increasing body stiffness, as in the cases of straight and divergent glottal channels. However, for the case of a -5 divergence angle (square symbols in Fig. 2), as body stiffness increased from 6 to 7 in Fig. 2, the phonation onset frequency abruptly increased to a much higher value, whereas the phonation threshold pressure only increased slightly. As the body stiffness further increased, the phonation onset frequency increased gradually, but the phonation threshold pressure started to decrease.

The abrupt increase in phonation onset frequency in response to a slight change in body stiffness was due to the competition of two coexisting instabilities for dominance. Figure 8 shows the frequencies and growth rates of the first six eigenvalues of the coupled system for the two cases before ($E_b=6$) and after ($E_b=7$) the abrupt increase in F_0. As shown in Fig. 8, there were two groups of eigenmodes of strong interaction: the first group included eigenmodes 1 and 2, and the second group included eigenmodes 4, 5, and possibly 6. At close to onset, there were two eigenmodes (eigenmodes 2 and 4) that had growth rate close to zero and increasing, but with quite different frequencies. For small values of body stiffness ($E_b<7$), the interaction between the first two eigenmodes was strong, and onset occurred as the second eigenmode first reached a zero growth rate. With increasing body stiffness, this interaction between eigenmodes in the first group gradually weakened [as indicated by the increasing phonation threshold pressure for $E_b<7$ in Fig. 2(a)], whereas the interaction between eigenmodes of the second group became increasingly stronger [as indicated by the decreasing phonation threshold pressure for $E_b>7$ in Fig. 2(a)]. At a certain threshold ($E_b=7$), the interaction within the second group was so strong that the fourth eigenmode reached a zero growth rate before the second eigenmode did, and phonation onset occurred at the fourth eigenmode instead of the second eigenmode, leading to an abrupt increase in phonation onset frequency.

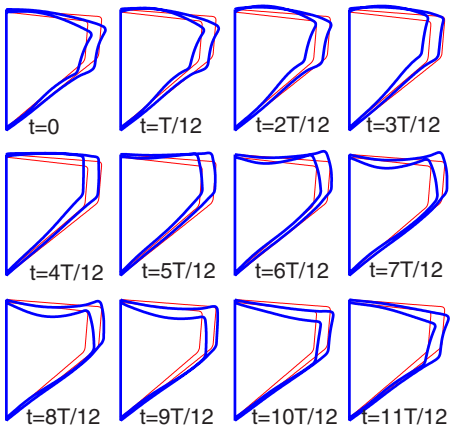


FIG. 9. (Color online) The vocal fold geometry during one oscillation cycle. $E_b/E_c=6$, $\alpha=-5$ and other model parameters are given in Eq. (5). The first frame in time is shown in the leftmost plot in the first row, and the last frame is shown in the rightmost plot in the last row. The thin lines correspond to the mean deformed vocal fold geometry at onset as obtained from solving the steady-state problem.

Figures 9 and 10 show the vocal fold motion during one oscillation cycle for the two cases $E_b=6$ and $E_b=7$, respectively. When vibrating at a higher-order eigenmode, the vocal fold vibration along the surface had a smaller wavelength in the case of $E_b=7$, as compared to the case of $E_b=6$. This led to a more wavelike motion in the case of $E_b=7$. The corresponding spatiotemporal plots are shown in Fig. 11. Compared to the case of $E_b=6$, the vocal fold vibration in the case of $E_b=7$ exhibited a large medial-lateral motion along the superior edge of the medial surface and reduced superior-inferior motion along the vocal fold surface (except the superior part of the medial surface, where the upheaval-like motion was observed in Fig. 10). Figure 12 compares the amplitudes of the medial-lateral and superior-inferior displacement of the normalized FSI eigenmode along the vocal fold surface at onset for the two cases $E_b=6$ (gray lines) and $E_b=7$ (dark thick lines). The motion was more restricted to the superior portion of the medial surface in the case of $E_b=7$. Note that the vocal fold vibration for $E_b=7$ was similar

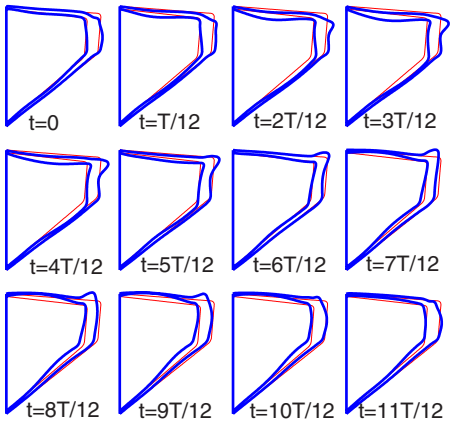


FIG. 10. (Color online) The vocal fold geometry during one oscillation cycle. $E_b/E_c=7$, $\alpha=-5$, and other model parameters are given in Eq. (5). The first frame in time is shown in the leftmost plot in the first row, and the last frame is shown in the rightmost plot in the last row. The thin lines correspond to the mean deformed vocal fold geometry at onset as obtained from solving the steady-state problem.

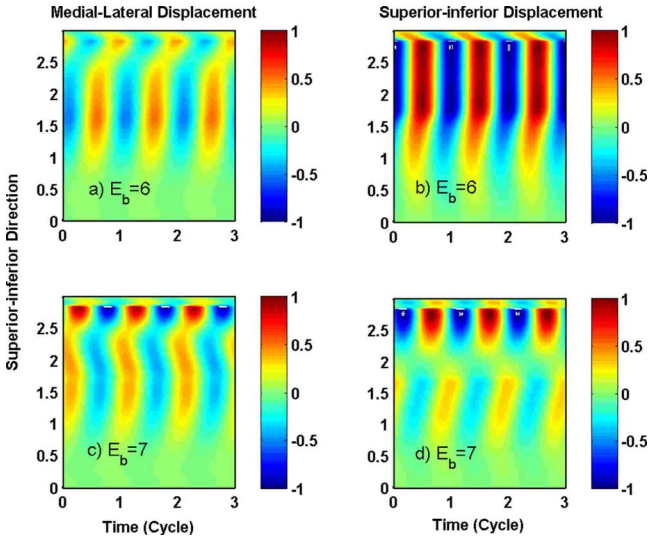


FIG. 11. (Color online) The spatiotemporal plot of the medial-lateral (left) and superior-inferior (right) components of the vocal fold surface displacement for $E_b/E_c=6$ (top) and $E_b/E_c=7$ (bottom). $\alpha=-5$ and other model parameters are given in Eq. (5). For each case, the two components were normalized by the maximum value of the two components along the surface.

to the case of $E_b=100$, in terms of the surface vibration pattern (e.g., small wavelength, wavelike motion, and restriction of large motion to the superior portion of the medial surface). However, these common features were achieved in different ways: one was induced by vibrating at a higher-order mode, while the other was induced by a stiff body.

The abrupt increase in phonation onset frequency was accompanied by a boost in sound production efficiency, as shown in Fig. 2(c). In fact, Fig. 2(c) shows that vibrating at a higher-order eigenmode was always more efficient in terms of sound production than vibrating at a lower-order mode. Although the motion was more uniformly spread over the vocal fold surface in the case of $E_b=7$, excitation of the

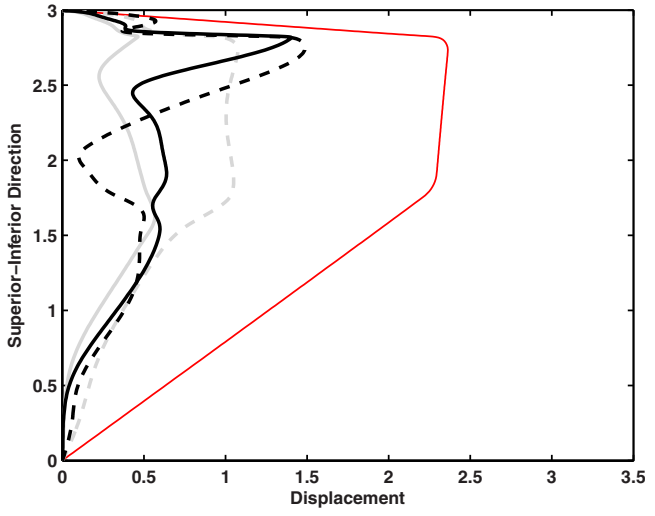


FIG. 12. (Color online) The amplitudes of the medial-lateral (thick solid lines) and superior-inferior components (dashed lines) of the vocal fold surface displacement along the flow direction for $E_b/E_c=6$ (gray lines) and $E_b/E_c=7$ (dark lines). $\alpha=-5$ and other model parameters are given in Eq. (5). For each case, the FSI eigenmode was normalized so that the vibrational energy was 1. The thin solid line denotes the vocal fold surface.

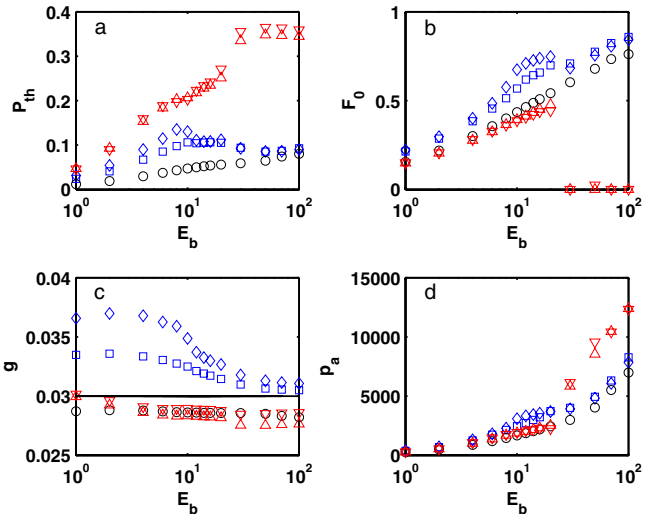


FIG. 13. (Color online) (a) Phonation threshold pressure P_{th}, (b) phonation onset frequency F_0, (c) prephonatory minimum glottal half-width g, and (d) amplitude of radiated acoustic pressure p_a as functions of body-cover stiffness ratio E_b/E_c for five different glottal channel divergence angles: ◇: −10, □: −5, ○: 0, ▽: 5, and △: 10. Also shown in (c) is the minimum glottal half-width at rest (solid line). $T=1$, $D_b=6D_c=1.2$, and other parameters are given in Eq. (5).

higher-order mode reduced the superior-inferior motion (compare $E_b=7$ in Fig. 12 to both $E_b=6$ in Fig. 12 and $E_b=100$ in Fig. 6), which allowed more energy to be spent on the medial-lateral motion at the superior portion of the medial surface and led to higher sound production efficiency.

D. Effects of other model parameters

As expected, the eigenmode synchronization pattern is highly dependent on the vocal fold geometry. As an example, Fig. 13 shows the phonation onset characteristics when the vocal fold depths were decreased [$D_b=1.2$, $D_c=0.2$, other parameters the same as in Eq. (5)], which is more similar to the conditions used in previous studies (Zhang *et al.*, 2007; Zhang, 2008). For the straight glottal channel, onset occurred due to the interaction of the first and second eigenmodes. For convergent glottal channels, onset occurred due to the interaction of the second and third eigenmodes for small body stiffnesses ($E_b<8$) but changed to the interaction of the first and second eigenmodes for large body stiffnesses ($E_b>20$). However, as the second eigenmode was involved in all cases, no abrupt change in phonation onset frequency was observed in Fig. 13. For divergent glottal channels, the synchronization pattern changed from interaction (coupled-mode flutter) between the first and second eigenmodes for small body stiffnesses ($E_b<20$) to static divergence (zero-frequency instability) at larger body stiffnesses ($E_b>20$), which is consistent with Zhang (2008). Note that for this geometry, the prephonatory minimum glottal half-width was consistently smaller than that in Fig. 2, indicating an increasing difficulty for the vocal fold to be deformed laterally with decreasing vocal fold depth (or aspect ratio).

Figures 14(a) and 14(b) show the phonation threshold pressure and phonation onset frequency as a function of the structural loss factor for $E_b=10$ and other parameter values in Eq. (5). Figures 14(c) and 14(d) show the phonation

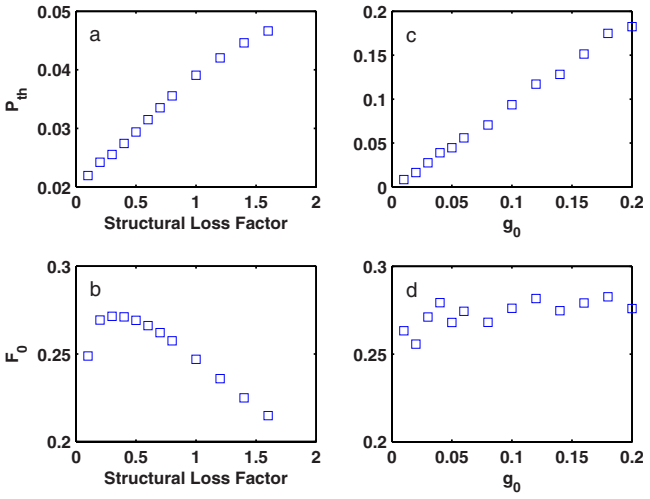


FIG. 14. (Color online) The left column shows the (a) phonation threshold pressure and (b) phonation onset frequency as functions of structural loss factor (E_b=10, g_0=0.03). The right column shows the (c) phonation threshold pressure and (d) phonation onset frequency as functions of minimum glottal half-width at rest (E_b=10, σ=0.4).

threshold pressure and phonation onset frequency as a function of the minimum glottal half-width at rest for E_b=10 and other parameter values in Eq. (5). Figure 14 shows an approximately linear dependence of the phonation threshold pressure on both the structural loss factor and the glottal half-width at rest. Although the values of model parameters do not exactly match, this almost-linear dependence is consistent with previous experimental observations (Titze *et al.*, 1995; Chan *et al.*, 1997). As there was no sudden change in the eigenmode synchronization pattern, the phonation frequency changed continuously in the case of Fig. 14. It decreased with increasing structural loss factor and slightly increased with increasing glottal half-width.

E. Implications on pitch control

The implications of the results of this study on pitch control are better illustrated using dimensional variables. Figure 15(a) shows four phonation onset frequency contours in the $\bar{E}_c - \bar{E}_b$ space corresponding to 100, 150, 200, and 250 Hz, respectively. The figure was generated from Fig. 2(b) for a straight glottal channel, a medial surface thickness T=3 mm, and a vocal fold density of 1000 kg/m^3. Figure 15 shows that the effectiveness of varying body and cover stiffness as a pitch control mechanism depends on the body-cover stiffness ratio. For very large body-cover stiffness ratios, which correspond to a case with maximum TA contraction and minimum CT contraction, phonation frequency can be effectively controlled by varying the cover stiffness, whereas varying the body stiffness as a pitch control was much less effective. For very small body-cover stiffness ratios, which correspond to a case when both the TA and CT contractions are weak or when the TA muscle is not active and the CT muscle contracts powerfully, phonation frequency can be effectively controlled by varying the body stiffness, whereas varying cover stiffness had little influence

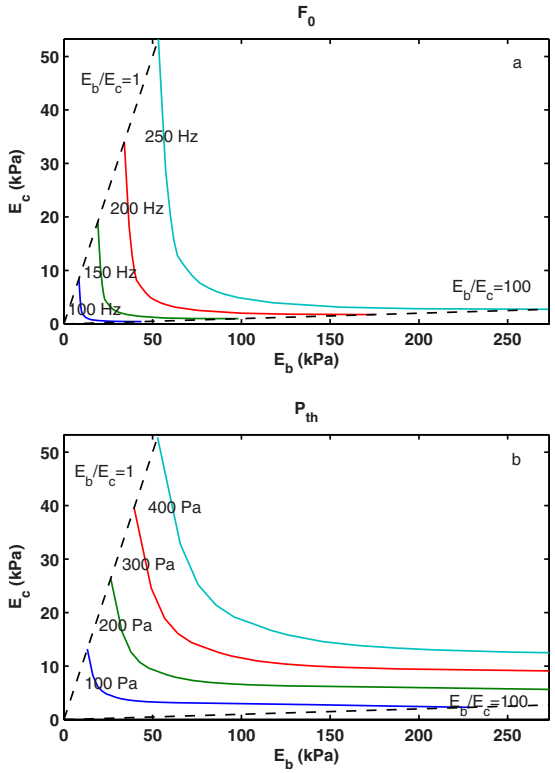


FIG. 15. (Color online) (a) Constant F_0 contours and (b) constant P_{th} contours in the $E_c - E_b$ space. T=3 mm and ρ_c=1000 kg/m^3. In (a), the four solid lines indicate F_0 of 100, 150, 200, and 250 Hz. In (b), the four solid lines indicate P_{th} of 100, 200, 300, and 400 Pa. The two dashed lines indicate constant body-cover stiffness ratio of E_b/E_c=1 and E_b/E_c=100. F_0 can be effectively controlled by changing the body stiffness at small body-cover stiffness ratios (E_b/E_c<10) and by changing the cover stiffness for large body-cover stiffness ratios (E_b/E_c>10).

on phonation frequency. For medium values of body-cover stiffness ratio, phonation frequency can be controlled by varying either the body or cover stiffness.

Figure 15(a) also shows that the same phonation onset frequency can be achieved by different combinations of body and cover stiffnesses. For example, an F_0 of 100 Hz can be achieved by $[\bar{E}_b, \bar{E}_c]$=[8.52, 8.52] kPa for E_b/E_c=1, $[\bar{E}_b, \bar{E}_c]$=[10.24, 2.05] kPa for E_b/E_c=5, or $[\bar{E}_b, \bar{E}_c]$=[43.7, 0.44] kPa for E_b/E_c=100. To produce the same F_0 of 100 Hz, the condition of E_b/E_c=5 required moderate values of both the body and cover stiffnesses. For body-cover stiffness ratios below or above 5, producing the same F_0 required a dramatic increase in either the cover stiffness (increased from 2.05 to 8.52 kPa) or the body stiffness (from 10.24 to 43.7 kPa). In human phonation, this dramatic increase would require strong contraction of either the CT or TA muscles in human phonation, which may be less desirable.

Figure 15(b) shows similar contour plots for phonation threshold pressure. As the phonation threshold pressure is directly related to the phonation onset frequency, a similar behavior can be also noted as to the effectiveness of varying body or cover stiffness as a control mechanism of phonation threshold pressure.

Note that in reality, the variations in body and cover stiffnesses are often accompanied by changes in the vocal

fold geometry. This effect needs to be taken into consideration using a proper muscular model (e.g., Titze and Hunter, 2007) in order to obtain a complete understanding of pitch control mechanisms in human phonation.

IV. DISCUSSION

The mucosal wave along the vocal fold surface has long been observed and considered an essential element of vocal fold vibration. It is generally assumed that mucosal wave propagation causes a time delay in the movement from bottom to top of the vocal folds (Titze, 1988) or a phase difference between the upper and lower masses in the two-mass model. However, the present study and a previous study (Zhang *et al.*, 2007) show that the phase difference of the vocal fold vibration along the vocal fold surface was the consequence of eigenmode synchronization at the same frequency but different phases. A wavelike motion appears only for vibrations of small wavelengths, which occurred in this study at large body-cover stiffness ratios (Sec. III A) or when the vocal fold vibrated at a higher-order mode (Sec. III C). In general, the vocal fold vibration does not necessarily exhibit a wavelike motion. More often, two regions of almost-in-phase vocal fold motion were observed along the vocal fold surface, and noticeable phase change only occurred in the transition region. There was no wave propagation (or rather infinitely fast wave motion) within each in-phase region. In other words, the presence of mucosal wave is not a necessary component of the self-sustained vocal fold vibration. However, the presence of mucosal wave may be desirable in phonation, for example, to achieve high sound production efficiency, as shown in this study.

This study shows that the concept of eigenmode and eigenmode synchronization may provide a theoretic framework toward a better understanding of the correspondence between biomechanical and geometric properties of the vocal folds and the resulting phonation characteristics. In this study, the vocal fold vibration was calculated as the weighted combination of the natural modes of the vocal fold structure, with the weights (the generalized coordinates q) determined by the fluid-structure interaction or the eigenmode synchronization process. With different combinations of weights, various types of vocal fold vibration can be generated, as demonstrated in this study. Therefore, further research on phonation onset can be pursued in two aspects. The first aspect focuses on the natural modes and how they would be affected by the changes in geometric and biomechanical properties of the vocal fold structure. Such study would yield valuable information on the characteristic vocal fold vibration patterns and the associated frequencies (e.g., Titze and Strong, 1975; Berry and Titze, 1996; Cook and Mongeau, 2007). For example, the restriction of motion toward the cover layer and the superior portion of the medial surface with increasing body stiffness can be explained by similar features in the first few natural modes of the vocal fold structure as the body stiffness increases. The second aspect aims to investigate eigenmode synchronization due to the fluid-structure interaction and to determine the weights used to calculate the final vocal fold vibration (e.g., Zhang *et al.*, 2007). Such studies would reveal which modes have a strong interaction and eventually synchronize to induce phonation onset and at what conditions bifurcations [e.g., the abrupt frequency change in Fig. 2(b)] in the behavior of the coupled system would occur.

This study shows that a slight change in body stiffness can cause an abrupt change in phonation onset frequency and vocal fold vibration pattern. Although this study focuses on phonation onset, it is reasonable to expect that a similar mechanism may also be present in finite-amplitude vibrations beyond onset and may play a role in register change. This mechanism requires that the coupled system have two or more coexisting instabilities so that, given appropriate changes in certain system parameters, the vocal fold vibration can switch from one self-oscillating state to another. For example, Tokuda *et al.* (2007) showed that a qualitative change in the vocal fold vibration and phonation frequency in a three-mass model was observed when two synchronizing eigenmodes switched from the first and second eigenmodes to the second and third eigenmodes. Note that these instabilities could be two near-field FSI instabilities (instabilities due to two pairs of synchronizing eigenmodes, as in the case of Fig. 8) but could also be one near-field FSI instability and one due to the coupling of the vocal fold vibration to sub- or supraglottal acoustics (Zhang *et al.*, 2006a, 2006b).

Although such abrupt change in vocal fold vibration pattern in Tokuda *et al.* (2007) and the present study was induced by a slight change in vocal fold stiffness, a similar abrupt change may be also induced by changes in other system parameters, which may affect the relative strength of the coexisting instabilities. Such parameters include flow separation point (as induced by change in flow rate or vocal fold geometry, Zhang, 2008), coupling to the sub- or supraglottal acoustics (Zhang *et al.*, 2006a, 2006b), and vocal fold geometry (Titze, 1994). For example, the abrupt change in vocal fold vibration pattern and phonation onset frequency, as observed in Fig. 2(b), can be also induced when the medial-surface shape changes from straight to convergent, without changing the body stiffness. This can be caused by the activation of the TA muscle, which may cause the inferior portion of the medial surface to bulge out. Such change in vocal fold geometry as induced by the contraction of the TA muscle has been suggested as a possible mechanism of register transition from chest to falsetto (Titze, 1994).

V. CONCLUSIONS

Using a linear stability analysis, phonation threshold pressure, phonation onset frequency, vocal fold vibration pattern, and sound production efficiency were investigated as a function of the mechanical and geometric parameters of a body-cover vocal fold model. The conclusions are as follows:

(1) Increasing body-cover stiffness ratio gradually restricted the vocal fold motion to the cover layer and the medial surface where the vocal fold motion is the most effective in terms of flow modulation, leading to increased voice production efficiency.

(2) A wave like motion was observed for vocal fold surface vibration of small wavelengths, which occurred at high body-cover stiffness ratios or when the vocal fold vibrated at a higher-order mode.
(3) In addition to a reduced wavelength along the vocal fold surface, self-oscillations at higher-order modes exhibited reduced superior-inferior motion. This allowed more energy to be spent on the medial-lateral motion along the superior portion of the medial surface and therefore higher sound production efficiency than that when the vocal fold vibrated at low-order modes.
(4) For small body-cover stiffness ratios, phonation onset frequency can be effectively controlled by varying the body stiffness, whereas for larger body-cover stiffness ratios, phonation onset frequency can be more effectively controlled by varying the cover stiffness.
(5) There was more than one group of eigenmodes that synchronized toward phonation onset in the coupled continuum system so that at least two potential instabilities (coupled-mode flutter in this study) existed. At certain conditions, the phonation threshold pressures associated with two such instabilities may be close to each other, and a slight change in the mechanical or geometric parameters of the system would cause phonation onset to switch from one instability to another, leading to sudden changes in (a) phonation onset frequency, (b) vocal fold vibration pattern, and (c) sound production efficiency. It is hypothesized that a similar mechanism may play a role in register change.

ACKNOWLEDGMENTS

This study was supported by research Grant Nos. R01 DC009229 and R01 DC003072 from the National Institute on Deafness and Other Communication Disorders, the National Institutes of Health.

Alipour, F., Berry, D. A., and Titze, I. R. (**2000**). "A finite-element model of vocal-fold vibration," J. Acoust. Soc. Am. **108**, 3003–3012.
Berry, D. A., Herzel, H., Titze, I. R., and Krischer, K. (**1994**). "Interpretation of biomechanical simulations of normal and chaotic vocal fold oscillations with empirical eigenfunctions," J. Acoust. Soc. Am. **95**, 3595–3604.
Berry, D. A., and Titze, I. R. (**1996**). "Normal modes in a continuum model of vocal fold tissues," J. Acoust. Soc. Am. **100**, 3345–3354.
Chan, R., Titze, I. R., and Titze, M. (**1997**). "Further studies of phonation threshold pressure in a physical model of the vocal fold mucosa," J. Acoust. Soc. Am. **101**, 3722–3727.
Colton, R. H., and Casper, J. K. (**1996**). *Understanding voice problems: A physiological perspective for diagnosis and treatment* (Lippincott Williams & Wilkins, Baltimore, MD).
Cook, D., and Mongeau, L. (**2007**). "Sensitivity of a continuum vocal fold model to geometric parameters, constraints, and boundary conditions," J. Acoust. Soc. Am. **121**, 2247–2253.
De Oliveira Rosa, M., Pereira, J. C., Grellet, M., and Alwan, A. (**2003**). "A contribution to simulating a three-dimensional larynx model using the finite element method," J. Acoust. Soc. Am. **114**, 2893–2905.
Decker, G. Z., and Thomson, S. (**2007**). "Computational simulations of vocal folds vibration: Bernouli versus Navier-Stokes," J. Voice **21**, 273–284.
Dollinger, M., Tayama, N., and Berry, D. A. (**2005**). "Empirical eigenfunctions and medial surface dynamics of a human vocal fold," Methods Inf. Med. **44**, 384–391.
Hirano, M. (**1974**). "Morphological structure of the vocal cord as a vibrator and its variations," Folia Phoniatr. **26**, 89–94.
Ishizaka, K., and Flanagan, J. L. (**1972**). "Synthesis of voiced sounds from a two-mass model of the vocal cords," Bell Syst. Tech. J. **51**, 1233–1267.
Ishizaka, K., and Flanagan, J. L. (**1977**). "Acoustic properties of longitudinal displacement in vocal cord vibration," Bell Syst. Tech. J. **56**, 889–918.
Isshiki, N. (**1998**). "Mechanical and dynamic aspects of voice production as related to voice therapy and phonosurgery," J. Voice **12**, 125–137.
Lous, N. J. C., Hofmans, G. C. J., Veldhuis, R. N. J., and Hirschberg, A. (**1998**). "A symmetrical two-mass vocal-fold model coupled to vocal tract and trachea, with application to prosthesis design," Acta Acust. **84**, 1135–1150.
Story, B. H., and Titze, I. R. (**1995**). "Voice simulation with a body-cover model of the vocal folds," J. Acoust. Soc. Am. **97**, 1249–1260.
Svec, J. G., Schutte, H. K., and Miller, D. G. (**1999**). "On pitch jumps between chest and falsetto registers in voice: Data from living and excised human larynges," J. Acoust. Soc. Am. **106**, 1523–1531.
Tao, C., and Jiang, J. J. (**2006**). "Simulation of vocal fold impact pressures with a self-oscillating finite-element model," J. Acoust. Soc. Am. **119**, 3987–3994.
Thomson, S. L., Mongeau, L., and Frankel, S. H. (**2005**). "Aerodynamic transfer of energy to the vocal folds," J. Acoust. Soc. Am. **118**, 1689–1700.
Titze, I. R. (**1988**). "The physics of small-amplitude oscillation of the vocal folds," J. Acoust. Soc. Am. **83**, 1536–1552.
Titze, I. R. (**1994**). *Principles of Voice Production* (Prentice-Hall, Englewood Cliffs, NJ).
Titze, I. R., and Hunter, E. J. (**2007**). "A two-dimensional biomechanical model of vocal fold posturing," J. Acoust. Soc. Am. **121**, 2254–2260.
Titze, I. R., Jiang, J., and Drucker, D. G. (**1988**). "Preliminaries to the body-cover theory of pitch control," J. Voice **1**, 314–319.
Titze, I. R., Schmidt, S., and Titze, M. (**1995**). "Phonation threshold pressure in a physical model of the vocal fold mucosa," J. Acoust. Soc. Am. **97**, 3080–3084.
Titze, I. R., and Strong, W. J. (**1975**). "Normal modes in vocal cord tissues," J. Acoust. Soc. Am. **57**, 736–744.
Titze, I. R., and Talkin, D. T. (**1979**). "A theoretical study of the effects of various laryngeal configurations on the acoustics of phonation," J. Acoust. Soc. Am. **66**, 60–74.
Tokuda, I. T., Horacek, J., Svec, J. G., and Herzel, H. (**2007**). "Comparison of biomechanical modeling of register transitions and voice instabilities with excised larynx experiments," J. Acoust. Soc. Am. **122**, 519–531.
Zhang, Z. (**2008**). "Influence of flow separation location on phonation onset," J. Acoust. Soc. Am. **124**, 1689–1694.
Zhang, Z., Mongeau, L., and Frankel, S. H. (**2002**). "Experimental verification of the quasi-steady approximation for aerodynamic sound generation by pulsating jets in tubes," J. Acoust. Soc. Am. **112**, 1652–1663.
Zhang, Z., Neubauer, J., and Berry, D. A. (**2006a**). "Aerodynamically and acoustically driven modes of vibration in a physical model of the vocal folds," J. Acoust. Soc. Am. **120**, 2841–2849.
Zhang, Z., Neubauer, J., and Berry, D. A. (**2006b**). "The influence of subglottal acoustics in laboratory models of phonation," J. Acoust. Soc. Am. **120**, 1558–1569.
Zhang, Z., Neubauer, J., and Berry, D. A. (**2007**). "Physical mechanisms of phonation onset: A linear stability analysis of an aeroelastic continuum model of phonation," J. Acoust. Soc. Am. **122**, 2279–2295.

Perceptual recalibration of speech sounds following speech motor learning

Douglas M. Shiller[a)]
School of Communication Sciences and Disorders, McGill University, 1266 Pine Avenue West, Montreal, Quebec H3G 1A8, Canada and Centre for Research on Language, Mind and Brain, McGill University, 3640 de la Montagne, Montreal, Quebec H3G 2A8, Canada

Marc Sato
Départment Parole et Cognition, GIPSA-lab, UMR CNRS 5216, Grenoble Universités, 1180 Avenue Centrale, BP 25, 38040 Grenoble Cedex 9, France

Vincent L. Gracco
School of Communication Sciences and Disorders, McGill University, 1266 Pine Avenue West, Montreal, Quebec H3G 1A8, Canada; Centre for Research on Language, Mind and Brain, McGill University, 3640 de la Montagne, Montreal, Quebec H3G 2A8, Canada; and Haskins Laboratories, 300 George Street, Suite 900, New Haven, Connecticut 06511

Shari R. Baum
School of Communication Sciences and Disorders, McGill University, 1266 Pine Avenue West, Montreal, Quebec H3G 1A8, Canada and Centre for Research on Language, Mind and Brain, McGill University, 3640 de la Montagne, Montreal, Quebec H3G 2A8, Canada

(Received 6 June 2008; revised 26 November 2008; accepted 8 December 2008)

The functional sensorimotor nature of speech production has been demonstrated in studies examining speech adaptation to auditory and/or somatosensory feedback manipulations. These studies have focused primarily on flexible motor processes to explain their findings, without considering modifications to sensory representations resulting from the adaptation process. The present study explores whether the perceptual representation of the /s-ʃ/ contrast may be adjusted following the alteration of auditory feedback during the production of /s/-initial words. Consistent with prior studies of speech adaptation, talkers exposed to the feedback manipulation were found to adapt their motor plans for /s/-production in order to compensate for the effects of the sensory perturbation. In addition, a shift in the /s-ʃ/ category boundary was observed that reduced the functional impact of the auditory feedback manipulation by increasing the perceptual "distance" between the category boundary and subjects' altered /s/-stimuli—a pattern of perceptual adaptation that was not observed in two separate control groups. These results suggest that speech adaptation to altered auditory feedback is not limited to the motor domain, but rather involves changes in both motor output and auditory representations of speech sounds that together act to reduce the impact of the perturbation. © *2009 Acoustical Society of America.* [DOI: 10.1121/1.3058638]

PACS number(s): 43.70.Mn, 43.71.Es, 43.70.Bk [AL] Pages: 1103–1113

I. INTRODUCTION

The functional sensorimotor nature of speech production has been demonstrated in a number of studies employing manipulations of both somatosensory and auditory feedback. Introducing unexpected auditory or somatosensory perturbations during speech production results in rapid (online) *compensatory* motor changes (Abbs and Gracco, 1984; Gracco and Abbs, 1985; Kawahara, 1995) while using more predictable and constant changes in auditory or somatosensory feedback results in a *recalibration* (or relearning) of the mapping between sensory signals and motor output (Baum and McFarland, 1997; Houde and Jordan, 1998; Jones and Munhall, 2000; Houde and Jordan, 2002; Jones and Munhall, 2003; Tremblay *et al.*, 2003; Nasir and Ostry, 2006; Purcell and Munhall, 2006a, 2006b; Villacorta *et al.*, 2007). For example, in a seminal investigation by Elman (1981), participants were presented with normal or frequency-shifted auditory feedback in real time during a task requiring them to instantaneously reproduce (i.e., shadow) shifts in fundamental frequency (F0) in synthetic vowel stimuli presented to them. Results revealed that in the frequency-shifted condition, participants shifted their F0 production in a compensatory fashion in an effort to achieve a target F0 in the feedback they received. A comparable compensatory effect was demonstrated using sentential stimuli (Elman, 1981).

Since that early study, numerous investigations have made use of real-time alterations in auditory feedback to explore resulting output characteristics (as well as patterns of generalization), including studies of shifts in fundamental frequency (F0—Kawahara, 1995; Jones and Munhall, 2000, 2005) and formant frequencies (Houde and Jordan, 1998, 2002; Purcell and Munhall, 2006a, 2006b; Villacorta *et al.*,

[a)]Author to whom correspondence should be addressed. Electronic mail: doug.shiller@mail.mcgill.ca

2007) during vowel production. All of these investigations have reported significant (but incomplete) compensatory effects in the speech motor output, as measured via kinematic or acoustic means, suggesting—not surprisingly—a strong link between auditory perception and motor output. Importantly, these studies have demonstrated not only motor corrections to counteract the effect of the perturbation, but also a persistence of those corrections (i.e., an aftereffect) once the perceptual manipulation is removed (e.g., Houde and Jordan, 1998; Jones and Munhall, 2000; Houde and Jordan, 2002; Jones and Munhall, 2005; Purcell and Munhall, 2006a, 2006b; Villacorta *et al.*, 2007). The fact that changes in the motor system do not disappear immediately likely reflects a change in motor representations due to a global remapping of the auditory-motor relationship (i.e., sensorimotor adaptation). These results are typically interpreted in the framework of forward and inverse internal models (Guenther, 1995; Perkell *et al.*, 1997; Kawato, 1999). Modulated response outputs during the manipulation of somatosensory or auditory feedback are thought to reflect feedback control mechanisms in which the expected sensory consequences of the speech-motor act are evaluated against the actual sensory input in order to further control production. These mechanisms also help to distinguish the sensory consequences of our own actions from sensory signals due to changes in the outside world (see Guenther, 2006, for a review).

Although these findings suggest an important relationship between an individual's perceptual "space" and his/her speech motor patterns, they fail to reveal the extent to which the perception and production systems may be truly integrated in producing such adaptation effects. One recent study attempted to draw a stronger association between perception and production, by examining how individual differences in auditory discrimination abilities may influence the degree to which speakers adapt to altered sensory feedback (Villacorta *et al.*, 2007). The investigation demonstrated that those individuals who exhibited greater sensitivity in auditory discrimination of relevant acoustic features (i.e., F1 frequencies) produced greater degrees of compensation to alterations in F1 in real-time auditory feedback. However, an important issue that has not been addressed is whether speech-motor adaptation to alterations in sensory feedback modifies the perceptual representation of speech sounds. Indeed, while the above-mentioned studies of speech adaptation to altered sensory feedback have indicated that sensory input affects speech motor control, the extent to which speech motor processes influence the speech sound representations that are central to how we perceive and produce speech has not been previously explored. That is, with changes in speech-motor output to alterations in auditory feedback, is there a concomitant adjustment in the perceptual space?

A significant body of speech perception research has demonstrated that sensory representations of speech sounds are flexible in response to changes in the sensory and linguistic aspects of speech input (Ladefoged and Broadbent, 1957; Miller and Liberman, 1979; Mann and Repp, 1980; Bertelson *et al.*, 2003; Norris *et al.*, 2003; Kraljic and Samuel, 2005). Listeners have been shown to rapidly compensate for short-term changes in speaking rate (Miller and Liberman, 1979) and phonetic context (Mann and Repp, 1980, 1981), as well as individually varying vocal tract characteristics (i.e., speaker normalization (Ladefoged and Broadbent, 1957; Nearey, 1989), in order to maintain perceptual accuracy in the face of such variability. Perceptual learning has been observed following exposure to foreign-accented talkers, resulting in improved word identification performance over time (Clarke and Garrett, 2004; Bradlow and Bent, 2008). In addition, a number of studies have shown changes in perceptual speech sound representations in listening tasks involving lexically ambiguous stimuli (Bertelson *et al.*, 2003; Norris *et al.*, 2003; Kraljic and Samuel, 2005; McQueen *et al.*, 2006).

In spite of this significant body of speech perception research demonstrating that sensory representations of speech sounds are flexible in response to changes in the sensory and linguistic aspects of speech input, studies of adaptation in speech production have focused primarily on the flexibility of motor processes in order to explain their findings, without regard for the possible contribution of changes in sensory representations that are presumed to constitute the acoustic "target" of speech movements. In the present study, we utilized a speech adaptation paradigm to explore whether auditory representations of speech sounds are, in fact, not static following speech production under conditions of altered auditory feedback, but rather can be adjusted to reflect changes in a talker's own speech output. The experimental procedure involved the real-time alteration of auditory feedback during the production of the sibilant /s/ in brief /s/-initial words. During an intensive period of speech practice under feedback-altered conditions, /s/-productions were examined for evidence of compensation for the manipulation. The persistence of any change in output following the sudden removal of the perturbation was also explored, as a reflection of a change in the speech-motor representation (Baum and McFarland, 1997; Houde and Jordan, 1998; Baum and McFarland, 2000; Houde and Jordan, 2002; Jones and Munhall, 2003; Tremblay *et al.*, 2003; Jones and Munhall, 2005; Purcell and Munhall, 2006a; Villacorta *et al.*, 2007). In addition to motor adaptation, the perceptual representation of the /s-ʃ/ contrast was examined immediately prior to and following the period of speech adaptation using a phoneme labeling task in order to determine whether boundary shifts would emerge, suggesting a perceptual adaptation to reduce the functional impact of the auditory feedback alteration.

II. METHODS

All procedures were approved by the Institutional Review Board of the Faculty of Medicine at McGill University and all subjects provided informed consent prior to testing.

A. Subjects

Thirty subjects were tested, all females (in order to reduce between-subject variability in fricative centroid frequency and vowel F0), between 19 and 30 years of age, and native speakers of North American English. Subjects had no reported history of speech or language disorder and no hear-

ing loss (confirmed by a pure-tone hearing screening conducted prior to testing). Subjects were also evaluated informally to rule out the presence of a functional speech disorder involving the production of /s/ (i.e., interdental, dentalized, lateral, or palatal lisp) by one of the authors (Shiller), who is a certified speech-language pathologist. The 30 subjects were randomly assigned to one of three groups, with ten subjects in each group: (1) a group that produced speech under conditions of *altered* auditory feedback (group AF), (2) a group that produced speech under conditions of *unaltered* auditory feedback (group UF), and (3) a group that passively listened to a sequence of frequency-altered speech stimuli that was matched to the stimuli perceived by subjects in the AF group, only in this case without speaking (PL group).

B. Audio recording

All groups were tested while seated in a sound attenuating testing room (Industrial Acoustics Company, Bronx, NY). For tasks involving speech production (AF and UF groups), subjects spoke into a condenser microphone (ME-66, Sennheiser, Germany) positioned 10 cm from the subject's mouth. The signal from the microphone was amplified to line level using a microphone preamplifier (model 302, Symetrix, Mountlake Terrace, WA) and then passively split into two identical channels, one of which was used to digitally record the subject's unprocessed speech signal, and the other which was sent to a digital signal processor (DSP) for processing. The output of the DSP was again passively split into two identical channels, one of which was digitally recorded (simultaneously with the unaltered speech signal) and the other which was presented back to the subject through circumaural headphones (SR-80, Grado Labs, Brooklyn, NY). The two channels of audio (unprocessed and processed by the DSP) were analog low-pass filtered at 22.0 kHz, and then digitized at 44.1 kHz (16 bit resolution) using an analog-to-digital converter (Transit, M-Audio, Irwindale, CA) attached via USB to a Toshiba laptop computer. The digitized audio signals were captured directly into MATLAB (v.7.4, Mathworks, Natick, MA) using the DATA ACQUISITION TOOLBOX (v. 2.10, Mathworks, Natick, MA).

C. Manipulation of auditory feedback

Subjects in the AF group produced a sequence of individual words under conditions of altered auditory feedback. Auditory feedback was manipulated using a DSP to gradually modify in real time the frequency spectrum of short words containing the initial fricative /s/ (e.g., "see"). Under conditions of maximal acoustic perturbation, the DSP shifted the first spectral moment (or centroid) of the fricative down by three semitones (averaging −1430 Hz across subjects), resulting in an acoustic signal that, while still categorically an /s/, was now closer in centroid frequency to the fricative /ʃ/ (as in "she"). The vowel spectrum was shifted to the same degree (reducing the fundamental frequency and all formants), which had the effect of lowering the perceived pitch of the voice. This modified acoustic signal was amplified and fed back to subjects through headphones.

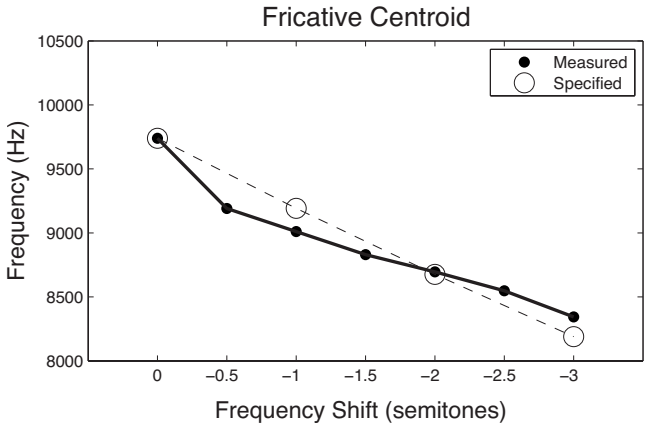


FIG. 1. Empirical test of the ability of the DSP to alter fricative centroid frequency. The dashed line (open circles) indicates the specified magnitude of spectral shifting (in steps of −0.5 semitones). The solid line (filled circles) indicates the measured centroid frequency following processing by the DSP. This test confirms the ability of the DSP to significantly reduce the centroid frequency of *s* within a reasonably small margin of error (<5% at the maximum specified shift of −3.0 semitones).

The DSP used to perform the real-time manipulation of auditory feedback was a commercial device designed for the manipulation of speech acoustic signals (SPX-1000, Yamaha, Japan). The device digitally samples an input signal at 44.1 kHz and uses a proprietary method to "pitch shift" an input speech signal at a delay of 10 ms. The ability of the device to manipulate the parameter of interest in the present study—fricative first spectral moment—was assessed empirically in order to verify the operation of the device. A speech utterance "see" was recorded on a digital audio tape (44.1 kHz, 16 bit sampling) and then played back through the DSP with the device set at 0 frequency shift (baseline, unmodified signal) and then at six linearly decreasing steps of frequency shift, ranging from −0.5 to −3.0 semitones (the maximum employed in the present study) in steps of −0.5 semitones. For each level of spectral shift, the output of the DSP was digitally recorded and the fricative centroid frequency was estimated using the same method as was employed for the analysis of speech production in the study (see Sec. II H). Figure 1 shows the recovered values of fricative centroid frequency at each level of frequency shift (the open circles show the specified shift and the filled circles indicate the recovered frequency). A close correspondence was observed between the specified and recovered frequencies; at the maximum level of spectral shifting (−3.0 semitones, which corresponds to the manipulation employed in the present study), the deviation between specified and recovered centroids was less than 5%. This was satisfactory for the purpose of the present study, as it confirmed the ability to significantly reduce the fricative centroid frequency with a reasonably small margin of error in the magnitude of the specified shift.

In previous studies of speech adaptation involving manipulations of auditory feedback during vowel production, masking noise has been mixed with the subject's altered speech signal before presenting it back to them through headphones in order to minimize the subject's perception of their unmodified speech output via air and bone conduction.

TABLE I. Speech production stimuli.

/s/-stimuli	sue, see, saw, sigh, say, sot, so, seep, sip, sop, sock, suit, seat, soy, soup
/ʃ/-stimuli	shoe, she, shaw, shy, shay, shot, show, sheep, ship, shop, shoot, sheet, shape, shake, shut

In the present study, no masking noise was added to the modified speech signal due to its potential impact on the noise spectrum of the voiceless fricatives that were of primary interest. Rather, the modified acoustic signal was simply gained sufficiently to limit the subjects' perception of their air/bone conducted speech signal. The amplifier gain levels and speaking volume employed in the present study were determined in pilot tests during which subjects were instructed to produce words at a comfortable speaking volume while the gain levels on the microphone preamplifier and DSP were adjusted. The goal was to achieve a sound level that was perceived by the subject to be loud (but not uncomfortable), that was reported by subjects to limit the perception of their own air/bone conduction speech signal, and that was found to yield clear evidence of speech adaptation in the fricative. With the microphone positioned at a fixed distance of 10 cm from the subjects, a gain level that achieved these goals was determined, at which point the peak level indicated on the volume unit (VU) meter of the microphone amplifier was noted (+2, on a scale from −7 to +3). For all subsequent subjects, with the variable gain settings on all sound equipment fixed, a consistent loudness level was achieved by ensuring that the subjects' mouths were positioned 10 cm from the microphone and that their speech yielded a target peak VU level of +2 [approximately 65 dB sound pressure level (SPL), as measured at the microphone using a handheld SPL meter]. Feedback regarding speaking volume was provided to subjects during a brief practice period prior to testing, as well as throughout the course of testing (using visual feedback from the experimenter), as required.

During the experiment, the DSP was controlled by the laptop computer using an external USB MIDI interface (MIDIsport 2×2, M-Audio). Control of the DSP was coordinated with the presentation of visual stimuli and audio recording using custom software written in MATLAB.

D. Task sequence: AF and UF groups (speech production)

For the AF and UF groups, the speech production task involved reading a sequence of words, presented one at a time on a computer display (65 point font) at a distance of 1.5 m. Each word was presented on the display for 3 s, followed by a 1 s period in which the display was blank, for a total interstimulus interval (onset to onset) of 4 s. Stimuli consisted of single-syllable English words of the form consonant-vowel (CV) or consonant-vowel-consonant (CVC). The total set of stimuli consisted of 30 words (Table I) 15 of which had /ʃ/ as the onset sound (not used for speech training, as described below) and 15 of which had /s/ as the onset sound (used for speech training). The stimulus word set included a range of vowel sounds; however, the final consonant sounds were restricted to the unvoiced stops: /p/, /t,/ or /k/ (see Table I). Subjects were instructed to read the words with the initial sound prolonged, following a model provided by the experimenter. In pilot studies, prolonged fricatives tended to show less overall variability in duration and amplitude as well as a more consistently present steady-state region at the fricative center. The modification also served to increase the subjects' exposure to the auditory feedback manipulation (in the AF group). The mean duration of baseline /s/-productions was 599 ms.[1]

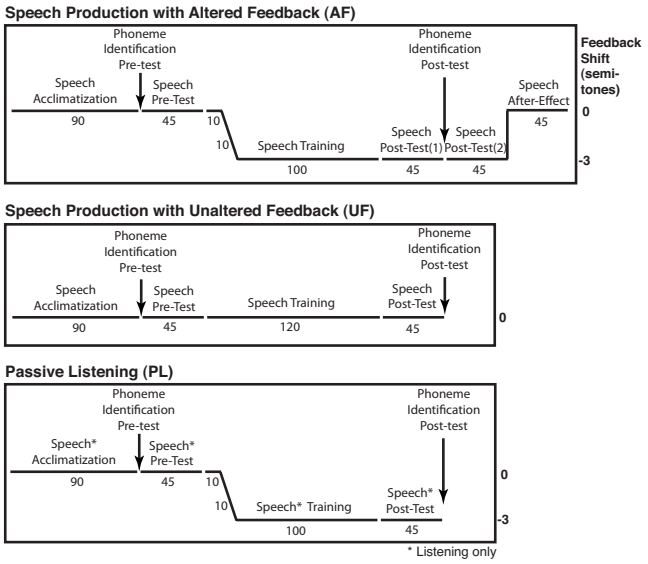


FIG. 2. Schematic depicting the sequence of procedures for each of the three groups of subjects: (1) speech production with altered auditory feedback (AF, top), (2) speech production with unaltered auditory feedback (UF, middle), and (3) passive listening (PL, bottom). The numbers underneath the horizontal lines indicate the number of words spoken. See text (Secs. II D and II E) for details.

For both the AF and UF groups, subjects underwent the following sequence of six procedures (see Fig. 2 for schematic).

(1) *Acclimatization period*. Subjects read aloud 90 words into a microphone while listening to their amplified (but otherwise unaltered) speech acoustic signal through headphones. The stimuli consisted of an equal proportion of words beginning with /s/ and /ʃ/, drawn from the full set of 30 stimulus items. Each word was presented three times in a fully randomized order.
(2) *Phoneme identification pretest*. Following the acclimatization period, subjects underwent the first of two phoneme identification tasks, which involved listening to synthetic speech stimuli through headphones and assigning a phoneme label to each token by responding on a computer keyboard (see *Phoneme identification* below).
(3) *Speech production pretest*. Subjects then underwent an assessment of speech production (the first of four for the AF group and the first of two for the UF group). The assessment involved the production of a restricted set of CV speech stimuli, consisting of the fricative /s/ combined with the three English "point" vowels: /u/ (as in "sue"), /i/ (as in "see") and /a/ (as in "saw"). Each word was produced ten times in a fully randomized order.

While the fricative /s/, produced in a range of vowel contexts, was the primary focus of the assessment, an additional 15 tokens involving the fricative /ʃ/ were also included in the assessment in order to evaluate the possible generalization of *s*-production training to /ʃ/. Due to space limitations, the analysis of /ʃ/-words is not reported in the present paper.

(4) *Speech training*. Subjects produced a sequence of words containing the fricative *s* exclusively (drawn from the full set of 15 /s/-stimuli). For the AF group, /s/-words were produced under conditions of altered auditory feedback, whereas for the UF group, /s/-words were produced with similarly amplified but otherwise unaltered auditory feedback. For the AF group, this training period began with 10 trials under unaltered feedback conditions, followed by the introduction of the acoustic perturbation (linearly ramped on over 10 trials), and then 100 trials under conditions of maximal acoustic perturbation (−3.0 semitones). For the UF group, the training period consisted of 120 trials under conditions of unaltered auditory feedback.

(5) *Speech production post-test (1)*. Both groups then underwent a second assessment of speech production consisting of the same 45 stimuli that were used in the speech production pretest (with the *s*-stimuli presented in a different random order). For both groups, auditory feedback conditions for this assessment remained unchanged from the preceding training period. Hence, for the AF group, subjects continued to experience the maximum level of acoustic perturbation, while for the UF group, feedback remained unaltered. For both groups, motor adaptation was assessed as the difference in /s/-centroid frequency between this test and the speech production pretest (item 3 above).

(6) *Phoneme identification post-test*. Following the speech production post-test, subjects in both groups underwent a second phoneme identification procedure (same as procedure 2 described above), but utilizing a different randomized order of perceptual stimuli.

For subjects in the AF group, two additional procedures were carried out following the phoneme identification post-test in order to examine the presence of a motor learning aftereffect.

(7) *Speech production post-test (2)*. A replication of the speech production post-test (item 5 above) was carried out (using a new randomized stimulus order) under conditions of maximal auditory feedback perturbation in order to ensure that any changes in speech output following speech production training were maintained during the phoneme identification post-test (a 7–8 min period during which the subject listened to speech stimuli without speaking).

(8) *Speech production aftereffect*. Immediately following the replicated speech production post-test (2), the perturbation of auditory feedback was suddenly and unexpectedly removed and a final assessment of speech production was carried out under conditions of unaltered feedback. Once again, the assessment consisted of the same stimuli as used in all previous assessments of speech production, but in a different randomized order.

E. Task sequence: PL group (passive listening)

Subjects in the PL group were seated in the sound-attenuating testing room and viewed the presentation of individual words on a computer display (using same text size, timing, and distance as described above). Simultaneous with each visual stimulus, a corresponding spoken word was presented auditorily through headphones.

The auditory stimuli were tokens digitally recorded from the output of the DSP for one subject who had participated in the study as a member of the AF group. The subject was selected on the basis of the following criteria: (1) a degree of /s/ motor adaptation that was similar to, but did not exceed, the average degree of adaptation for the entire AF group, and (2) a degree of token-to-token variability that did not exceed the average for the AF group. The subject that was selected exhibited a mean /s/-adaptation effect of 358 Hz (group mean: 527 Hz) and an across-token standard deviation of /s/-centroid frequency that ranged from 354.4 Hz in the baseline phase to 300.9 Hz at the end of training (group means: 503.9 and 452.4 Hz, respectively).

Following the same sequence as the AF group, subjects in the PL group first listened to 90 "acclimatization" trials and then underwent the phoneme identification pretest, using the same procedure administered to the AF and UF groups. Subjects then listened to the speech production pretest (45 trials), speech training (10 unaltered, 10 ramping on, and then 100 trials with maximal spectral shifting), and speech production post-test (45 trials at maximal spectral shift), after which they underwent a second test of phoneme identification (phoneme identification post-test).

In order to ensure that subjects in the PL group attended to the auditory and visual stimuli, they participated in a simple task in which they were instructed to indicate (using a computer keyboard) the number of letters contained in each word that was presented on the screen. Because the auditory presentation typically followed the visual presentation by approximately 1 s, subjects were instructed to respond with a key press only after the word had been presented auditorily.

F. Phoneme identification

A phoneme identification procedure was carried out in order to characterize the boundary between each subject's /s/ and /ʃ/ categories. The task employed a set of synthetic speech stimuli, which differed from each other along a nine-step acoustic spectral continuum from /s/ to /ʃ/. In the task, individual speech utterances, consisting of the sound embedded within the carrier: "a _ed" (e.g., "a shed" or "a said") were presented to subjects at a comfortable volume through headphones. Following each stimulus presentation, subjects labeled the fricative by pressing a corresponding key (labeled "s" or "sh") on a computer keyboard using the index and middle fingers of their dominant hand. Subjects were instructed to respond as quickly and accurately as possible following the onset of the stimulus. Key order was counterbalanced such that in each group, half of the subjects used each order. In total, ten tokens of each stimulus were pre-

sented to each subject in random order. Ten practice trials (randomly selected) were added at the beginning of each labeling session, yielding a total of 100 tokens per test (10 practice+90 test).

G. Synthetic speech stimuli

The synthetic speech stimuli used in the phoneme identification task were the same as those used in a recent study by Lane *et al.* (2007); thus, a detailed description and further references may be found in that paper. Briefly, speech stimuli were synthesized on the basis of natural exemplars of the phrases: "a said" and "a shed," produced by a female speaker. Two synthetic fricative segments were produced using a Klatt formant synthesizer, such that the frequency and amplitude of the formants (spectral peaks) were closely matched to the formants of the naturally produced fricatives. The parameters used to generate the two synthetic segments (/s/ and /ʃ/) were then adjusted slightly so that they differed only in terms of formant amplitude (i.e., formant frequencies were aligned). Finally, using these two synthetic fricatives as the boundary stimuli, a /s-ʃ/ continuum was generated by interpolating the formant amplitudes over seven intermediate steps. The interpolation was not perfectly linear, due to the constraint of having to use integer values to specify formant amplitudes in the Klatt synthesizer, as well as the desire for a continuum that was roughly balanced for the number of stimulus steps on either side of the category boundary. Finally, the remaining portions of the utterance (preceding vowel "a" and following coda "-ed") were concatenated with the fricative to yield the complete stimuli (e.g., "a said") (see Lane *et al.*, 2007 for further details).

H. Data analysis

1. Acoustics

Adaptation of *s*-production was assessed by examining changes in the first spectral moment: a stable acoustic property of fricatives (Behrens and Blumstein, 1988; Jongman *et al.*, 2000) that has been used to evaluate the accuracy of *s*-production in a number of studies involving the manipulation of sensory feedback (Baum and McFarland, 1997, 2000). The first spectral moment, or centroid frequency, is a measure of central tendency in the spectral domain and is computed as the amplitude-weighted mean of the frequency spectrum (obtained by discrete Fourier transform). For each utterance, the mean frequency centroid was obtained for a 100 ms window about the midpoint of the fricative.

In order to evaluate the magnitude of changes in fricative production within each group of subjects, mean centroid estimates were first obtained separately for each subject. For each of the speech assessments (four for the AF group, and two for the UF group—as described above), an average centroid was computed across all /s/-utterances (collapsing across the three vowel contexts[2]). Following the calculation of these within-subject mean values, difference scores were computed between relevant assessments. Two difference scores, in particular, were of interest: First, for all subjects in the AF and UF groups, the difference between the pretraining (baseline) and post-training (immediately following the training period) assessments was obtained in order to measure the direction and magnitude of the speech training effect; second, for all subjects in the AF group, the difference between the initial baseline assessment and the final assessment (immediately following removal of the shift) was obtained in order to measure the direction and magnitude of the speech adaptation aftereffect (the persistence of motor learning effects in the absence of auditory perturbation). Statistical pairwise comparisons were carried out using *t*-tests, corrected for multiple comparisons (familywise $p<0.05$) using Holm's sequential Bonferroni procedure.

2. Phoneme identification function

The set of response data from the phoneme identification task (selection of "s" or "sh" for ten tokens of nine different stimuli) was used to estimate parameters for the /s-ʃ/ identification function for the two tests (pre- and postpractice) within each subject. This was done by first computing the proportion of "s" responses for each stimulus (1.0=100% "s" response), linearly interpolating to an interval of 0.1 stimulus steps, and then fitting a four-parameter logistic function (sigmoid) to the resulting data points. The sigmoid parameters include the stimulus step at which the proportion of "s" responses is 0.5, which was taken as the boundary between the "s" and "sh" responses. As in the case of the centroid frequency measure, estimates of the location of the sigmoid boundary from the two assessments (pre- and post-training) were converted to a difference score for each subject in order to measure the direction and magnitude of the training effect. Pairwise comparisons between the AF and other groups were carried out using independent-samples t-tests, corrected for multiple comparisons (familywise $p<0.05$) using Holm's sequential Bonferroni procedure.

III. RESULTS

On average, the auditory feedback manipulation resulted in a 1430 Hz reduction in /s/-centroid frequency across subjects in the AF group. Following the period of speech practice under these altered auditory conditions, these subjects adjusted their production of /s/ to counteract the effect of the sensory perturbation (Fig. 3, left bar). Specifically, subjects showed an increase in /s/-centroid frequency (i.e., in the direction opposite that of the spectral shift) averaging 529 Hz following training. The mean change in fricative centroid frequency within each of the three tested vowel contexts (/u/, /i/, and /a/) was found to vary somewhat (averaging 567.5, 462.7, and 556.0 Hz, respectively); however, the difference between contexts was not statistically reliable ($F[2,18]=0.542$, $p>0.05$). The overall change in speech output was found to persist following the sudden removal of the perturbation (Fig. 3, middle bar), averaging 490 Hz across a final block of 30 /s/-word trials with unaltered auditory feedback. Note that a deadaptation effect of 113 Hz was observed between the first five and final five /s/-words in this final testing block; however, the reduction was not statistically reliable [$t(9)=0.49$, $p>0.05$]. The persistence of motor adaptation following removal of the sensory perturbation indicates that the changes in speech production were not the result of on-

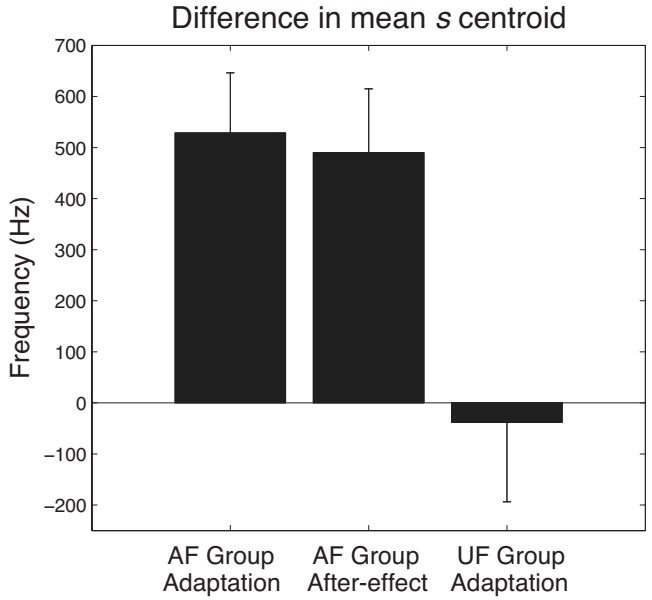


FIG. 3. Observed changes in speech output. The observed changes in /s/-*centroid* frequency associated with speech practice under conditions of altered auditory feedback (AF group; *left* and *middle* bars) and under conditions of unaltered auditory feedback (UF group; *right bar*). The error bars show one standard error. See text for details.

line corrections, but rather resulted from changes in the neurally specified motor plan for the production of the /s/-sound. The observed adaptation and aftereffects for the AF group were found to be reliably larger than the change in /s/-production observed in a group of control subjects (UF group) who underwent an identical period of intensive speech practice with *s*-initial words under conditions of *unaltered* auditory feedback [AF adaptation versus UF adaptation: $t(18)=2.9$, $p<0.05$; AF aftereffect versus UF adaptation: $t(18)=2.7$, $p<0.05$]. For this control group, little change in speech output was observed following speech practice [mean: −39 Hz, $t(9)=0.25$, $p>0.05$], indicating that the effects observed in the experimental group did not arise from aspects of the experimental setup unrelated to the spectral manipulation (such as the high concentration of /s/-words in the speech corpus) (Fig. 3, right bar).

In addition to assessing changes in /s/-production resulting from the auditory feedback manipulation, an evaluation of speech perception was carried out for both the AF and UF groups prior to and following the period of intensive /s/-practice in order to examine changes in the representation of speech-sound categories resulting from the manipulation. Mean /s-ʃ/ identification functions and their associated boundary locations (the stimulus step at which the proportion of "s" responses is 0.5) are shown in Figs. 4(a) and 4(b). Following the period of /s/-word practice with altered auditory feedback (in which the fricative centroid frequency was shifted lower, toward /ʃ/), subjects in the AF group exhibited a change in their /s-ʃ/ identification function that acted to reduce the impact of the perturbation. Specifically, the location of the /s-ʃ/ category boundary was shifted toward a lower centroid frequency (in the direction of the /ʃ/ category),

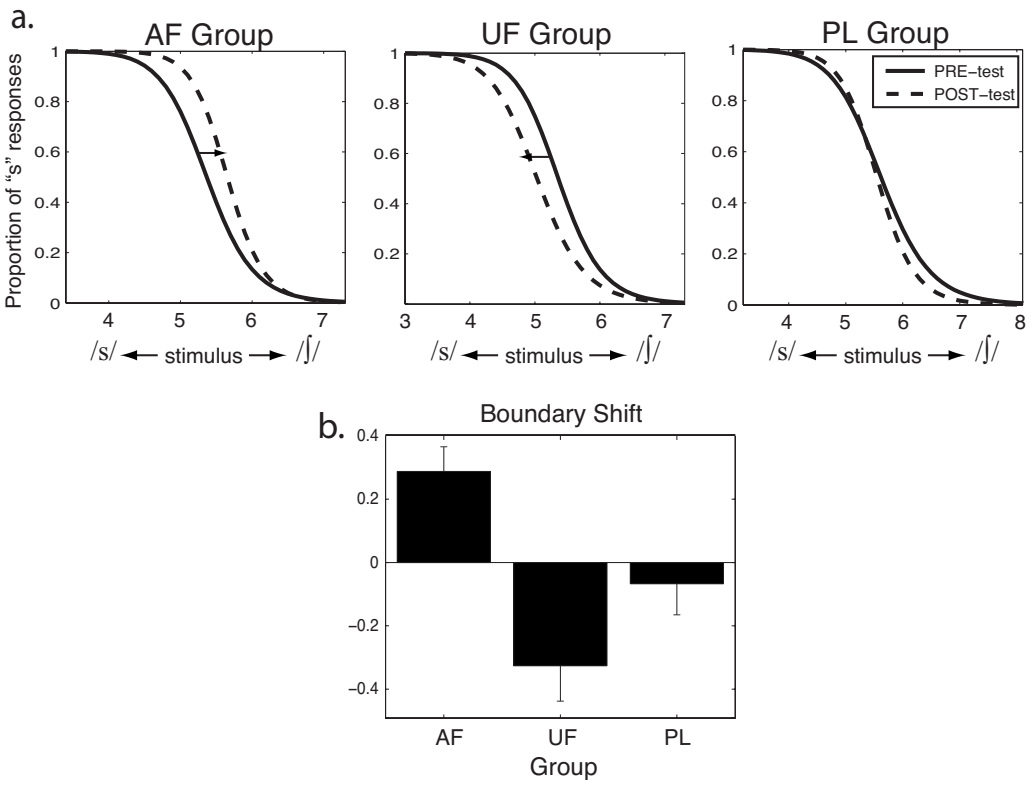


FIG. 4. Evaluation of speech-sound perception. (a) Mean pretest and post-test identification functions for each group, based on average sigmoid boundary and slope parameters (see Sec. II H). Notably, for subjects in the AF group, the boundary is seen to shift in the direction of the auditory perturbation (toward *sh*) following speech practice. In contrast, for the UF group, the boundary shifts toward the *s*-stimuli following speech practice, while for the PL group, passive listening of frequency-altered speech stimuli results in little change in boundary location. (b) Mean change in boundary location following the period of speech practice (AF and UF groups), or passive listening to altered stimuli (PL group). The error bars show one standard error.

thereby reducing the discrepancy between the spectrally altered *s* acoustics and the perceptual representation of the *s* category. The boundary shift for the AF group differed reliably from that of the UF group, for whom auditory feedback was not spectrally altered [AF boundary shift versus UF boundary shift: $t(18)=4.5$, $p<0.05$]. In contrast with the AF group, subjects in the UF group exhibited a shift in the /s-ʃ/ category boundary toward stimuli containing a *higher* centroid frequency (i.e., further toward the /s/-category) following the period of /s/-word practice [$t(9)=2.9$, $p<0.05$].

In order to determine whether the perceptual adaptation observed in the AF group might have been due solely to perceptual exposure to frequency-shifted /s/-words, the effects observed in the AF group were compared with those of a third group of subjects (PL group) who passively listened to a sequence of frequency altered speech stimuli—matched to the stimuli perceived by the AF group—without producing any words themselves. Compared with the relatively small boundary shift observed for the PL group, a significantly larger change in the boundary location was observed for the AF group [AF versus PL groups: $t(18)=2.84$, $p<0.05$], as shown in Figs. 4(a) and 4(b).

IV. DISCUSSION

In the present study, we have investigated speech motor adaptation to altered auditory feedback during the production of the sibilant /s/. Talkers exposed to the auditory manipulation (AF group) were found to alter their speech output to compensate for the effects of the sensory perturbation. That is, they produced the sibilant with an increased centroid frequency in order to offset the decrease in frequency in the fed-back acoustic signal. Furthermore, the compensatory change was found to persist following the sudden removal of the perturbation, indicating that the change in output was not the result of online feedback-based correction but rather reflected a change in the underlying motor plan for /s/. The finding of speech-motor learning is consistent with numerous studies of speech adaptation involving manipulations of auditory feedback (e.g., involving vowel formants and fundamental frequency—Houde and Jordan, 1998, 2002; Jones and Munhall, 2005; Purcell and Munhall, 2006a; Villacorta *et al.*, 2007) and somatosensory feedback (e.g., palatal or dental prosthesis, altered jaw path—Baum and McFarland, 1997, 2000; Jones and Munhall, 2003; Tremblay *et al.*, 2003). In comparison with the subjects in the AF condition, a group of subjects who underwent an identical set of speech production procedures without any alteration to auditory feedback (UF group) did not show a shift in speech output following the period of speech practice with /s/-initial words.

Previous studies of speech-motor learning using purely auditory manipulations of sensory feedback have examined the production of vowel formants and fundamental frequency (Houde and Jordan, 1998; Jones and Munhall, 2000; Houde and Jordan, 2002; Jones and Munhall, 2005; Purcell and Munhall, 2006a; Villacorta *et al.*, 2007). In contrast, the present study provides a demonstration of speech adaptation following a purely auditory perturbation of a consonantal speech sound. Studies of motor adaptation involving the production of sibilants have, of course, been carried out using perturbations that alter the shape of the vocal tract (e.g., using palatal or dental prostheses—Baum and McFarland, 1997, 2000; Jones and Munhall, 2003). The impact of these physical manipulations is complex—altering both tactile/somatosensory feedback as well as speech acoustics—therefore it is impossible to determine the extent to which subjects in these studies were compensating for changes in a particular sensory modality. In contrast, the observation of motor adaptation in the present study offers direct support for the role of auditory input in the production of the sibilant /s/.

In addition to the demonstration of speech motor adaptation in the present study, an investigation of the perceptual representation of the /s-ʃ/ contrast revealed a shift in the phoneme identification boundary following speech practice under feedback-altered conditions (AF group). Specifically, the boundary was found to shift in the same direction as the feedback manipulation (toward a lower centroid frequency, i.e., the /ʃ/ category). Such a perceptual shift is adaptive, since it has the effect of reducing the functional impact of the auditory feedback manipulation by increasing the perceptual distance between the category boundary and the subjects' altered /s/-stimuli. The extent to which this perceptual adaptation effect might have arisen solely due to the perceived change in acoustic properties of /s/-words was explored by exposing a set of similarly altered stimuli to a group of control subjects who themselves did not produce any speech (PL group). Following the exposure to frequency-shifted /s/-words (using the same testing sequence as subjects in the AF group), no overall change in these subjects' perceptual boundaries between /s/ and /ʃ/ was observed, suggesting that the perceptual changes observed in the AF group were indeed related to a perceptuomotor adaptation.

While subjects in the PL group exhibited no change in perceptual representations, subjects in the UF control group (who produced repeated /s/-words under conditions of unaltered auditory feedback) did exhibit a shift in the perceptual category boundary following the training phase, however, in this case toward a *higher* centroid frequency (i.e., in the direction of the /s/ category). Such a boundary shift is consistent with the well-known *selective adaptation effect* (Eimas and Corbit, 1973), a perceptual phenomenon in which repeated exposure to stimuli at one end of a phonetic continuum results in a shift in the phoneme identification boundary toward the repeated stimulus. Given that subjects in all three groups were exposed to repeated /s/-stimuli, the question remains as to why a selective adaptation effect was only observed in the UF group. One possibility is that the selective adaptation effect depends critically on the repeated presentation of canonical endpoint stimuli (i.e., unaltered /s/-words), as have typically been employed in studies investigating the phenomenon. Indeed, a small number of selective adaptation studies have included degraded or otherwise noncanonical speech stimuli and in spite of the subjects' tendency to categorize these stimuli as belonging to a particular phonemic category, a weaker or nonexistent selective adaptation effect was reported in comparison with end-

point stimuli (Sawusch and Pisoni, 1976; Blumstein *et al.*, 1977; Cheesman and Greenwood, 1995).

In addition to shifting perceptual category boundaries, the selective adaptation effect has been shown to impact the *production* of speech sounds in a number of studies (Cooper and Lauritsen, 1974; Cooper and Nager, 1975; Jamieson and Cheesman, 1987). Cooper and Lauritsen's (1974) seminal study demonstrated that repetitive listening to a CV syllable with an initial voiceless stop consonant (which presumably resulted in a perceptual selective adaptation effect) caused subjects to produce a shorter voice onset time (VOT) for voiceless stop consonants in CV syllables. Given this finding, it is perhaps surprising that the UF group in the present study did not show a corresponding shift in speech output. This discrepancy is likely explained by the substantially smaller number of trials in the current study as compared to that utilized by previous studies such as Cooper and Lauritsen's (1974), in which a significant shift in motor output was observed following many hundreds of trials. In addition, the impact of the selective adaptation effect on speech production has only been demonstrated in the context of a single speech parameter—VOT for voiceless plosives—with a failure to demonstrate a corresponding change in VOT output following selective adaptation to their voiced counterparts (Cooper and Lauritsen, 1974; Cooper and Nager, 1975; Jamieson and Cheesman, 1987). The apparent phonetic specificity of this perceptuomotor selective adaptation effect leaves open the question of whether a similar phenomenon would be expected in the case of fricatives.

A number of prior studies have demonstrated changes in the perception of phoneme categories under a range of listening conditions (Ladefoged and Broadbent, 1957; Miller and Liberman, 1979; Mann and Repp, 1980; Bertelson *et al.*, 2003; Norris *et al.*, 2003; Kraljic and Samuel, 2005; McQueen *et al.*, 2006). The question arises whether similar input processes might underlie the perceptual adaptation observed in the present study. For example, it has been shown that when listeners are exposed to phonetic stimuli that are perceptually ambiguous, e.g., a word containing a fricative that is acoustically midway between two categories (as in Norris *et al.*, 2003; Kraljic and Samuel, 2005; McQueen *et al.*, 2006), their phoneme category boundary may shift in order to reduce the ambiguity. In the present study, we investigated the possible contribution of a lexically based perceptual adaptation effect by examining the change in the /s-ʃ/ perceptual boundary in a group of subjects (PL group) who listened passively to a sequence of /s/-words that were frequency-altered in the same manner as the speech feedback presented to the AF group. The PL group showed no overall shift in their /s-ʃ/ category boundary following the perceptual exposure. The lack of a lexically based perceptual adaptation effect in the present study is likely related to the nature of the auditory stimuli involved. While previous perceptual adaptation studies have employed acoustic stimuli that were designed to be maximally acoustically and perceptually ambiguous, the /s/-productions in the present study were not degraded to such a large degree. As a result, the stimuli in the present study—while not typical /s/ tokens—remained within the /s/ category and hence were not, in fact, perceptually ambiguous.[3]

The results of the present study provide new insights into aspects of previous studies of speech adaptation that have been poorly understood, in particular, the common finding that adaptation in motor output is less than complete (Baum and McFarland, 1997; Houde and Jordan, 1998, 2002; Tremblay *et al.*, 2003; Jones and Munhall, 2005). The finding of incomplete adaptation in prior studies has been discussed primarily in terms of factors that might impact the motor planning process (although, see Houde and Jordan, 2002). These include articulatory constraints, auditory acuity, the balance of feedback and feed-forward control mechanisms, modulation of the salience of auditory feedback over time, and, in the case of purely auditory perturbations, the possible impact of unaltered somatosensory targets for speech sounds (e.g., Savariaux *et al.*, 1995; Baum and McFarland, 2000; Houde and Jordan, 2002; Purcell and Munhall, 2006a; Villacorta *et al.*, 2007). In addition, individual differences in the degree of motor adaptation may be related to speech production characteristics such as token-to-token variability and the distinctiveness of phoneme production—properties that could influence the impact of a given feedback manipulation. In contrast with these accounts, the present results suggest that speech adaptation to altered auditory feedback is not limited to the motor domain, but rather involves changes in both motor output and auditory representations of speech sounds that together act to reduce the impact of the perturbation. Thus, the contribution of both sensory and motor adaptive processes, in conjunction with those factors already suggested, might offer a better account of speech adaptation to altered sensory conditions.

These results appear in keeping with the long-standing proposal that speech perception and speech production are closely linked processes, as first detailed in the motor theory of speech perception (Liberman *et al.*, 1967; Liberman and Mattingly, 1985; Liberman and Whalen, 2000; see also Galantucci *et al.*, 2006), itself an extension of earlier theories of perception in which motor actions were viewed as possible integral components of the perceptual process (Berkeley, 1709; Washburn, 1926; Festinger *et al.*, 1967). The findings also appear consistent with a growing number of behavioral, neuroimaging, and neurophysiological studies providing supporting evidence for sensorimotor interactions during both speech perception and production (Cooper and Lauritsen, 1974; Fadiga *et al.*, 2002; Watkins *et al.*, 2003; Wilson *et al.*, 2004; Sams *et al.*, 2005; Pulvermuller *et al.*, 2006; Gentilucci and Bernardis, 2007; Meister *et al.*, 2007; Skipper *et al.*, 2007; Tourville *et al.*, 2008). The results of the present study provide new behavioral evidence showing not only a link between the processes underlying speech perception and production, but a functional and plastic change involving both input and output processes simultaneously.

ACKNOWLEDGMENTS

Supported by research grants from NSERC-Canada, a postdoctoral fellowship from the Centre for Research on

Language, Mind and Brain, McGill University, Canada (D.M.S.), and a Richard H. Tomlinson postdoctoral fellowship from McGill University, Canada (M.S.).

[1]Importantly, the AF and UF groups did not differ appreciably in /s/ duration (584 ms versus 613 ms). In addition, for the AF group, the mean fricative duration was very similar prior to and following the period of training under conditions of altered auditory feedback (584 ms versus 593 ms), which indicates that differences in duration did not play a role in the observed adaptation effects.

[2]As described in Sec. III, vowel context was not found to have a reliable impact on the magnitude of the /s/ motor adaptation effect in the AF group.

[3]In order to verify that the three-semitone frequency shift did not result in ambiguous stimuli, a small control study was carried out as follows: Single tokens of the words "sue" and "shoe" were selected from the baseline productions of the first five subjects in the AF group. The five tokens of "sue" were then played back through the DSP, applying a frequency shift of −3.0 semitones, and the output was digitally recorded. All speech stimuli (unshifted "sue" and "shoe," plus the frequency shifted "sue") were normalized in order to remove any differences in fricative amplitude and duration within each of the five speakers. Vowel differences were also eliminated by concatenating the normalized fricatives with a single vowel (taken from the unfiltered production of "sue") for each subject, resulting in three stimuli per talker that differed solely in the frequency composition of the fricatives. A perceptual test was carried out in which 6 female listeners made phoneme identity judgments ("s" versus "sh") on 85 speech stimuli played back through headphones under conditions comparable to those used for perceptual testing in the main study. The stimulus set contained five repetitions of the 15 different speech stimuli presented in randomized order (totaling 25 repetitions of each fricative), plus 10 initial practice trials (randomly selected). As expected, all eight listeners consistently labeled the unaltered /s/-stimuli as "s" (averaging 99% "s" responses across listeners) and the unaltered /ʃ/-stimuli as "sh" (2% "s" responses). Notably, subjects were also highly consistent in their labeling of the frequency-shifted /s/-stimuli as "s" (99% "s" responses). Thus, while the /s/-productions were altered by the DSP, the magnitude of the frequency shift was not great enough to render the stimuli phonemically ambiguous.

Abbs, J. H., and Gracco, V. L. (**1984**). "Control of complex motor gestures: Orofacial muscle responses to load perturbations of lip during speech," J. Neurophysiol. **51**, 705–723.

Baum, S. R., and McFarland, D. H. (**1997**). "The development of speech adaptation to an artificial palate," J. Acoust. Soc. Am. **102**, 2353–2359.

Baum, S. R., and McFarland, D. H., (**2000**). "Individual differences in speech adaptation to an artificial palate," J. Acoust. Soc. Am. **107**, 3572–3575.

Behrens, S. J., and Blumstein, S. E. (**1988**). "Acoustic characteristics of English voiceless fricatives: A descriptive analysis," J. Phonetics **16**, 295–298.

Berkeley, G., (**1709**). "An essay towards a new theory of vision," in *The Works of George Berkeley*, edited by A. A. Luce and T. E. Jessop (Nelson, London).

Bertelson, P., Vroomen, J., and De Gelder, B., (**2003**). "Visual recalibration of auditory speech identification: A McGurk aftereffect," Psychol. Sci. **14**, 592–597.

Blumstein, S. E., Stevens, K. N., and Nigro, G. N. (**1977**). "Property detectors for bursts and transitions in speech perception," J. Acoust. Soc. Am. **61**, 1301–1313.

Bradlow, A. R., and Bent, T., (**2008**). "Perceptual adaptation to non-native speech," Cognition **106**, 707–729.

Cheesman, M. F., and Greenwood, K. G. (**1995**). "Selective adaptation by context-conditioned fricatives," J. Acoust. Soc. Am. **97**, 531–538.

Clarke, C. M., and Garrett, M. F., (**2004**). "Rapid adaptation to foreign-accented English," J. Acoust. Soc. Am. **116**, 3647–3658.

Cooper, W. E., and Lauritsen, M. R. (**1974**). "Feature processing in the perception and production of speech," Nature (London) **252**, 121–123.

Cooper, W. E., and Nager, R. M. (**1975**). "Perceptuo-motor adaptation to speech: An analysis of bisyllabic utterances and a neural model," J. Acoust. Soc. Am. **58**, 256–266.

Eimas, P. D., and Corbit, J. D. (**1973**). "Selective adaptation of linguistic feature detectors," Cogn. Psychol. **4**, 99–109.

Elman, J. L. (**1981**). "Effects of frequency-shifted feedback on the pitch of vocal productions," J. Acoust. Soc. Am. **70**, 45–50.

Fadiga, L., Craighero, L., Buccino, G., and Rizzolatti, G., (**2002**). "Speech listening specifically modulates the excitability of tongue muscles: A TMS study," Eur. J. Neurosci. **15**, 399–402.

Festinger, L., Burnham, C. A., Ono, H., and Bamber, D. (**1967**). "Efference and the conscious experience of perception," J. Exp. Psychol. **74**, 1–36.

Galantucci, B., Fowler, C. A., and Turvey, M. T., (**2006**). "The motor theory of speech perception reviewed," Psychon. Bull. Rev. **13**, 361–377.

Gentilucci, M., and Bernardis, P., (**2007**). "Imitation during phoneme production," Neuropsychologia **45**, 608–615.

Gracco, V. L., and Abbs, J. H. (**1985**). "Dynamic control of the perioral system during speech: Kinematic analyses of autogenic and nonautogenic sensorimotor processes," J. Neurophysiol. **54**, 418–432.

Guenther, F. H. (**1995**). "Speech sound acquisition, coarticulation, and rate effects in a neural network model of speech production," Psychol. Rev. **102**, 594–621.

Guenther, F. H., (**2006**). "Cortical interactions underlying the production of speech sounds," J. Commun. Disord. **39**, 350–365.

Houde, J. F., and Jordan, M. I. (**1998**). "Sensorimotor adaptation in speech production," Science **279**, 1213–1216.

Houde, J. F., and Jordan, M. I., (**2002**). "Sensorimotor adaptation of speech I: Compensation and adaptation," J. Speech Lang. Hear. Res. **45**, 295–310.

Jamieson, D. G., and Cheesman, M. F. (**1987**). "The adaptation of produced voice-onset time," J. Phonetics **15**, 15–27.

Jones, J. A., and Munhall, K. G., (**2000**). "Perceptual calibration of F0 production: Evidence from feedback perturbation," J. Acoust. Soc. Am. **108**, 1246–1251.

Jones, J. A., and Munhall, K. G., (**2003**). "Learning to produce speech with an altered vocal tract: The role of auditory feedback," J. Acoust. Soc. Am. **113**, 532–543.

Jones, J. A., and Munhall, K. G., (**2005**). "Remapping auditory-motor representations in voice production," Curr. Biol. **15**, 1768–1772.

Jongman, A., Wayland, R., and Wong, S., (**2000**). "Acoustic characteristics of English fricatives," J. Acoust. Soc. Am. **108**, 1252–1263.

Kawahara, H. (**1995**). "Hearing voice: Transformed auditory feedback effects on voice pitch control," Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI'95) Workshop on Computational Auditory Scene Analysis, pp. 143–148.

Kawato, M. (**1999**). "Internal models for motor control and trajectory planning," Curr. Opin. Neurobiol. **9**, 718–727.

Kraljic, T., and Samuel, A. G., (**2005**). "Perceptual learning for speech: Is there a return to normal?," Cognit Psychol. **51**, 141–178.

Ladefoged, P., and Broadbent, D. E. (**1957**). "Information conveyed by vowels," J. Acoust. Soc. Am. **29**, 98–104.

Lane, H., Denny, M., Guenther, F. H., Hanson, H. M., Marrone, N., Matthies, M. L., Perkell, J. S., Stockmann, E., Tiede, M., Vick, J., and Zandipour, M., (**2007**). "On the structure of phoneme categories in listeners with cochlear implants," J. Speech Lang. Hear. Res. **50**, 2–14.

Liberman, A. M., Cooper, F. S., Shankweiler, D. P., and Studdert-Kennedy, M. (**1967**). "Perception of the speech code," Psychol. Rev. **74**, 431–461.

Liberman, A. M., and Mattingly, I. G. (**1985**). "The motor theory of speech perception revised," Cognition **21**, 1–36.

Liberman, A. M., and Whalen, D. H., (**2000**). "On the relation of speech to language," Trends Cogn. Sci. **4**, 187–196.

Mann, V. A., and Repp, B. H. (**1980**). "Influence of vocalic context on perception of the [zh]-[s] distinction," Percept. Psychophys. **28**, 213–228.

Mann, V. A., and Repp, B. H. (**1981**). "Influence of preceding fricative on stop consonant perception," J. Acoust. Soc. Am. **69**, 548–558.

McQueen, J. M., Norris, D., and Cutler, A., (**2006**). "The dynamic nature of speech perception," Lang Speech **49**, 101–112.

Meister, I. G., Wilson, S. M., Deblieck, C., Wu, A. D., and Iacoboni, M., (**2007**). "The essential role of premotor cortex in speech perception," Curr. Biol. **17**, 1692–1696.

Miller, J. L., and Liberman, A. M. (**1979**). "Some effects of later-occurring information on the perception of stop consonant and semivowel," Percept. Psychophys. **25**, 457–465.

Nasir, S. M., and Ostry, D. J., (**2006**). "Somatosensory precision in speech production," Curr. Biol. **16**, 1918–1923.

Nearey, T. M. (**1989**). "Static, dynamic, and relational properties in vowel perception," J. Acoust. Soc. Am. **85**, 2088–2113.

Norris, D., McQueen, J. M., and Cutler, A., (**2003**). "Perceptual learning in speech," Cognit Psychol. **47**, 204–238.

Perkell, J. S., Matthies, M. L., Lane, H., Guenther, F. H., Wilhelms-

Tricarico, R., Wozniak, J., and Guiod, P. (**1997**). "Speech motor control: Acoustic goals, saturation effects, auditory feedback and internal models," Speech Commun. **22**, 227–250.
Pulvermuller, F., Huss, M., Kherif, F., Moscoso del Prado Martin, F., Hauk, O., and Shtyrov, Y., (**2006**). "Motor cortex maps articulatory features of speech sounds," Proc. Natl. Acad. Sci. U.S.A. **103**, 7865–7870.
Purcell, D. W., and Munhall, K. G., (**2006a**). "Adaptive control of vowel formant frequency: Evidence from real-time formant manipulation," J. Acoust. Soc. Am. **120**, 966–977.
Purcell, D. W., and Munhall, K. G., (**2006b**). "Compensation following real-time manipulation of formants in isolated vowels," J. Acoust. Soc. Am. **119**, 2288–2297.
Sams, M., Mottonen, R., and Sihvonen, T., (**2005**). "Seeing and hearing others and oneself talk," Brain Res. Cognit. Brain Res. **23**, 429–435.
Savariaux, C., Perrier, P., and Orliaquet, J. P. (**1995**). "Compensation strategies for the perturbation of the rounded," J. Acoust. Soc. Am. **98**, 2428–2422.
Sawusch, J. R., and Pisoni, D. B. (**1976**). "Simple and contingent adaptation effects for place of articulation in stop consonants," Percept. Psychophys. **23**, 125–131.
Skipper, J. I., van Wassenhove, V., Nusbaum, H. C., and Small, S. L., (**2007**). "Hearing lips and seeing voices: How cortical areas supporting speech production mediate audiovisual speech perception," Cereb. Cortex **17**, 2387–2399.
Tourville, J. A., Reilly, K. J., and Guenther, F. H., (**2008**). "Neural mechanisms underlying auditory feedback control of speech," Neuroimage **39**, 1429–1443.
Tremblay, S., Shiller, D. M., and Ostry, D. J., (**2003**). "Somatosensory basis of speech production," Nature (London) **423**, 866–869.
Villacorta, V. M., Perkell, J. S., and Guenther, F. H., (**2007**). "Sensorimotor adaptation to feedback perturbations of vowel acoustics and its relation to perception," J. Acoust. Soc. Am. **122**, 2306–2319.
Washburn, M. F. (**1926**). "Gestalt psychology and motor psychology," Am. J. Psychol. **37**, 516–520.
Watkins, K. E., Strafella, A. P., and Paus, T., (**2003**). "Seeing and hearing speech excites the motor system involved in speech production," Neuropsychologia **41**, 989–994.
Wilson, S. M., Saygin, A. P., Sereno, M. I., and Iacoboni, M., (**2004**). "Listening to speech activates motor areas involved in speech production," Nat. Neurosci. **7**, 701–702.

The interaction of vocal characteristics and audibility in the recognition of concurrent syllables[a)]

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

Nicholas R. C. Fyson
Centre for the Neural Basis of Hearing, Department of Physiology, Development and Neuroscience, University of Cambridge, Downing Street, Cambridge CB2 3EG, United Kingdom and Bristol Centre for Complexity Science, Department of Engineering Mathematics, University of Bristol, Queen's Building, University Walk, Bristol BS8 1TR, United Kingdom

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

(Received 26 March 2008; revised 19 November 2008; accepted 24 November 2008)

In concurrent-speech recognition, performance is enhanced when either the glottal pulse rate (GPR) or the vocal tract length (VTL) of the target speaker differs from that of the distracter, but relatively little is known about the trading relationship between the two variables, or how they interact with other cues such as signal-to-noise ratio (SNR). This paper presents a study in which listeners were asked to identify a target syllable in the presence of a distracter syllable, with carefully matched temporal envelopes. The syllables varied in GPR and VTL over a large range, and they were presented at different SNRs. The results showed that performance is particularly sensitive to the combination of GPR and VTL when the SNR is 0 dB. Equal-performance contours showed that when there are no other cues, a two-semitone difference in GPR produced the same advantage in performance as a 20% difference in VTL. This corresponds to a trading relationship between GPR and VTL of 1.6. The results illustrate that the auditory system can use any combination of differences in GPR, VTL, and SNR to segregate competing speech signals.
 [DOI: 10.1121/1.3050321]

PACS number(s): 43.71.An, 43.71.Bp, 43.66.Ba [JES] Pages: 1114–1124

I. INTRODUCTION

In multispeaker environments, listeners need to attend selectively to a target speaker in order to segregate their speech from distracting speech sounds uttered by other speakers. This paper is concerned with two speaker-specific acoustic cues that listeners use to segregate concurrent speech: glottal pulse rate (GPR) and vocal tract length (VTL). GPR and VTL are prominent markers of a speaker's size and sex in adults (Peterson and Barney, 1952; Fitch and Giedd, 1999; Lee *et al.*, 1999). Brungart (2001) reported an intriguing concurrent-speech study in which the target and distracting speakers were either the same sex or different sexes. He showed that listeners were better at identifying concurrent speech when the competing voice was from the opposite sex, but it was not possible to determine whether it was the GPR or the VTL difference that was more valuable. Darwin *et al.* (2003) reported separate effects of GPR and VTL on target recognition, and showed that there is additional improvement when the target and distracter differ in both vocal characteristics.

In these and other studies, where the speech stimuli are sentences, it is not possible to match the temporal envelopes of the competing speech sounds. As a result, listeners can monitor the concurrent speech for clean segments of target speech, and string the sequence of temporal "glimpses" together to improve performance (Cooke, 2006). To the extent that temporal glimpsing is successful, it reduces the listener's dependence on speaker differences, which in turn reduces the sensitivity of the experiment to the effects of speaker differences. The purpose of the current study was to enhance sensitivity to the value of speaker differences with a concurrent-syllable paradigm, in which the temporal envelope of the speech from the target is matched to the envelope of the distracter in an effort to reduce the potential for temporal glimpsing. Moreover, the domain of voices was extended to include combinations where the GPR was either unusually large or unusually small relative to the VTL, and vice versa.

A. Acoustic cues for segregating concurrent speech

The most significant cue for speech segregation is almost undoubtedly audibility, which is determined by the signal-to-noise ratio (SNR). When the speech material is sentences and they are matched for overall SNR, there are,

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

nevertheless, momentary fluctuations in SNR that allow the listener to hear the target clearly. Miller and Licklider (1950) reported that listeners are capable of detecting segments of the target speech during relatively short minima of a competing temporally fluctuating background noise. Cooke (2006) used a missing-data technique to model the effect of temporal glimpsing, and concluded that it can account for the intelligibility of speech in a wide range of energetic masking conditions.

Vocal characteristics such as GPR and VTL also provide cues that support segregation of competing speech signals. GPR is heard as voice pitch, and a number of studies have demonstrated that performance on a concurrent-speech task increases with the pitch difference between the voices up to about four semitones (STs) (e.g., Chalikia and Bregman, 1993; Qin and Oxenham, 2005; Assmann and Summerfield, 1990, 1994; Culling and Darwin, 1993). The experiments reported in this paper focus on the effects of VTL differences on concurrent syllable recognition, and how VTL differences interact with differences in GPR. The syllable envelopes were matched to minimize temporal glimpsing, and the relative level of the target and distracter was varied across the range where they interact strongly, in order to determine whether SNR affects the form of the interaction between GPR and VTL in segregation.

B. The role of vocal characteristics in speech segregation

The role of pitch in concurrent speech has been investigated in many psychophysical studies, and in many cases the pitch is specified in terms of the fundamental frequency (F0) of the harmonic series associated with the GPR. Chalikia and Bregman (1993) used concurrent vowels to show that a difference in F0 leads to better recognition of both vowels. Furthermore, they showed that a difference in F0 contour can lead to better recognition in situations where the harmonicity of the constituents is reduced. Assmann and Summerfield (1994) used concurrent vowels to show that small departures, from otherwise constant F0 tracks, can improve vowel recognition, especially when the F0 difference is small. Qin and Oxenham (2005) used concurrent vowels to show that performance reached its maximum when the difference in F0 was about four STs. They also found that when the spectral envelope was smeared with a channel vocoder, an F0 difference no longer improved vowel recognition. Summerfield and Assmann (1991) argued that the advantage of an F0 difference derives from the difference in pitch *per se* and not from the difference in spectral sampling of the formant frequencies, or glottal pulse asynchrony. In a series of related experiments, de Cheveigné and co-workers developed a harmonic cancellation model tuned to the periodicity of the distracter (de Cheveigné *et al.*, 1997b, 1997a; de Cheveigné, 1997, 1993). They showed that the advantage of an F0 difference in double-vowel recognition depends primarily on the harmonicity of the distracter. Culling and Darwin (1993) showed that when F0 tracks of concurrent speech are ambiguous, listeners can use the formant movements of competing diphthongs to disambiguate concurrent speech. They showed that listeners were only able to judge whether the F0 tracks of the speakers crossed or bounced off each other when the constituents had different patterns of formant movement.

When the F0 difference is small or the pitch is otherwise ill-defined, listeners have to use other acoustic cues to segregate concurrent speech. Brungart (2001) used noise and speakers of different sex as distracters in a concurrent-speech experiment. The distracting speech came from (a) the same speaker, (b) a different speaker of the same sex, or (c) a speaker of the opposite sex. Performance was measured with the coordinate response measure (CRM) that consists of sentences in the following form: "Ready */call-sign/* go to */color/ /number/* now," where the call-sign is a name like Charlie or Ringo (Bolia *et al.*, 2000; Moore, 1981). Brungart (2001) found that the psychometric functions for noise and speech distracters had different shapes. A clear performance advantage was observed when the distracter was a different speaker from the target, and the biggest advantage arose when the distracter was of a different sex. For all of the speech distracters, the worst performance was at 0 dB SNR; at negative SNR, performance recovered somewhat. For noise maskers, performance was better overall, and no recovery was found for negative SNR. The results were interpreted in terms of informational masking, and they suggest that voices are more distracting than noise even when the noise is modulated. However, the release from masking at negative SNR only occurred for identification of the color coordinate in the target sentence. Brungart (2001) speculated that this might be because the numbers appear last in the CRM sentences and so might not overlap in time as much as the color coordinates. Thus, more temporal glimpsing was possible with the numbers than with the colors. This might explain the morphological difference between the psychometric functions for color and number. Glimpsing could also explain the relative inefficiency of the noise masker in that study. In Brungart *et al.* (2001), similar results were found for multiple distracting speakers except for the recovery phenomenon, indicating that it probably has to do with the extent to which words overlap in CRM.

Darwin *et al.* (2003) investigated the effects of F0 and VTL in a study on concurrent speech using the CRM corpus. They used a pitch-synchronous overlap-add technique as implemented in PRAAT to separately manipulate F0 and VTL in a set of sentences and produced a range of speakers with natural combinations of F0 and VTL. For an F0 difference of 12 STs, at 0 dB SNR, they reported an increase in speech recognition of 28%, most of which (~20%) was already apparent at an F0 difference of four STs (see Fig. 1 in Darwin *et al.*, 2003). They also found that individual differences in intonation can help identify speech of similar F0, corroborating the findings of Assmann and Summerfield (1994) mentioned above. For a 38% change in VTL, Darwin *et al.* (2003) reported an increase in recognition of ~20% at 0 dB SNR (see Fig. 6 in Darwin *et al.*, 2003). The largest performance increase was found for a combined difference in GPR and VTL, and they concluded that F0 and VTL interact in a synergistic manner. However, a large asymmetry was reported with regard to the effect of VTL. When the VTL of the target was smaller than the VTL of the distracter, the

effect was much larger than when the VTL of the target was larger than the VTL of the distracter (by the same relative amount). As in the study of Brungart (2001), they made no attempt to control glimpsing.

In two related studies involving a lexical decision task, Rivenez *et al.* (2006, 2007) showed that differences in both F0 and VTL between two competing voices presented dichotically facilitated the use of priming cues in an unattended contralateral signal. In both their studies, an advantage was observed in terms of faster response time to the target stimuli, and the results were interpreted to lend support to the notion that early perceptual separation of the competing voices is a necessary prerequisite for lexical processing of the unattended voice.

The results described above support the hypothesis, originally proposed by Ladefoged and Broadbent (1957), that listeners construct a model of the target and distracting speakers, and that they use speaker-specific acoustic cues such as VTL and GPR as part of the model. Smith and Patterson (2005) showed that listeners can judge the relative size/age, and the sex of a speaker based on their vowels even when the GPR and VTL values were well beyond the range of normal speech. Collins (2000) showed that female listeners can make accurate judgments about the weights of male speakers based solely on their voices. To recognize a speech sound, the listener needs to normalize for phonetically irrelevant speaker-dependent acoustic variability (Nearey, 1989) such as that associated with variation in VTL and GPR. For unmasked speech, this is a trivial task for most listeners, and it leads to remarkable robustness in speech perception to speaker differences (Smith *et al.*, 2005). Smith *et al.* (2005) pointed out that the robustness of speech recognition is unlikely to be entirely due to learning, since speakers with highly unusual combinations of GPR and VTL are understood almost as well as those with the most common combinations of GPR and VTL. This is compatible with the suggestion that the auditory system operates early preprocessing stages that detect and normalize for GPR and VTL (Irino and Patterson, 2002). The preprocessing is applied to all sounds and is hypothesized to operate irrespective of attention and task relevance. In the current paper, it is argued that listeners use GPR and VTL in their speaker models and that this facilitates the segregation of competing voices. We hypothesize that the reason why it is possible to attend selectively to a particular speaker in a multispeaker environment is that the processing of VTL and GPR cues is automatic and occurs at an early point in the hierarchy of speech processing.

In natural speech, speakers vary GPR by changing the tension of the vocal folds, and they use GPR to convey prosody information within a range determined largely by the anatomical constitution of the laryngeal structures (Titze, 1989; Fant, 1970). By contrast, it is only possible to change the VTL by a small amount, either by pursing the lips or by lowering or raising the larynx, which require training, and both of which produce an audible change to the quality of the voice. The relative stability of the VTL cue suggests that VTL is likely to be at least as important for tracking a target speaker as GPR.

The purpose of the current study was to investigate the relative contribution of GPR and VTL in the recognition of concurrent speech, while carefully controlling other potential cues. The aims were (1) to quantify the effects of VTL and GPR, and (2) to model the relationship between them.

II. METHOD

The participants were required to identify syllables spoken by a target speaker in the presence of a distracting speaker. Performance was measured as a function of the difference between the target and distracting speakers along three dimensions: GPR, VTL, and SNR. In order to prevent the listeners from taking advantage of temporal glimpses, the temporal envelopes of the target and distracter syllables were carefully matched, as described below.

A. Listeners

Six native English speaking adults participated in the study (four males and two females). Their average age was 21 years (19–22 years), and no subject had any history of audiological disorder. An audiogram was recorded at the standard audiometric frequencies to ensure that the participants had normal hearing. The experiments were done after informed consent was obtained from the participants. The experimental protocol was approved by the Cambridge Psychology Research Ethics Committee.

B. Stimuli

The experiments were based on the syllable corpus previously described by Ives *et al.* (2005) and von Kriegstein *et al.* (2006). It consists of 180 spoken syllables, divided into consonant-vowel (CV) and vowel-consonant (VC) pairs. There were 18 consonants, 6 of each of 3 *categories* (plosives, sonorants,[1] and fricatives), and each of the consonants was paired with 1 of the 5 vowels spoken in both CV and VC combinations. The syllables were analyzed and resynthesized with a vocoder (Kawahara and Irino, 2004) to simulate speakers with different combinations of GPR and VTL. Since all the voices were synthesized from a recording of a single speaker (Patterson), the only cues available for perceptual separation were the GPR and VTL differences introduced by the vocoder. Throughout the experiment the target voice was presented at 60 dB SPL, while the RMS level of the distracting voice varied to achieve an SNR of +6, 0, or −6 dB.

1. Vocal characteristics

The voice of the target (the reference voice) remained constant throughout the experiments, and its characteristics were chosen with reference to typical male and female voices. Peterson and Barney (1952) reported that the average GPRs of men and women are 132 and 223 Hz, respectively, and Fitch and Giedd (1999) reported that the average VTLs of men and women are 155 and 139 mm, respectively. The geometric means of these values were used to simulate an androgynous target speaker with a GPR of 172 Hz and a VTL of 147 mm. The VTL of the original speaker was estimated to be 165 mm, and this value was used as a reference

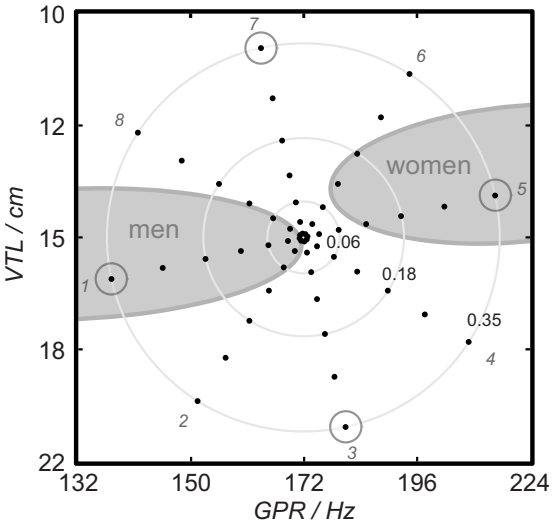


FIG. 1. The specification of the distracter voices forms an elliptical spoke pattern in the GPR-VTL plane. The ordinate has been compressed by a factor of 1.5 to illustrate the relationship between GPR and VTL used in the experiment. It is this, that makes the ellipses appear circular. The pitch of the distracter voices varied from 137 to 215 Hz, and the VTL varied from 11 to 21 cm. The target voice is in the center of the spoke pattern with a pitch of 172 Hz and a VTL of 15 cm. In the text, points in the plane are described in terms of their radial displacement from the target (a vector distance). For convenience, they are arbitrarily numbered 1–7 where 1 is closest to, and 7 is farthest away from, the reference voice in the center. The $RSD_{1.5}$ values for spoke points 3, 5, and 7 are shown; see Table I for the complete specification. Some of the distracter voices were used in the extended SNR experiment; they are marked with circles. The gray areas show the region of the plane occupied by 95% of male and female speakers in the population reported by Peterson and Barney (1952), and modeled by Turner *et al.* (2009).

to rescale the original recordings in order to produce the voices used in the experiment. The combinations of GPR and VTL chosen for the distracter are shown by the dots in Fig. 1, which form an elliptical spoke pattern radiating out across the GPR-VTL plane from the reference voice. The ellipse that joins the ends of the spokes had a radius of 26% (4 STs=$2^{4/12}$) along the GPR axis and 41% (6 STs=$2^{6/12}$) along the VTL axis. The VTL dimension is proportionately longer because the just noticeable difference (JND) for VTL is at least 1.5 times the JND for GPR (Ives *et al.*, 2005; Ritsma and Hoekstra, 1974). There were seven points along each spoke, spaced logarithmically in this logGPR-logVTL plane in order to sample the region near the target speaker with greater resolution. Spokes 1 and 5 were tilted by 12.4° to form a line joining the average man with the average woman (Turner *et al.*, 2009; Peterson and Barney, 1952; Fitch and Giedd, 1999). The tilt ensured that there was always variation in both GPR and VTL between the target and distracter voices, which reduces the chance of the listener focusing on one of the dimensions to the exclusion of the other (Walters *et al.*, 2008). In all, there were 56 different distracter voices with the vocal characteristics shown in Table I.

An assumption underlying the design is that there is a trading relationship between VTL and GPR in the perceptual separation between the target voice and any distracter voice, and that the perceptual distance between voices can be expressed by the radial scale displacement (RSD) between their points in the logGPR-logVTL plane. The RSD is the

TABLE I. Vocal characteristics of the distracter voice for the spoke numbers shown in Fig. 1. Spoke point refers to the position on the spokes ascending outwards from the reference voice in the center. The radial scale displacement ($RSD_{1.5}$) is shown for each spoke point; see text for details.

Spoke No.	Spoke point	1	2	3	4	5	6	7
	$RSD_{1.5}$	0.01	0.03	0.06	0.11	0.18	0.25	0.35
1	GPR (Hz)	170.9	168.6	164.7	159.5	153.0	145.5	137.0
	VTL (cm)	14.7	14.8	14.9	15.1	15.3	15.5	15.8
2	GPR (Hz)	171.3	170.0	167.8	164.9	161.1	156.7	151.6
	VTL (cm)	14.8	15.0	15.5	16.2	17.0	18.2	19.7
3	GPR (Hz)	171.9	172.4	173.3	174.5	176.1	178.1	180.4
	VTL (cm)	14.8	15.1	15.6	16.4	17.5	18.8	20.6
4	GPR (Hz)	172.4	174.5	178.0	183.0	189.6	198.1	208.6
	VTL (cm)	14.7	14.9	15.2	15.6	16.2	16.8	17.7
5	GPR (Hz)	172.5	174.9	179.0	184.8	192.7	202.7	215.2
	VTL (cm)	14.7	14.6	14.5	14.3	14.1	13.9	13.6
6	GPR (Hz)	172.1	173.5	175.7	178.8	183.0	188.1	194.5
	VTL (cm)	14.6	14.3	13.9	13.4	12.7	11.9	11.0
7	GPR (Hz)	171.5	171.0	170.1	168.9	167.4	165.6	163.4
	VTL (cm)	14.6	14.3	13.8	13.2	12.4	11.5	10.5
8	GPR (Hz)	171.0	169.0	165.7	161.1	155.5	148.8	141.3
	VTL (cm)	14.6	14.5	14.2	13.8	13.4	12.8	12.2

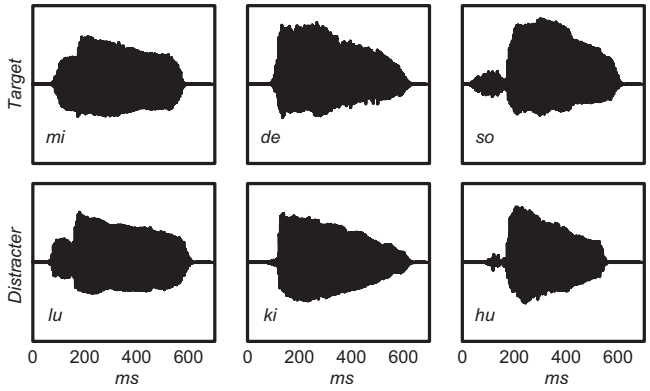


FIG. 2. Examples of temporal-envelope matching pairs of sonorant-vowel syllables (left column), plosive-vowel syllables (central column), and fricative-vowel syllables (right column). The syllables are *p*-centered to effect the optimum match for arbitrary syllable pairings within a CV group. See the text for details.

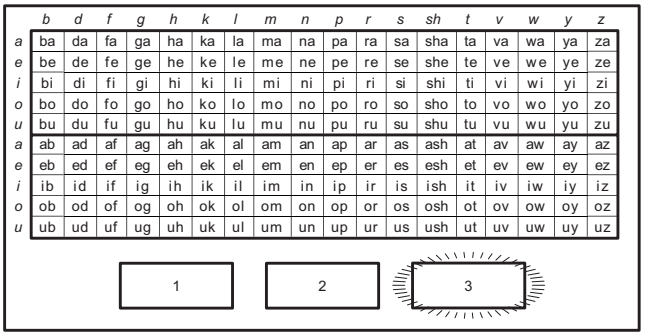


FIG. 3. Schematic illustration of the GUI. The response area, in which the listeners indicate their answer by a click with a computer mouse, is shown at the top. Underneath are shown the three visual interval indicators that light up synchronous with the three syllable intervals of each trial. In this example, interval 3 stays lit to indicate that the listeners should respond to the target syllable that was played in interval 3.

geometrical distance between the target and distracting voices,

$$\mathrm{RSD}_{\chi} = \sqrt{\chi^2(X_{\mathrm{target}} - X_{\mathrm{distracter}})^2 + (Y_{\mathrm{target}} - Y_{\mathrm{distracter}})^2}, \qquad (1)$$

where X is log(GPR), Y is log(VTL), and χ is the GPR-VTL trading value that is 1.5 in the design. The RSD values shown in Fig. 1 are for $\chi = 1.5$. The design only requires the trading value to be roughly correct; so long as the voices vary in combination over ranges that go from indistinguishable to readily distinguishable in all directions, then the optimum trading value can be estimated from the recognition data.

2. Envelope control

A combination of techniques was employed to limit temporal glimpsing. First, the perceptual centers (Marcus, 1981) of the syllables were aligned as described by Ives *et al.* (2005). Second, the target and distracter syllables were matched according to their phonetic specification in the following way: (1) the CV order of the target and distracter syllables was required to be the same, and (2) the consonants in a concurrent pair of target and distracter syllables were from the same category. The result of these manipulations is that the temporal envelopes of the target and distracter syllables were closely aligned and similar in shape, as illustrated in Fig. 2. Within the six categories of syllables, pairs of target and distracter syllables were chosen at random with the restriction that the pair did not contain either the same consonant or the same vowel. These restrictions leave 20 potential distracter syllables for each target syllable.

C. Procedure

The study consisted of three parts: (1) pre-experimental training, (2) the main experiment, and (3) an SNR extension. The procedure was the same in all three: The target syllables were presented in triplets; the first syllable (the precursor) was intended to provide the listener with cues to the GPR and VTL of the target speaker; the second and third target syllables were presented with a concurrent distracter syllable. The three intervals of each trial were marked by visual indicator boxes on the graphical user interface (GUI) illustrated in Fig. 3. After the third interval was complete, the box for interval 2 or 3—chosen at random—was illuminated to indicate which of the two target syllables the listener was required to identify. The multisyllable format was intended to promote perception of the stimuli as speech and the sense that there were two speakers, one of whom was the target. The format makes it possible to vary the task difficulty in order to maximize the sensitivity of the recognition scores to the variation in GPR and VTL. In the present implementation, the trial includes *two* concurrent syllable pairs after the precursor, because pilot work showed that the task was too easy with just one concurrent syllable pair, and that the memory load was too great with three or more concurrent syllable pairs.

Listeners indicated their answers by clicking on the orthographical representation of their chosen syllable in the response grid on the screen. The participants were seated in front of the response screen in a double-walled IAC (Winchester, UK) sound-attenuated booth, and the stimuli were presented via AKG K240DF headphones.

1. Pre-experimental training

The ambiguity of English orthography meant that the response grid shown in Fig. 3 required some introduction. Listeners had to learn that the notation for the vowels was like that in the majority of European languages; that is, "a" is [ɑː], "e" is [ɛː], "i" is [iː], "o" is [oː] and "u" is [uː]. Moreover, they had to learn to find the response syllables on the grid rapidly and with confidence. In the first training session, target syllables without distracters were presented to the listeners who were instructed to identify the syllable in interval 3. This training comprised 15 runs with visual feedback. Each run was limited to a subset of the syllable database in order to gradually introduce the stimuli and their orthography. Then followed 380 trials without distracters in which the target syllable was in either interval 2 or interval 3. In addition, the visual feedback was removed to demonstrate that the listeners could operate the GUI and that the complexity of the GUI was not the basis of subsequent errors. In

a second training session, the distracters were progressively introduced with the SNR starting at 15 dB and then gradually decreased to −9 dB. In this session, the target syllable was in either interval 2 or interval 3. There were 10 runs of 48 trials in this session with visual feedback.

During training, performance criteria were used to ensure that the listener could perform the task at each SNR before proceeding to the next level. The pass marks were set based on pilot work that had shown that the performance of most listeners rose rapidly to more than 80% correct for unmasked syllables, and more than 30% correct when distracters were included at −9 dB SNR. Consequently, the criterion for the initial training session with targets but no distracters was 80%, and the criterion for the second training session with targets and distracters decreased from 70% to 30% as the SNR decreased from 15 to −9 dB. If a listener did not meet the criterion on a particular run, it was repeated until performance reached criterion. In total, the listeners did, at least, 1050 trials before commencing the main experiment, by which time they were very familiar with the GUI.

2. Main experiment

In the main experiment, recognition performance was measured as a function of GPR, VTL, and SNR for the CV syllables only. There were 56 different distracter voices (cf. Fig. 1 and Table I) and 3 SNRs (−6, 0, and 6 dB). The RSD between the target and distracter voices was varied over trials in a consistent fashion, from large to small and back to large. In this way, the distracter voice cycled through spoke points 7 to 1 and back to 7 so that the task became progressively harder and then easier in an alternating way. This was done to ensure that the listener was never subjected to a long sequence of difficult trials with small RSD values. The conditions at the ends of the RSD dimension were not repeated as the oscillation proceeded, so one complete cycle contained 12 trials. The main experimental variable—the combination of GPR and VTL in the distracter voice—was varied randomly without replacement from the eight values with the current RSD value (one on each spoke). The main experiment consisted of 120 runs of 48 trials (four cycles as explained above). Between runs, the SNR cycled through the three SNR values. The combination of RSD oscillation and controlled spoke randomization meant that when all runs had been completed, all of the RSD values, other than the endpoints, had been sampled 40 times at each SNR, and the endpoints had been sampled 20 times.

3. SNR extension

The final part of the study measured the psychometric function for six distracters at SNR values of −15, −9, 0, +9, and +15 dB using an up-down procedure. This was done in order to demonstrate that the effect of RSD decreases as SNR moves out of the range (−6 to +6 dB) used in the main experiment because performance becomes dominated by energetic masking in one way or another. Above +6 dB, the distracter is becoming ever less audible, and below −6 dB, the target is becoming ever less audible. At the same time, the SNR extension ensures that the main experiment was performed in the region where the paradigm was most sensitive to the interaction of GPR, VTL, and SNR.

Four of the distracters were the outermost points of spokes 1, 3, 5, and 7, marked with circles in Fig. 1. The remaining two distracters were the target voice itself and a noise masker. The noise maskers were created by extracting the temporal envelopes of distracters chosen in the usual way, and then filling the envelopes with speech-shaped broadband noise (Elberling *et al.*, 1989). Randomization was performed so that after 24 runs of 40 trials each, all midpoints in the SNR range had been measured 40 times for each distracter.

III. RESULTS

Two measures of performance were analyzed: (1) target syllable recognition rate, and (2) distracter intrusion rate, where the listener reported the distracter rather than the target syllable. The effects of vocal characteristics and audibility on these target and distracter scores in the main experiment were analyzed with a three-way repeated-measures analysis of variance (ANOVA) [3(SNRs) × 8(spoke numbers) × 7(spoke points)]. In the SNR extension, the effects of vocal characteristics and audibility were analyzed with a two-way repeated-measures ANOVA for the four different voices [5(SNRs) × 4(spoke numbers)]. Paired comparisons with Sidak correction for multiple comparisons were used to analyze the effects of all six distracters (four different voices, one identical voice, and one noise masker). Greenhouse–Geisser correction for degrees of freedom was used to compensate for lack of sphericity, and partial eta squared values (η_p^2) are quoted below to report the effect sizes.

For target recognition in the main experiment, there were significant main effects of SNR ($F_{2,10}=201.4$, $p<0.001$, $\varepsilon=0.64$, $\eta_p^2=0.98$) and spoke point ($F_{6,30}=123.2$, $p<0.001$, $\varepsilon=0.47$, $\eta_p^2=0.96$), and an interaction between SNR and spoke point ($F_{12,60}=5.4$, $p=0.004$, $\varepsilon=0.33$, $\eta_p^2=0.52$). There was no main effect of spoke number and no interaction between spoke number and either SNR or spoke point. For the distracter intrusions, there was a similar pattern of results: There were the same significant main effects of SNR ($F_{2,10}=50.2$, $p=0.001$, $\varepsilon=0.54$, $\eta_p^2=0.91$) and spoke point ($F_{6,30}=168.1$, $p<0.001$, $\varepsilon=0.47$, $\eta_p^2=0.96$), and the same interaction between SNR and spoke point ($F_{12,60}=26.2$, $p<0.001$, $\varepsilon=0.19$, $\eta_p^2=0.84$). There was no main effect of spoke number and no interaction between spoke number and either SNR or spoke point. Notice that it is spoke *point* that indicates the radial distance between voices, whereas spoke *number* specifies the angular relationship between sets of voices in the VTL-GPR plane. For target recognition in the SNR extension, there was again a significant main effect of SNR ($F_{4,10}=114.5$, $p<0.001$, $\varepsilon=0.53$, $\eta_p^2=0.96$). There was no main effect of distracter voice and no interaction between distracter voice and SNR. For distracter intrusions, there was a similar pattern of results: a significant main effect of SNR ($F_{4,10}=7.0$, $p=0.011$, $\varepsilon=0.54$, $\eta_p^2=0.58$), no main effect of distracter voice, and no interaction between distracter voice and SNR.

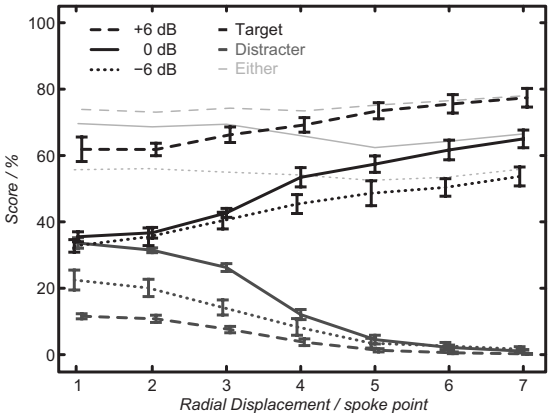


FIG. 4. Recognition scores as a function of spoke point (see Fig. 1). The black lines at the top show target recognition scores, and the dark gray lines in the lower part of the figure show distracter intrusion scores. The sums of the target and distracter scores for each SNR are shown in the lightest gray lines at the top of the figure. They are largely independent of spoke point.

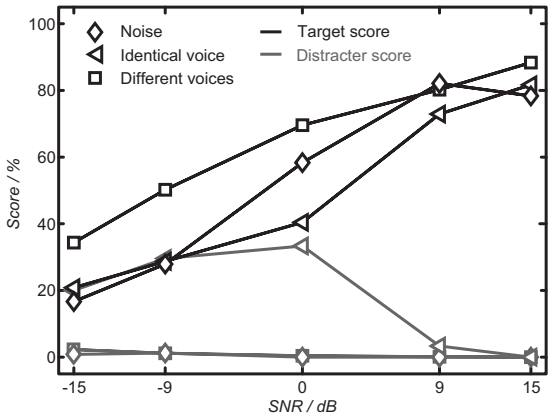


FIG. 5. Psychometric functions for the speech distracters and the noise masker. The black lines at the top show target recognition scores, and the dark gray lines in the lower part of the figure show distracter intrusion scores. Speech distracters for the different voices are shown with squares, the speech distracter with identical vocal characteristics is shown with triangles, and the noise masker is shown with diamonds.

A. Effects of vocal characteristics

The fact that there was no main effect of spoke number, and no interaction of spoke number with the other variables, for either of the performance measures, means that the best estimate of the main effect of RSD is provided by collapsing performance across spoke number. The results are shown in Fig. 4, where average performance is plotted as a function of spoke point. The target recognition scores are shown by dashed, solid, and dotted black lines in the upper part of the figure. Performance is best, as would be expected, for conditions where the target and distracter voices are maximally dissimilar (spoke point 7). The effect of vocal characteristics (i.e., spoke point) is greatest for 0 dB SNR, where the cumulative effect along the spoke-point function is 29%. For the +6 and −6 dB SNRs, the cumulative effect is somewhat smaller, 15% and 22%, respectively. The convergence of the 0 dB function with the −6 dB function as spoke point decreases between points 4 and 3 shows that listeners can use the level difference in the −6 dB condition to prevent performance falling as low as it might when there is little, or no, difference in vocal characteristics.

The distracter intrusion scores are shown by the dashed, solid, and dotted dark gray lines in the lower part of Fig. 4. Distracter intrusions occur most often, as would be expected, when the voices are similar (spoke point 1). When the voices are dissimilar (spoke points 5–7), distracter intrusions are very rare. The intrusion rate for similar voices (point 1) is greatest (34%) when the SNR is 0 dB, and least (12%) when the SNR is +6 dB. The fact that the intrusion rate is higher at 0 dB SNR than it is at −6 dB SNR (22%) shows that listeners derive a consistent advantage from a loudness difference even when the distracter is louder than the target.

The sum of the target and distracter scores is presented in the upper part of Fig. 4, separately for each SNR, by thin, light gray lines, using the same line type. These lines exhibit no effect of vocal characteristics; the only difference between them is their vertical position, which reflects the effect of SNR. This means that for a given SNR, the listeners identified the syllables in a pair at a roughly constant level, but they were only successful at segregating the voices when there was a perceptible difference in vocal characteristics.

B. Effect of SNR

An extended view of the effect of SNR is presented in Fig. 5, which shows performance as a function of SNR from −15 to +15 dB, for speech distracters with different voices (lines with squares), a speech distracter with identical vocal characteristics (lines with triangles), and a noise masker (lines with diamonds). The speech distracters with different voices were the four speakers farthest from the target voice on spokes 1, 3, 5, and 7. The black lines in the upper part of the figure show average target recognition, which drops from between 78% and 88% at 15 dB SNR to between 17% and 34% at −15 dB SNR; the largest drop is seen for the noise masker. Performance with speech distracters that differ in vocal characteristics from the target is always better than performance when the distracter has the same vocal characteristics, even when the SNR is large. The gradient is steepest for the noise masker between 0 and −9 dB, and steepest for the identical voice between +9 and 0 dB. In these regions performance drops about 3 % / dB.

With the exception of the case where the distracter was identical to the target, recognition performance for speech distracters did not vary with the spoke of the distracter speaker, that is, the angle of the spoke in the GPR-VTL plane. Paired comparisons showed that the target scores for all of the different voices were significantly higher than the target scores for the noise masker for the three lowest SNRs, and lower than the target scores for the identical voice for −9 and 0 dB SNRs. The target scores for the noise masker were significantly higher than the target scores for the identical voice for 0 and 9 dB SNRs.

The distracter scores are shown in the lower part of Fig. 5 with gray lines. Listeners made this type of error only if the target and distracter voices were identical. As the SNR decreased from 15 dB, the proportion of distracter intrusions increased to a maximum (33%) at 0 dB SNR where there is no loudness cue to distinguish the target from the distracter.

The rate of distracter intrusions then decreased from 33% at 0 dB SNR to 20% at −15 dB SNR. Paired comparisons of the distracter scores for the identical voice showed that the maximal distracter intrusion rate at 0 dB SNR was significantly higher than the distracter intrusion rates at −15, 9, and 15 dB SNRs. These results confirm that the listeners were able to attend selectively to the target when the only difference between the target and distracter voices was loudness, even in adverse listening conditions where the distracter was much louder than the target.

IV. MODELING THE INTERACTION OF VOCAL CHARACTERISTICS AND AUDIBILITY

The fact that there is no main effect of spoke angle, and no interaction between spoke angle and either spoke point or SNR, suggests that it might be possible to produce a relatively simple model of the target recognition performance in this experiment. Such a model would provide an economical summary of the data and enable us to provide performance surfaces to illustrate how speaker differences interact with SNR differences in the performance space. The model would also allow us to test the assumption made when generating the stimuli that a difference in VTL has to be about 1.5 times a difference in GPR in order to produce the same effect on recognition performance.

The psychometric function that relates target recognition to RSD (along a spoke) should be asymmetric because the RSD variable is limited to positive values, by definition, and the slope of the psychometric function should, therefore, be 0 for an RSD of 0. This means that the traditional cumulative Gaussian would not provide a good fit in this experiment. Accordingly, the psychometric function was modeled with a cumulative gamma function rather than a cumulative Gaussian. In this way, the probability of target recognition as a function of RSD was defined as

$$p(\mathrm{RSD}) = \lambda + (\mu - \lambda)\int_0^{\mathrm{RSD}} \Gamma(x|\alpha,\beta)dx, \qquad (2)$$

where Γ is the gamma function and α and β are its shape and scale parameters. The gamma function rises from 0 to 1 over its range, so the function was offset and scaled on the p-axis by λ and μ, where λ is the intercept on the p-axis, and μ is the asymptotic recognition score for large RSD. The shape parameter, α, was restricted to be greater than 2 to ensure that the cumulative gamma function would have a gradient of zero at the p-axis intercept. The other parameters were limited only by their theoretical maximum range.

The psychometric function was fitted to the data using maximum likelihood estimation (MLE), in which the trading value between VTL and GPR, χ, was included as a free parameter whose optimum value was estimated in the process. Specifically, MLE was used to fit the cumulative gamma function to all of the individual sets of data, and between iterations, the value of χ was varied (that is, the relative lengths of the GPR and VTL dimensions were varied) to find the value of χ that was most likely to have produced the observed data. In one case, the procedure was applied separately to the three groups of eight data sets associated with each SNR (+6, 0, and −6 dB), and in a second case, the procedure was applied to the collective set of data from the three SNRs taken together. In this way, a trading value was derived for each SNR as well as an optimum trading value describing the trading relation for all SNRs. The psychometric functions for the collective fit are presented along with the data in Fig. 6; the optimum trading value for this collective fit was 1.9. Since the trading value is different from the value of 1.5 used to generate the stimuli, the points from different spokes occur at different $\mathrm{RSD}_{1.9}$ values. The trading values for the three SNRs, fitted individually, were 1.9, 1.6, and 3.2 for SNR values of 6, 0, and −6 dB, respectively.

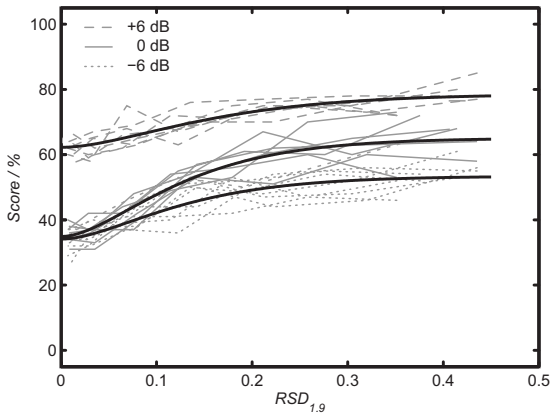


FIG. 6. Cumulative gamma functions fitted by MLE to the behavioral data associated with the eight stimulus spokes at each of the three SNR values are shown in thick black lines. The GPR-VTL trading parameter was allowed to vary along with the parameters describing the gamma functions. The optimized trading value was 1.9. Performance is plotted as a function of the RSD ($\mathrm{RSD}_{1.9}$) in logarithmic units. The mean data for the individual spokes are shown by the gray lines. The dashed lines are for an SNR of 6 dB; the solid lines are for 0 dB SNR and the dotted lines are for −6 dB SNR.

In order to illustrate the effects of the trading relationship between GPR and VTL, the trading value from the collective fit was used to generate performance surfaces, which are shown in Fig. 7. The surfaces are sets of elliptical, equal-performance contours fitted to the values from the eight spokes associated with each RSD value, separately for the three SNRs (+6, 0, and −6 dB). The dotted lines in the GPR-VTL plane below the surfaces show the combinations of GPR and VTL that defined the distracters. The surfaces show the main effects of SNR, GPR, and VTL on performance reported above. The main effect of SNR is illustrated by the displacement of the surfaces around their outer edges, where the radial distance between the target and distracter is maximum; the average performance for the largest speaker difference was 77%, 64%, and 52% when the SNR was +6, 0, and −6 dB, respectively. The effect of speaker difference is illustrated by the indentation at the center of each surface. In each case, as the radial distance between the target and distracter voice decreases, recognition performance decreases, and in each case, the worst performance occurs when there is minimal difference between the target and distracter in terms of GPR and VTL. The indentation in the surface is deepest for the 0 dB condition where the loudness cue is smallest. The

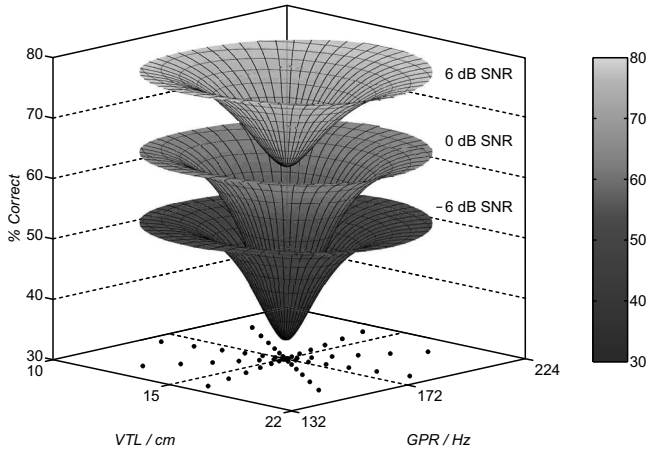


FIG. 7. Performance surfaces for −6, 0, and 6 dB SNRs from below to above, respectively. The tip of the middle surface for 0 dB SNR extends down almost to the tip of the surface for −6 dB below it. The vocal specifications of the distracters are indicated in the GPR-VTL plane below the surfaces. The surfaces were modeled by MLE of the psychometric functions along the spokes and by interpolating between the spokes; see the text for details.

average performance at the center of the surface drops 16% when the SNR is +6 dB, 29% when the SNR is 0 dB, and 20% when the SNR is −6 dB.

V. DISCUSSION

The main experiment showed that listeners take advantage of VTL differences as well as GPR differences when recognizing competing syllables. The effect is most notable when the SNR is 0 dB and there is no loudness cue to assist in tracking the target speaker. In noisy environments where extra vocal effort is required, normally hearing speakers tend to allow the level of their voice to fall to an SNR of around 0 dB (Lombard, 1911; Plomp, 1977). Hence, the 0 dB SNR condition of the current experiment represents a common condition in multispeaker environments, and the data show the value of the vocal characteristics in speaker segregation.

A. Trading relationship between VTL and GPR

In the current experiment, the trading values for the three SNRs were found to be 1.9, 1.6, and 3.2 for SNR values of 6, 0, and −6 dB, respectively. The values were determined by the minima of the cost functions for the MLE. However, the cost functions were highly asymmetric about their minima. They drop rapidly as χ increases from 0 toward their minimum, but beyond the minimum the cost functions rise slowly. The trough for the 0 dB cost function was the sharpest, which indicates that the trading value of 1.6 is the most reliable. The trough was least sharp for the −6 dB cost function, indicating that the trading value of 3.2 is the least reliable.

The fact that the smallest trading relationship is found for 0 dB SNR is consistent with the finding that the effects of vocal characteristics are largest when there are few other cues. Pitch is a very salient cue, so it may be that, as soon as there are both pitch and loudness cues, it requires a relatively large VTL difference to improve performance further. The trading value for 0 dB SNR, 1.6, means that a two-ST GPR difference provides the same performance advantage as a 20% difference in VTL.

Darwin *et al.* (2003) reported data on the interaction of VTL and F0 in their study on concurrent speech. At 0 dB SNR, for a 38% difference in VTL, the performance advantage was 20%, and for a four-ST difference in F0, the performance advantage was 20%.[2] This corresponds to a trading relationship of 1.4, just below that observed at 0 dB SNR in the current experiment. They also found that when the VTL and GPR values both shifted toward larger speakers, there was more benefit than when they both shifted toward smaller speakers. Given their results, one might have expected to see that the improvement in recognition performance along spoke 6 was greater than along spoke 2. Darwin *et al.* (2003) also reported that there was synergy between GPR and VTL; a result that could have led to the expectation that performance would increase less along spokes 1, 3, 5, and 7 (where there is mainly variation in one variable) than along spokes 2, 4, 6, and 8 (where there is comparable variation in both variables). However, we do not find any asymmetry in target recognition performance about the target speaker in the GPR-VTL plane. Our results show that there was no statistical effect of spoke number; the only effect of vocal characteristics was an effect of spoke point, i.e., the radial distance from the target speaker. At this point, it is not clear why Darwin *et al.* (2003) found asymmetries that we did not.

Van Dinther and Patterson (2006) varied the equivalent of GPR and VTL in musical instrument sounds and measured the amount of change in the variables (pulse rate and resonance scale) required to discriminate the relative size of instruments from their sustained note sounds. The design allowed them to estimate the trading value between the variables, which was found to be 1.3 with these sounds. This suggests that when listeners compare sounds with different combinations of pulse rate and resonance rate, pulse rate has a somewhat larger effect in judgments of relative instrument size.

B. Determinants of distraction

The present experiments show that the vocal characteristics of a competing speaker have a large effect on the amount of distraction caused by that speaker. This was evident in Fig. 5, which showed the psychometric functions for six different distracters. At 0 dB SNR, the noise masker reduced performance to 58%; but when the distracter was the identical voice, performance dropped further to 40%. This phenomenon (a drop in recognition performance without a change in SNR) is sometimes referred to as "informational masking" (for a recent review, see Watson, 2005). The idea is that the degree of disturbance depends not only on the distracter's ability to limit audibility (energetic masking), but also on its ability to pull attention away from an otherwise audible or partially audible target sound. In other words, informational masking is a property separate from energetic masking—a property which differentiates the effects of equally intense distracters. Note, however, that as the vocal characteristics of the distracting voice become more and

more different from those of the target voice in our study, performance rises above that achieved when the target voice is presented in broadband, envelope-matched noise. When the SNR is 0 dB, performance rises from 40% when the target and distracter are the same voice, to 58% when the distracter is a noise masker, and on up to 70% when the distracter is one of the different voices. Hence, the vocal characteristics of the distracter caused a drop in performance (relative to a noise masker of the same level) when the voices were similar, and an increase in performance when the voices were dissimilar.

The difference between masking and distraction is also evident in the distribution of errors, and the interaction of error type with SNR. When the voice of the distracter caused a drop in recognition performance, it was often because the listener reported the distracter syllable rather than the target syllable (the gray lines in Fig. 4). If these errors are scored as correct, performance is observed to be largely independent of speaker characteristics for a given SNR. In other words, the main effect of reducing the RSD between the target and distracter is that listeners are increasingly unsuccessful at segregating the syllable streams of the competing voices. Target-distracter confusions occurred most often for 0 dB SNR as illustrated in Fig. 5. In a recent study on the release of informational masking based on spatial separation of the competing sources, Ihlefeld and Shinn-Cunningham (2008) also reported that most target-distracter confusions occurred for 0 dB SNR.

For conditions where the SNR was not zero, listeners appear to derive some value from loudness cues, even when the SNR is negative. The depth of the indentation in the performance surface (Fig. 7) for data with an SNR of −6 dB is less than for an SNR of 0 dB. A different view of this effect is provided in Fig. 4 in which performance drops off most for an SNR of 0 dB as RSD decreases. The improvement in relative performance at negative values of SNR is similar to, although less pronounced than, the recovery phenomenon reported by Brungart (2001) at negative SNRs mentioned earlier. The data suggest that listeners can use loudness to reject the distracter and focus on the target even when it is softer than the distracter.

The depth of the indentation in the performance surface for data with an SNR of 0 dB confirms the importance of differences in vocal characteristics when there is no loudness cue. It is in this condition that listeners are observed to derive the most advantage from differences in VTL, GPR, or any combination of the two. A similar result was reported by Ihlefeld and Shinn-Cunningham (2008). They showed that recognition performance was poorer for an SNR of 0 dB than it was for an SNR of −10 dB, and that the largest relative advantage provided by spatial separation occurred when the SNR was 0 dB.

VI. SUMMARY AND CONCLUSION

The experiments in this paper show how two speaker-specific properties of speech (GPR and VTL) assist a listener in segregating competing speech signals. In multispeaker environments, where there are substantial differences between speakers in GPR and VTL, the performance for a particular SNR depends critically on these speaker differences. When they are not available, target recognition is severely reduced as evidenced by the indentations in the performance surfaces in Fig. 7. The results also show that, when there is a large difference between the speaker-specific characteristics of the target and distracter voices, performance is primarily determined by SNR. As speaker-specific differences between the target and distracter are reduced, performance decreases from the level imposed by the SNR by as much as 30%.

There is a strong interaction between the effects of GPR and VTL that takes the form of a relatively simple tradeoff. When the two variables are measured in logarithmic units, and there are no loudness cues to assist in tracking the target speaker, then a change in VTL has to be about 1.6 times a change in GPR to have the same effect on performance. The results are consistent with the notion that audibility is the prime determinant of performance, and that GPR and VTL are particularly effective when there are no loudness cues. The study has also demonstrated that the auditory system can use any combination of these cues to segregate competing speech signals.

ACKNOWLEDGMENTS

The research was supported by the UK Medical Research Council (Grant Nos. G0500221 and G9900369) and the European Office of Aerospace Research and Development (Grant No. FA8655-05-1-3043). The authors thank Joan Sussman and two anonymous reviewers for constructive comments on earlier versions of the manuscript.

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

[2]Darwin *et al.* (2003) stated that a nine-ST F0 difference corresponded to the advantage derived from a 38% VTL difference. However, in their Fig. 1, it is apparent that a four-ST difference yielded a 20% performance advantage while a nine-ST difference yielded an advantage of approximately 25%.

Assmann, P. F., and Summerfield, Q. (**1990**). "Modeling the perception of concurrent vowels: Vowels with different fundamental frequencies," J. Acoust. Soc. Am. **88**, 680–697.

Assmann, P. F., and Summerfield, Q. (**1994**). "The contribution of waveform interactions to the perception of concurrent vowels," J. Acoust. Soc. Am. **95**, 471–484.

Bolia, R. S., Nelson, W. T., Ericson, M. A., and Simpson, B. D. (**2000**). "A speech corpus for multitalker communications research," J. Acoust. Soc. Am. **107**, 1065–1066.

Brungart, D. S. (**2001**). "Informational and energetic masking effects in the perception of two simultaneous talkers," J. Acoust. Soc. Am. **109**, 1101–1109.

Brungart, D. S., Simpson, B. D., Ericson, M. A., and Scott, K. R. (**2001**). "Informational and energetic masking effects in the perception of multiple simultaneous talkers," J. Acoust. Soc. Am. **110**, 2527–2538.

Chalikia, M. H., and Bregman, A. S. (**1993**). "The perceptual segregation of simultaneous vowels with harmonic, shifted, or random components," Percept. Psychophys. **53**, 125–133.

Collins, S. A. (**2000**). "Men's voices and women's choices," Anim. Behav. **60**, 773–780.

Cooke, M. (**2006**). "A glimpsing model of speech perception in noise," J. Acoust. Soc. Am. **119**, 1562–1573.

Culling, J. F., and Darwin, C. J. (**1993**). "The role of timbre in the segregation of simultaneous voices with intersecting f0 contours," Percept. Psychophys. **54**, 303–309.
Darwin, C. J., Brungart, D. S., and Simpson, B. D. (**2003**). "Effects of fundamental frequency and vocal-tract length changes on attention to one of two simultaneous talkers," J. Acoust. Soc. Am. **114**, 2913–2922.
de Cheveigné, A. (**1993**). "Separation of concurrent harmonic sounds: Fundamental frequency estimation and a time-domain cancellation model of auditory processing," J. Acoust. Soc. Am. **93**, 3271–3290.
de Cheveigné, A. (**1997**). "Concurrent vowel identification. III. A neural model of harmonic interference cancellation," J. Acoust. Soc. Am. **101**, 2857–2865.
de Cheveigné, A., McAdams, S., and Marin, C. M. H. (**1997a**). "Concurrent vowel identification. II. Effects of phase, harmonicity and task," J. Acoust. Soc. Am. **101**, 2848–2856.
de Cheveigné, A., Kawahara, H., Tsuzaki, M., and Aikawa, K. (**1997b**). "Concurrent vowel identification. I. Effects of relative amplitude and f_0 difference," J. Acoust. Soc. Am. **101**, 2839–2847.
Elberling, C., Ludvigsen, C., and Lyregaard, P. E. (**1989**). "Dantale: A new Danish speech material," Scand. Audiol. **18**, 169–176.
Fant, G. C. M. (**1970**). *Acoustic Theory of Speech Production* (Mouton, The Hague).
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**, 1511–1122.
Ihlefeld, A., and Shinn-Cunningham, B. G. (**2008**). "Spatial release from energetic and informational masking in a selective speech identification task," J. Acoust. Soc. Am. **123**, 4369–4379.
Irino, T., and Patterson, R. D. (**2002**). "Segregating information about the size and shape of the vocal tract using a time-domain auditory model: The stabilised wavelet-mellin transform," Speech Commun. **36**, 181–203.
Ives, D. T., Smith, D. R., and Patterson, R. D. (**2005**). "Discrimination of speaker size from syllable phrases," J. Acoust. Soc. Am. **118**, 3816–3822.
Kawahara, H., and Irino, T. (**2004**). "Underlying principles of a high-quality speech manipulation system straight and its application to speech segregation," in *Speech Separation by Humans and Machines*, edited by P. L. Divenyi (Kluwer Academic, Boston, MA).
Ladefoged, P., and Broadbent, D. E. (**1957**). "Information conveyed by vowels," J. Acoust. Soc. Am. **29**, 98–104.
Lee, S., Potamianos, A., and Narayanan, S. (**1999**). "Acoustics of children's speech: Developmental changes of temporal and spectral parameters," J. Acoust. Soc. Am. **105**, 1455–1468.
Lombard, E. (**1911**). "Le signe de l'élévation de la voix (The characteristics of the raised voice)," Ann. Mal. Oreil. Larynx Nez Pharynx **37**, 101–119.
Marcus, S. M. (**1981**). "Acoustic determinants of perceptual center (p-center) location," Percept. Psychophys. **30**, 247–256.
Miller, G. A., and Licklider, J. C. R. (**1950**). "The intelligibility of interrupted speech," J. Acoust. Soc. Am. **22**, 167–173.
Moore, T. J. (**1981**). "Voice communications jamming research," in *AGARD Conference Proceedings 311: Aural Communication in Aviation* (AGARD, Neuilly-Sur-Seine, France), pp. 2:1–2:6.
Nearey, T. M. (**1989**). "Static, dynamic, and relational properties in vowel perception," J. Acoust. Soc. Am. **85**, 2088–2113.
Peterson, G. E., and Barney, H. L. (**1952**). "Control methods used in a study of the vowels," J. Acoust. Soc. Am. **24**, 175–184.
Plomp, R. (**1977**). "Acoustical aspects of cocktail parties," Acustica **38**, 186–191.
Qin, M. K., and Oxenham, A. J. (**2005**). "Effects of envelope-vocoder processing on f0 discrimination and concurrent-vowel identification," Ear Hear. **26**, 451–460.
Ritsma, R. J., and Hoekstra, A. (**1974**). "Frequency selectivity and the tonal residue," in *Facts and Models in Hearing*, edited by E. Zwicker and E. Terhardt (Springer, Berlin).
Rivenez, M., Darwin, C. J., Bourgeon, L., and Guillaume, A. (**2007**). "Unattended speech processing: Effect of vocal-tract length," J. Acoust. Soc. Am. **121**, EL90–EL95.
Rivenez, M., Darwin, C. J., and Guillaume, A. (**2006**). "Processing unattended speech," J. Acoust. Soc. Am. **119**, 4027–4040.
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.
Summerfield, Q., and Assmann, P. F. (**1991**). "Perception of concurrent vowels: Effects of harmonic misalignment and pitch-period asynchrony," J. Acoust. Soc. Am. **89**, 1364–1377.
Titze, I. R. (**1989**). "Physiologic and acoustic differences between male and female voices," J. Acoust. Soc. Am. **85**, 1699–1707.
Turner, R. E., Walters, T. C., and Monaghan, J. J. M. (**2009**). "A statistical formant-pattern model for segregating vowel type and vocal tract length in developmental formant data," J. Acoust. Soc. Am. (in press).
van Dinther, R., and Patterson, R. D. (**2006**). "The perception of size in musical instruments," J. Acoust. Soc. Am. **120**, 2158–2176.
von Kriegstein, K., Warren, J. D., Ives, D. T., Patterson, R. D., and Griffiths, T. D. (**2006**). "Processing the acoustic effect of size in speech sounds," Neuroimage **32**, 368–375.
Walters, T. C., Gomersall, P., Turner, R. E., and Patterson, R. D. (**2008**). "Comparison of relative and absolute judgments of speaker size based on vowel sounds," Proc. Meet. Acoust. **1**, 050003.
Watson, C. S. (**2005**). "Some comments on informational masking," Acta. Acust. Acust. **91**, 502–512.

Identifying isolated, multispeaker Mandarin tones from brief acoustic input: A perceptual and acoustic study

Chao-Yang Lee[a)]
School of Hearing, Speech and Language Sciences, Ohio University, Athens, Ohio 45701

(Received 28 January 2008; revised 8 November 2008; accepted 24 November 2008)

Lexical tone identification relies primarily on the processing of F0. Since F0 range differs across individuals, the interpretation of F0 usually requires reference to specific speakers. This study examined whether multispeaker Mandarin tone stimuli could be identified without cues commonly considered necessary for speaker normalization. The *sa* syllables, produced by 16 speakers of each gender, were digitally processed such that only the fricative and the first six glottal periods remained in the stimuli, neutralizing the dynamic F0 contrasts among the tones. Each stimulus was presented once, in isolation, to 40 native listeners who had no prior exposure to the speakers' voices. Chi-square analyses showed that tone identification accuracy exceeded chance as did tone classification based on F0 height. Acoustic analyses showed contrasts between the high- and low-onset tones in F0, duration, and two voice quality measures (F1 bandwidth and spectral tilt). Correlation analyses showed that F0 covaried with the voice quality measures and that tone classification based on F0 height also correlated with these acoustic measures. Since the same acoustic measures consistently distinguished the female from the male stimuli, gender detection may be implicated in F0 height estimation when no context, dynamic F0, or familiarity with speaker voices is available. © *2009 Acoustical Society of America.* [DOI: 10.1121/1.3050322]

PACS number(s): 43.71.An, 43.71.Bp, 43.71.Es [AJ] Pages: 1125–1137

I. INTRODUCTION

A. Overview

Dealing with speaker variability is an integral part of spoken language processing (Johnson, 2005). To uncover the linguistic representation intended by a speaker, a listener has to disentangle the linguistic information from speaker-specific information in the acoustic signal. The speaker variability issue is pertinent to lexical tone perception, which relies primarily on the detection of fundamental frequency (F0) (Abramson, 1978; Gandour and Harshman, 1978; Howie, 1976; Tseng, 1981). Since F0 range differs across individuals, absolute F0 values of a particular tone may vary by speaker. For example, a phonologically low tone produced by a female speaker could be acoustically equivalent to a phonologically high tone for a male speaker. What is considered high or low, then, has to be determined with respect to a speaker's F0 range.

Dealing with acoustically impoverished stimuli is also a common challenge for speech perception. Studies using degraded acoustic stimuli have revealed contextual and secondary cues to lexical tone identification (Abramson, 1972; Blicher *et al.*, 1990; Gottfried and Suiter, 1997; Lee, 2000; Lee *et al.*, 2008; Liu and Samuel, 2004; Whalen and Xu, 1992; Xu, 1994). For example, Gottfried and Suiter (1997) showed that "silent-center" Mandarin tones, where the majority of F0 information is missing, could be identified as accurately as intact tones. Overall, these studies indicate that listeners are capable of integrating the limited F0 information in the signal or resorting to secondary cues to uncover tone identity when F0 information is not available.

How well can listeners cope with the dual challenge of speaker variability and limited acoustic input in tone perception? This article reports a perceptual experiment and acoustic analyses that examined the roles of speaker variability and limited acoustic input in Mandarin tone identification. The research question is whether tones produced by multiple speakers can be identified without prior exposure to the speakers' voices, when syllable-internal F0 information is reduced to a minimum and when the syllable-external context is not available for speaker normalization. To this end, multispeaker, naturally produced Mandarin tones were digitally processed such that only the onset consonant and the first six glottal periods remained in the stimuli. Acoustic analyses showed that the signal processing effectively removed the dynamic F0 information that distinguishes the four Mandarin tones (Tone 1: level; Tone 2: rising; Tone 3: dipping; Tone 4: falling). While these tones may still be identified based on F0 height contrasts (Tones 1 and 4: high; Tones 2 and 3: low), the height judgment depends on the knowledge of a speaker's F0 range, which should be difficult to estimate without prior exposure or an external context. If, however, listeners can identify the tones from these incomplete stimuli with accuracy exceeding chance, the question is what in the acoustic signal enables them to do so. The basis of the tone identification performance will be explored by acoustic analyses.

B. Speaker variability in lexical tone perception

Information about a speaker's F0 range can be obtained from a sentential context (Leather, 1983; Moore and Jong-

[a)]Electronic mail: leec1@ohio.edu

man, 1997; Wong and Diehl, 2003). For contour tones presented in isolation (such as the Mandarin rising, dipping, and falling tones), speaker variability may not be a problem because these tones have distinct dynamic F0 patterns and may not require reference to F0 height for identification (Moore and Jongman, 1997). However, F0 height judgment is relevant for the perception of noncontour tones such as the Cantonese level tones (Wong and Diehl, 2003). F0 height may also be implicated in the perception of contour tones that have similar F0 patterns but are located in different registers of a speaker's F0 range. Even for tones in connected speech, F0 height estimation may still be relevant because the canonical F0 contour of a tone often changes substantially as a result of tonal coarticulation (Xu, 1994, 1997). For example, the Mandarin dipping tone becomes phonetically a low tone in nonfinal positions. Consequently, the low tone may contrast with the high tone (a level tone) only in F0 height. Under these circumstances, syllable-intrinsic F0 patterns may not provide sufficiently distinctive information and a listener will have to resort to F0 height for tone identification.

There is ample evidence that F0 height is involved in the perception of Mandarin tones even in isolation (Gandour, 1983; Gandour and Harshman, 1978; Massaro *et al.*, 1985). The role of F0 height implies the use of F0 range information. Since the tone stimuli used in these studies were presented in isolation, F0 range information could not have come from an external context. However, the synthetic tone stimuli used in these studies included contour tones; therefore the listeners could have estimated F0 height by calibrating F0 range syllable internally. The listeners could also have developed familiarity with the synthetic voice and thus the F0 range through repeated exposure to the stimuli (Honorof and Whalen, 2005; Nygaard and Pisoni, 1998; Palmeri *et al.*, 1993). Since only one speaker was modeled in these studies, it is not known whether F0 height could be judged when multiple speakers are present.

Other studies presented tones in carrier phrases of various F0 levels as a way of simulating different speakers (Fox and Qi, 1990; Leather, 1983; Lin and Wang, 1984; Moore and Jongman, 1997; Wong and Diehl, 2003). Leather (1983) and Moore and Jongman (1997) showed that synthesized Mandarin tone stimuli were identified with reference to the perceived F0 range obtained from naturally produced carrier phrases. Wong and Diehl (2003) similarly found that a naturally produced Cantonese tone stimulus was identified as different tone categories depending on the F0 height of synthesized carrier phrases. Importantly, by using level tone stimuli only, Wong and Diehl (2003) eliminated the potential contribution of syllable-intrinsic dynamic F0 patterns to the F0 range estimation. Since no syllable-internal dynamic F0 information was available, F0 height estimation must have been achieved by obtaining information from the carrier phrases.

If F0 range information can only be derived from an external context, removing the context should effectively prevent F0 range estimation, which should compromise tone identification that requires F0 height information. However, there is some indication that multispeaker, isolated tones lacking dynamic F0 information can be identified without context. In Wong and Diehl's (2003) Experiment 1, the three Cantonese level tones (high, mid, and low) produced by seven speakers were presented in isolation. The results showed that tone identification was more accurate when the stimulus presentation was blocked by the speaker (80.3%) than when it was mixed across the speakers (48.6%). The accuracy values suggest that the listeners were able to identify the F0 level in these isolated, multispeaker tone stimuli beyond chance (33.3%). However, familiarity with voices could have contributed to the results. In particular, the listeners in the experiment heard the stimulus set 12 times. Given that the tone stimuli span across a substantial F0 range used for Cantonese tones, it is likely the multiple exposures allowed the listeners to learn the F0 range of the speakers (Honorof and Whalen, 2005; Palmeri *et al.*, 1993; Nygaard and Pisoni, 1998).

A more stringent test for speaker normalization without context, then, is to present isolated, multispeaker tone stimuli lacking dynamic F0 information without repeated exposure. No data currently exist on lexical tone materials, but Honorof and Whalen (2005) showed that English-speaking listeners were able to locate an F0 reliably within a speaker's F0 range without context or prior exposure to a speaker's voice. In their study, isolated vowel [ɑ] tokens, produced by 20 English speakers with varying F0s, were presented to listeners to judge where each token was located in the speakers' F0 ranges. The results showed significant correlations between the perceived F0 location and the actual location in the speakers' F0 ranges. Voice quality and gender detection were implicated (Honorof and Whalen, 2005). If listeners can demonstrate such an ability for pitches that are not linguistically significant, this ability could also be implicated in the perception of lexical tones. Considering the lexically contrastive function of F0 in tone languages, the ability to locate within an F0 range isolated pitches produced by unfamiliar voices can be quite useful.

C. Tone identification from incomplete acoustic input

Since three of the four Mandarin tones are contour tones with dynamic F0 patterns, the issues of F0 height estimation and speaker normalization may not seem relevant for Mandarin tone identification. That is, Mandarin listeners should be able to rely on dynamic F0 information for tone identification. Nonetheless, research has consistently shown that detection of F0 height is implicated in Mandarin tone perception (Fox and Qi, 1990; Gandour, 1983; Gandour and Harshman, 1978; Leather, 1983; Lin and Wang, 1984; Massaro *et al.*, 1985; Moore and Jongman, 1997). To isolate the role of F0 height estimation in tone identification, a potentially useful research strategy is to use brief tone segments extracted from intact tones as stimuli such that dynamic F0 information can be minimized or removed. Although no study has been carried out that specifically examined speaker normalization in this way, there are data on how well brief tone segments can be identified. Whalen and Xu (1992) examined Mandarin tone identification from segments ranging from 40 to 100 ms extracted from various parts of a syllable.

Tone segments that do not have much F0 change were most often heard as Tone 1 (level). A low, unchanging F0 was often heard as Tone 3 (dipping). Recall that a nonfinal Tone 3 is usually realized as a low tone. These results therefore suggest that the listeners were referring to F0 height when dealing with the brief tone segments. Whalen and Xu (1992) also predicted that very short syllables such as those in running speech would show more of such "register tendency." Sensitivity to register, as noted, implies the ability to locate a particular pitch within a speaker's F0 range.

Gottfried and Suiter (1997) showed that isolated Mandarin tone stimuli excised from the first six glottal pulses of a syllable could be identified quite accurately (Tone 1:91%; Tone 2:28%; Tone 3:65%; Tone 4:32%). This finding was replicated by Lee *et al.* (2008) with a similar procedure but a different set of Mandarin syllables (Tone 1:68%; Tone 2:46%; Tone 3:87%; Tone 4:90%). The contrast in the actual accuracy values from these two studies could be attributed to the differences in stimuli, participants, and response format. Although it has been established that isolated Mandarin tones do not require the entire syllable to identify (Tseng, 1981; Yang, 1992), the accuracy values reported by Gottfried and Suiter (1997) and Lee *et al.* (2008) are still surprisingly high. In particular, the F0 contours remaining in the six glottal pulses were basically flat and did not show distinct patterns among the four tones (Lee *et al.*, 2008). Even if there were distinct F0 contours acoustically, the dynamic character of the pitch contour cannot be perceived at such a short duration (Greenberg and Zee, 1979). On the other hand, the acoustic analyses of Lee *et al.* (2008) did show distinct groupings of the four tones in F0 height: the high-onset tones included Tones 1 and 4; the low-onset tones included Tones 2 and 3. These acoustic data are consistent with the finding that tones belonging to the same height group were more confusable perceptually. It appears that the listeners were indeed sensitive to F0 height contrasts when dynamic F0 information was not available. Again, being able to judge F0 height implies the ability to locate pitches in a speaker's F0 range.

The sensitivity to F0 height, however, was not observed in Wu and Shu's (2003) data. Using the gating paradigm (Grosjean, 1996), Wu and Shu (2003) presented isolated Mandarin syllable fragments in 40 ms increments, starting with an 80 ms gate. The task for the participants was to propose a monosyllabic word candidate at each gate that best matched the acoustic input. Analyses of the candidates proposed at each gate showed that the most common tone error for Tone 1 stimuli was Tone 4 up to 120 ms of input, which is consistent with the confusion patterns from the previous three studies (Whalen and Xu, 1992; Gottfried and Suiter, 1997; Lee *et al.*, 2008). For stimuli generated from the other three tones, however, the most common error was invariably Tone 1. While these results are consistent with Greenberg and Zee's (1979) finding that tone contours cannot be perceived with durations shorter than 130 ms, the predominant Tone 1 response, even for the low-onset tones (Tones 2 and 3), did not indicate any sensitivity to F0 height.

As with earlier studies, only one speaker was used to generate the stimuli in Whalen and Xu (1992), Gottfried and Suiter (1997), Lee *et al.* (2008), and Wu and Shu (2003). For the three studies that did show signs of sensitivity to F0 height, the listeners had ample opportunity to become familiar with the speaker's voice through repeated exposures. Familiarity, as noted, could contribute to F0 range estimation. Therefore it is not clear if the high accuracy reported in some of the studies truly reflects the ability to detect F0 height. Nonetheless, these studies showed that the use of brief, gated tone stimuli can be an effective way of neutralizing dynamic F0 information in Mandarin tones.

D. Summary and predictions

The available evidence suggests that lexical tone perception involves processing F0 information with reference to a speaker's F0 range (Leather, 1983; Moore and Jongman, 1997; Wong and Diehl, 2003). However, how speaker normalization is accomplished remains underspecified. The acoustic basis for the perceptual ability also remains unsubstantiated. The finding that nonlinguistic, isolated pitches produced by an unfamiliar voice can be reliably located within a speaker's F0 range (Honorof and Whalen, 2005) invites the question of whether a similar ability is implicated in the perception of lexically contrastive tones.

The present study aims to address this issue by examining the identification of brief portions of Mandarin tones and the acoustic characteristics of the tone stimuli. Only six glottal pulses were preserved in the tone stimuli such that syllable-internal dynamic F0 information was neutralized. The stimuli were presented in isolation such that no contextual information was available. Each of the multispeaker tone stimuli was presented only once such that familiarity with speaker voice was minimized. Since dynamic F0, context, and familiarity are known to contribute to speaker normalization, eliminating all of them from the stimuli should make speaker normalization very difficult. Consequently, tone identification from these multispeaker stimuli should not exceed chance level. If so, this result will indicate that dynamic F0, context, and/or familiarity are indeed necessary for speaker normalization.

If, however, these tone stimuli can be identified with accuracy exceeding chance, listeners must be able to obtain useful information from the isolated stimuli for speaker normalization. One possibility is that some acoustic properties in the stimuli covary with F0 to facilitate F0 height estimation (Honorof and Whalen, 2005). Previous research has shown that secondary cues such as duration and amplitude contour can be used for lexical tone identification when F0 information is compromised (Abramson, 1972; Blicher *et al.*, 1990; Liu and Samuel, 2004; Whalen and Xu, 1992). Voice quality has also been proposed as the basis for the ability to locate a pitch in a speaker's F0 range (Honorof and Whalen, 2005; Swerts and Veldhuis, 2001). In particular, voice quality may covary with F0, or it may be used for gender detection as a first pass for F0 height estimation. It is known that listeners are able to detect speaker gender from units of speech ranging from the sentence to individual sounds (Childers and Wu, 1991; Ingemann, 1968; Lass *et al.*, 1978). Honorof and Whalen (2005) also reported positive correlations between perceived pitch locations and gender-related

estimates, indicating the potential role of gender detection in F0 height estimation. Considering the female-male difference in average F0 (Peterson and Barney, 1952) and voice quality (Hanson, 1997; Hanson and Chuang, 1999), successful gender identification could mediate F0 height judgments. The potential contribution of these measures will be explored in the acoustic analyses.

Another possibility is that F0 height can be estimated based on some internal templates of speaking F0 range, which are formed based on experience with the prevailing F0 range of the speakers in a linguistic community (Dolson, 1994; Honorof and Whalen, 2005). Dolson (1994) noted that despite anatomical variabilities across individuals, the speaking F0 range is not wildly variable within a linguistic community. His survey showed that 80% of the male speakers within a linguistic community have an average speaking F0 within three semitones of the group mean, and the female speakers' F0 range is even narrower. The idea is that listeners can acquire an internal representation of pitch classes based on the prevailing speaking F0 range in their linguistic community, and that pitch perception and production are both influenced by this representation (Deutsch, 1991; Deutsch, *et al.*, 1999, 2004; Deutsch *et al.*, 1990). If the speaking F0 range in a linguistic community is sufficiently constrained, listeners may be able to locate a particular F0 based on these stored pitch templates.

II. TONE IDENTIFICATION EXPERIMENT

A. Method

1. Materials

The Mandarin syllable *sa*, produced with all four tones by 16 female and 16 male Beijing Mandarin speakers (age range: 23–35 years), was selected to generate the stimuli for the experiment. The syllables were drawn from an existing database collected at Ohio University in 2004 that includes the Mandarin syllables *sa*, *xia*, and *sha*. The *sa* syllable was chosen because the same materials were to be used in a subsequent study that would investigate brief tone identification by English-speaking listeners. To that end, the intention was to select a syllable that exists in both languages to avoid potential interference from segmental structure. Among the three fricatives, the alveolar fricative is phonologically closest to an actual phoneme in English. In contrast, the articulatory and acoustic characteristics of the other two Mandarin fricatives differ more substantially from the English palatoalveolar fricative (Ladefoged, 1996; Stevens *et al.*, 2004).

The recordings were made in a sound-treated booth with an Audio-technica AT825 field recording microphone connected through a preamplifier and A/D converter (USBPre microphone interface) to a Windows personal computer. The speakers were instructed to read the syllables in citation form. The recordings were digitized with the Brown Lab Interactive Speech System (BLISS, Mertus, 2000) at 44.1 kHz with 16 bit quantization. Each syllable was identified from the BLISS waveform display, excised from the master file, and saved as an individual audio file. The peak amplitude was normalized across syllables.

Each *sa* syllable was digitally processed with BLISS such that the stimuli included only the fricative consonant and the first six glottal periods. The cut was always made at a zero crossing. There were no perceptible clicks as a result of the signal processing; therefore no further tapering procedure was applied. A total of 128 stimuli (4 tones $\times$ 32 speakers) were used in the experiment.

2. Participants

Forty Beijing Mandarin speakers were recruited from the Ohio University community to participate in the experiment with cash compensation. The participants included 27 females (mean age=24 years, SD=4.5) and 13 males (mean age=21 years, SD=2.6). All spoke Mandarin on a daily basis and none reported any speech or hearing difficulties. Twenty-one participants reported speaking some dialect of Chinese other than Mandarin, but all identified Mandarin as their native language.

3. Procedure

The stimuli were imported to AVRunner, the subject-testing program in BLISS, for stimulus presentation and response data acquisition. To minimize the impact of familiarity with individual speaker voices, the 128 stimuli produced by the 32 speakers were assigned to four blocks such that each block included only one stimulus from a speaker. That is, each block had 32 stimuli and all stimuli were produced by different speakers. Within each block, the number of male and female speakers was balanced (i.e., 16 females and 16 males) as was the number of the four tones (i.e., eight stimuli for each of the four tones). No two participants received the same order of presentation. The order of presentation for the blocks was also randomized. The interstimulus interval was 5 s. Eight practice stimuli were given prior to the experiment to familiarize the participants with the procedure and response format. The practice stimuli, also excised *sa* syllables, were recorded by a female and a male speaker not used in the actual experiment.

Participants were tested individually in a sound-treated room. They listened to the stimuli through a pair of KOSS R80 headphones connected to a personal computer. The participants were told they would be listening to the syllable *sa* with all four tones produced by 32 female and male speakers. They were also told the syllables had been digitally processed such that only the very beginning of the syllables was audible. Their task was to identify the tone of each stimulus by pressing the buttons labeled "1," "2," "3," and "4" on the computer keyboard, representing the four Mandarin tones. All participants indicated they were familiar with the convention of designating Mandarin tones with the numbers. The participants were further told that they had 5 s to respond to each stimulus and that their response would be timed.

B. Results

To evaluate the null hypothesis that tone identification responses were not related to the stimulus tones, the expected and actual responses were tabulated to generate con-

TABLE I. Contingency tables showing the frequency counts of observed and expected (in parentheses) tone identification responses.

	Female stimuli Response			
Stimulus	Tone 1	Tone 2	Tone 3	Tone 4
Tone 1	281 (234.75)	102 (109.75)	59 (120)	194 (171.5)
Tone 2	239 (235.119)	133 (109.923)	140 (120.189)	125 (171.77)
Tone 3	187 (235.119)	108 (109.923)	213 (120.189)	129 (171.77)
Tone 4	232 (234.012)	96 (109.405)	68 (119.623)	238 (170.961)
	Male stimuli Response			
Tone 1	268 (226.821)	116 (106.416)	107 (150.132)	144 (151.631)
Tone 2	228 (227.536)	103 (106.751)	180 (150.605)	126 (152.108)
Tone 3	154 (226.464)	88 (106.249)	228 (149.895)	164 (151.392)
Tone 4	258 (227.179)	119 (106.584)	86 (150.368)	173 (151.869)

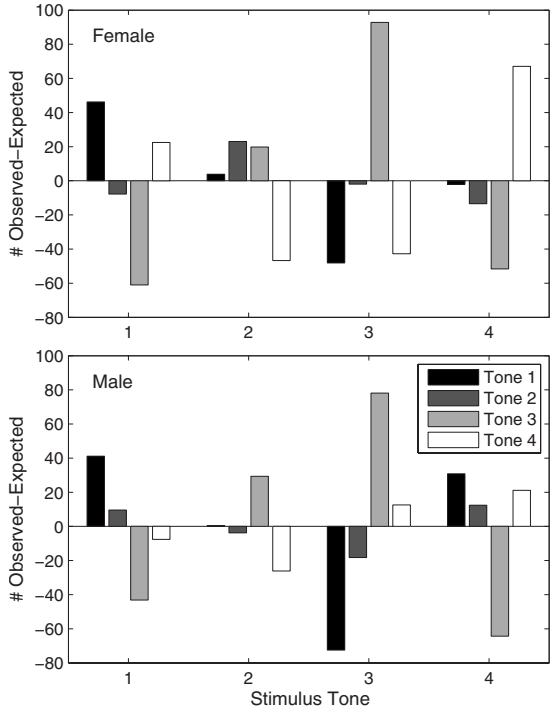


FIG. 1. Difference between the numbers of observed and expected tone responses as a function of stimulus tone in the tone identification experiment.

tingency tables for χ^2 tests. Table I shows the numbers of expected and actual tone responses to the tone stimuli. The expected frequencies were calculated as follows: Suppose E_{ij} is the expected frequency for the cell in row i and column j, R_i and C_j are the corresponding row and column totals (marginal totals), and N is the total number of observations, then $E_{ij}=R_iC_j/N$ (Howell, 1999; Cohen, 1996).

Separate χ^2 tests were conducted for female and male stimuli to test the null hypothesis that the responses were independent of the stimuli. Both tests showed that the null hypothesis should be rejected: female, χ^2 $(9,N=2544)=206.96$, $p<0.0001$; male, χ^2 $(9,N=2542)=135.66$, $p<0.0001$. (The unequal N resulted from 16 and 18 missing responses in the female and male data, respectively.) These results indicate that the listeners' tone identification responses were not random or totally unrelated to the tone stimuli. That is, the overall accuracy was different from chance level performance.

A second set of contingency tables was generated based on specific tones. To evaluate Tone 1 identification, for example, both the stimuli and responses were coded as Tone 1 or non-Tone 1. Therefore, all responses could be classified into "hit" (Tone 1 stimulus→Tone 1 response), "miss" (Tone 1 stimulus→non-Tone 1 response), "false alarm" (non-Tone 1 stimulus→Tone 1 response), and "correct rejection" (non-Tone 1 stimulus→non-Tone 1 response). For the remaining three tones, the same coding procedure was applied to generate similar 2×2 contingency tables. As before, χ^2 tests were performed to test the null hypothesis that the responses were not related to the stimuli. For the female stimuli, all four χ^2 tests indicate that the null hypothesis should be rejected: Tone 1, χ^2 $(1,N=2544)=19.26$, $p<0.0001$; Tone 2, χ^2 $(1,N=2544)=7.81$, $p<0.01$; Tone 3, χ^2 $(1,N=2544)=117.85$, $p<0.0001$; Tone 4, χ^2 $(1,N=2544)=47.94$, $p<0.0001$. For the male stimuli, all but the Tone 2 result indicate that the null hypothesis should be rejected: Tone 1, χ^2 $(1,N=2542)=15.5$, $p<0.0001$; Tone 2, χ^2 $(1,N=2542)=0.21$, n.s.; Tone 3, χ^2 $(1,N=2542)=71.01$, $p<0.0001$; Tone 4, χ^2 $(1,N=2542)=5.15$, $p<0.05$. Overall, these results show that tone identification performance was above chance level except for the responses to Tone 2 stimuli produced by the male speakers.

A final set of χ^2 tests was conducted based on F0 height of the tones: high-onset tones include Tones 1 and 4; low-onset tones include Tones 2 and 3. As the acoustic analyses will show, dynamic F0 information was absent from these brief stimuli, but the F0 height difference remained between the high- and low-onset tones. If the identification data show that the high-low distinction can be detected beyond chance, the ability to estimate F0 height is likely the basis for the tone identification performance. Indeed, when both the stimuli and responses were coded by F0 height, χ^2 tests showed that the high-low judgments were not random: for the female stimuli, χ^2 $(1,N=2544)=121.95$, $p<0.0001$; for the male stimuli, χ^2 $(1,N=2542)=47.77$, $p<0.0001$. These results indicate that the listeners were capable of judging the high-low distinction above chance.

The contingency tables also provide information about the confusion patterns among the four tones. To facilitate the interpretation of the patterns, Fig. 1 shows the difference between the expected and actual response counts as a function of stimulus tone. Each bar represents the result of subtracting the number of expected responses from the number of actual responses in each cell of Table I. Positive or greater values indicate that the response tone is more confusable

with the stimulus tone. By contrast, the smaller or more negative a value is, the less confusable the response tone is with the stimulus tone. There are some differences between the female and male stimuli, but some common patterns can still be seen. In particular, Tone 1 was least often misidentified as Tone 3; Tone 2 was most confused with Tone 3 and least identified as Tone 4; Tone 3 was rarely misidentified as Tone 1; Tone 4 was least misidentified as Tone 3. There are also some gender-specific patterns. Specifically, while the female data indicate that Tone 1 was most often misidentified as Tone 4, the male data show that Tone 4 was most often misidentified as Tone 1. These results suggest there might be gender differences in the stimuli, which will be examined in the acoustic analyses. Nonetheless, all these results are consistent with the observation that tone stimuli sharing a similar F0 height are more confusable, indicating that the high-low distinction can be detected by the listeners.

C. Discussion

When presented with multispeaker, isolated tone stimuli with six glottal periods, Mandarin listeners were able to identify the tones of the stimuli with accuracy exceeding chance. With the exception of Tone 2 produced by the male speakers, tone identification responses were not random but were contingent on the tone stimuli, indicating the ability to distinguish the target tone from the nontarget tones. With only six glottal periods in a stimulus, it is unlikely that dynamic F0 information was available for tone identification. Rather, detection of F0 height is most likely the basis for tone identification. This interpretation is further evaluated by acoustic analyses reported in Sec. III.

These results suggest that F0 height detection is the basis for the tone identification performance. Given that the tone stimuli were produced by 32 speakers and that the F0 height of a tone is likely to vary across individuals, it can be inferred that some sort of speaker normalization was involved in the tone identification process. Since no context was given, whatever allowed the speaker normalization must have come from the stimuli internally. Familiarity with individual speaker voices is unlikely to have contributed to the normalization since each tone stimulus was presented only once. Taken together, it is reasonable to conclude that some acoustic characteristics in the stimuli facilitated the speaker normalization and thus the tone identification.

As with most phonetic distinctions, there are multiple acoustic cues to tonal contrasts (e.g., Whalen and Xu, 1992). Since dynamic F0 information was presumably compromised in the stimuli, it is conceivable that some of the secondary cues such as duration and amplitude might be implicated. Acoustic analyses were conducted on the tone stimuli to explore the contribution of these cues. The acoustic analyses were also intended to verify several assumptions made earlier such as the lack of dynamic F0 in the stimuli, the high-versus low-onset distinction among the four tones, and the existence of speaker variability in F0 height. Finally, the acoustic analyses could also explicate the gender differences observed in the perceptual results. In particular, the χ^2 test for male Tone 2 stimuli failed to reach significance; the overall accuracy for male stimuli was also lower. These results suggest the tonal distinctions might not be as clear in the male stimuli. These observations will be evaluated by the acoustic analyses.

The Tone 2 stimuli in the current study appear to have a disadvantage compared to the other tones. In particular, the contingency table analyses showed the Tone 2 stimuli produced by the male speakers were the only stimuli among the four tones that failed to be identified with accuracy exceeding chance. A potentially confounding factor is lexical status/frequency. In particular, the syllable *sa* with Tone 2 happens to be an accidental gap in the Mandarin lexicon; that is, this syllable-tone combination is not associated with a real word in the language. Even when all the syllables that begin with the *sa* sequence are considered (*sa, sai, san, sang*, and *sao*), none of the syllable-tone combinations is associated with a real word (Wang, 1986). Given the demonstrated influence of lexical status on Mandarin tone categorization (Fox and Unkefer, 1985), the Tone 2 disadvantage could have resulted from the accidental gap. Alternatively, since Gottfried and Suiter (1997) and Lee *et al.* (2008) also reported the lowest accuracy and/or longest reaction time for onset-only Tone 2, it is equally plausible that the low accuracy for Tone 2 reflects the acoustic characteristics of the stimuli or global distributional patterns of the four tones in the language. Nonetheless, a syllable that is balanced in lexical frequency across all four tones should be used in future studies.

III. ACOUSTIC ANALYSES

Acoustic analyses were conducted on the F0, duration, amplitude, and voice quality of the stimuli. Examining F0 is an obvious choice since it is the primary acoustic correlate of Mandarin tones (e.g., Abramson, 1978). Duration is a direct measure of the amount of acoustic input and has been shown to be a cue to tone identification when F0 is not available (e.g., Liu and Samuel, 2004; Whalen and Xu, 1992). There is also evidence that amplitude contours could be used to identify Mandarin tones in the absence of F0 (Whalen and Xu, 1992). Although the peak amplitude had been normalized for all the syllables prior to the silencing procedure, it is possible that some differences among the four tones could still exist in the gated stimuli.

The use of voice quality to distinguish lexical tones has been noted in several acoustic studies (Andruski, 2006; Andruski and Ratliff, 2000; Huffman, 1987). Specifically, Andruski (2006) and Andruski and Ratliff (2000) showed that the amplitude difference between the first and second harmonic (H1-H2), a measure of glottal open quotient, is distinct among the modal, creaky, and breathy tones that occupy a crowded F0 space in Green Mong. For these tones that are minimally distinct in F0, voice quality carries part of the functional load for tonal distinctions. Although voice quality is not used phonemically for Mandarin tones, isolated Tone 3 has been noted to involve glottalization for some speakers (e.g., Liu and Samuel, 2004). In addition, acoustic measures of glottal configuration have been shown to distinguish between female and male voices (Hanson, 1997; Hanson and Chuang, 1999), which could be implicated in F0 height esti-

mation (Honorof and Whalen, 2005). To explore the potential contribution of glottal configuration, acoustic measures that have been shown to reflect the status of the glottis (open quotient, F1 bandwidth, and spectral tilt) were examined.

A. Method

The F0 of the first six glottal periods was measured cycle by cycle from the BLISS waveform display. The duration of the six glottal periods was also measured from the waveform display. Root mean square (rms) amplitude was calculated using MATLAB. Specifically, the sound files were imported into MATLAB as digital samples with amplitude values ranging from −1 to 1. The rms values were calculated for the frication noise and the six glottal periods separately. A normalized amplitude was also calculated by subtracting the amplitude of the frication noise from that of the six glottal periods.

Finally, three measures of voice quality were taken that have been shown to indicate glottal configuration. They included open quotient, F1 bandwidth, and spectral tilt (Hanson, 1997; Hanson and Chuang, 1999; Stevens, 1998). In particular, the amplitude difference between the first and second harmonic (H1-H2) is a measure of open quotient or the percent of a glottal cycle in which the glottis is open. A larger H1-H2 value indicates that the glottis is open for a significant portion of a cycle. The amplitude difference between the first harmonic and the strongest harmonic in the F1 range (H1-A1) is a measure of F1 bandwidth. A wider F1 bandwidth indicates greater acoustic loss at the glottis due to the incomplete glottal closure. The amplitude difference between the first harmonic and the strongest harmonic in the F3 range (H1-A3) is a measure of spectral tilt. A larger spectral tilt also indicates greater acoustic loss at the glottis, indicating that the glottis is never closed during a cycle. To obtain these voice quality measures, a Hamming window of 25.6 ms was placed at the beginning of acoustic periodicity for each stimulus. A power spectrum was generated with BLISS by applying discrete Fourier transform to the signal. The first harmonic (H1), second harmonic (H2), the strongest harmonic in the F1 range (A1), and the strongest harmonic in the F3 range (A3) were identified from the spectrum. The amplitude values were measured from the spectrum to derive the three measures of glottal configuration.

To evaluate whether these acoustic properties were different among the tones and between the genders, analyses of variance (ANOVAs) were conducted on the acoustic measures with stimulus tone (1, 2, 3, and 4) as a within-subject factor, gender (female and male) as a between-subject factor, and speakers as a random factor. When a main effect from an ANOVA was significant, the Bonferroni *post hoc* test was used for pairwise means comparisons to keep the familywise type I error rate at 5%.

B. Results

1. Fundamental frequency

Figure 2 shows the average F0 values for each of the six glottal periods. Each data point represents an average of 16 female or 16 male speakers. As the figure shows, all four F0 contours are fairly similar to each other, showing minimal contrasts. Specifically, the F0 drops somewhat during glottal cycles 1–3 and stays relatively flat during glottal cycles 3–6. As predicted, the dynamic F0 difference among the four tones has been neutralized.

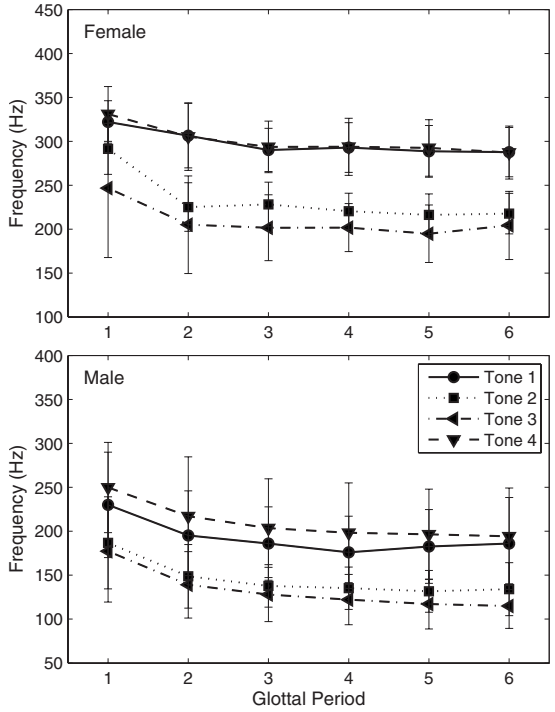


FIG. 2. Average F0 (±1SD) of each of the six glottal periods in the tone stimuli. Each data point represents an average of 16 female or 16 male speakers.

For both the female and male stimuli, the four tones form two distinct groups. In particular, Tones 1 and 4 have higher F0 values than Tones 2 and 3. Despite the individual variability shown by the error bars, the high-onset tones are fairly separated from the low-onset tones. This is particularly obvious for the female speakers, where the high-onset tones show little overlap with the low-onset tones. The grouping of the four tones is consistent with the traditional description of isolated Mandarin tones (that Tones 1 and 4 start higher than Tones 2 and 3) and with the confusion patterns found in the identification experiment.

To obtain a quantitative measure of the tone and gender differences, a mean F0 value for each tone was calculated by averaging across the six glottal periods. Since the F0 from one period to the next are similar, an average across the six glottal pulses is a reasonable representation of the tone. The ANOVA on the mean F0 showed a main effect of gender, $F(1,30)=67.15$, $p<0.0001$, and a main effect of tone, $F(3,90)=119.58$, $p<0.0001$. The gender effect is expected, with female speakers showing a higher average F0 than male speakers. For the tone effect, pairwise means comparisons showed all pairwise comparisons were significant except for the Tone 1–Tone 4 comparison. Specifically, Tones 1 and 4 have a higher F0 than Tone 2, which in turn has a higher F0 than Tone 3. That is, all tone pairs, except for Tones 1 and 4, are statistically distinct from each other. The lack of difference between Tones 1 and 4 is consistent with the confusion

pattern reported earlier. Furthermore, the result that Tone 3 occupies the very low end of the F0 range and contrasts maximally with Tones 1 and 4 is also compatible with the finding that Tone 3 was identified quite accurately (Recall in the contingency table analyses, Tone 3 had the greatest positive difference between observed and expected responses for both the female and male stimuli). As Whalen and Xu (1992) also reported, a low, unchanging F0 is often heard as Tone 3. The converging evidence suggests that the listeners were able to detect a low F0 fairly well and that they were likely to call it a Tone 3.

Finally, the existence of speaker variability in F0 height is evident in Fig. 2. In addition to the variability shown by the error bars, the comparison between the female and male stimuli is particularly telling. In particular, the low-onset tones for the female speakers overlap substantially with the high-onset tones for the male speakers. Acoustically, the female "low" tones are almost identical to the male "high" tones. Nonetheless, the contingency table analyses showed that all four tones involved (female: Tones 2 and 3; male: Tones 1 and 4) were identified with accuracy beyond chance. This observation suggests the possibility that the listeners were able to detect speaker gender first and then locate the stimulus F0 within the F0 range based on the gender judgment.

2. Duration

The average duration of the six glottal periods ranges from 20 to 49 ms. For the female stimuli, Tone 1 (21 ms, SD=2), Tone 2 (26 ms, SD=2), Tone 3 (31 ms, SD=8), and Tone 4 (20 ms, SD=2). For the male stimuli, Tone 1 (33 ms, SD=8), Tone 2 (44 ms, SD=8), Tone 3 (49 ms, SD=13), and Tone 4 (31 ms, SD=8). With such a short duration, the question arises whether any durational differences can be perceived. Although no data are available on the duration difference limen (ΔT) for lexical tones, several studies using noise and pure tone burst stimuli provided information on ΔT as a function of stimulus duration (Abel, 1972; Sinnott *et al.* 1987; Dooley and Moore, 1988). The results from these studies are generally consistent (Gelfand, 1988; Yost, 2007). Inspection of Fig. 1 of Abel (1972), for example, shows that for stimuli of 10–50 ms long, ΔT is in the range of 3–5 ms. Since the duration differences found in the current study range from 0 to 19 ms, some of the differences do exceed the ΔT reported in Abel (1972). Certainly the noise/pure tone discrimination results may not generalize to lexical tones, but if the duration difference can indeed be perceived, duration could play a role in tone or gender identification.

Statistically, the ANOVA on duration showed a main effect of gender, $F(1,30)=50.12$, $p<0.001$; a main effect of tone, $F(3,90)=56.88$, $p<0.001$; and a significant gender $\times$tone interaction, $F(3,90)=4.26$, $p<0.01$. The gender effect is expected. In particular, since the male speakers on average have a lower F0 and the period is the inverse of F0, the male duration would be longer compared to the female duration. The longer duration and presumably the greater amount of acoustic information in the male stimuli, however, did not translate to higher response accuracy. On the contrary, the identification results showed the opposite pattern, with the female stimuli generating more accurate responses. Assuming the duration difference can be perceived, this indicates that longer duration does not necessarily provide higher information value. The contrast between the female and male stimuli, if it could be perceived, also suggests that duration could be useful for gender detection: shorter duration for females and longer duration for males. Even though the short-long contrast is also a relative one, exposure to the 128 stimuli could have provided the "context" for the judgment.

For the tone effect, all pairwise comparisons were significant except for the Tone 1–Tone 4 comparison. The result that the duration of Tones 2 and 3 is longer than that of Tones 1 and 4 can be readily predicted from their F0 contrasts. Assuming the duration difference can be perceived and using duration as an index of the amount of information, we would predict Tones 2 and 3 to be identified better than Tones 1 and 4. This prediction is only partially consistent with the identification results. In particular, Tone 3 was identified quite accurately, but Tone 2 was identified with the lowest accuracy. Finally, although the gender $\times$ tone interaction is statistically significant, inspection of the interaction plot did not reveal any notable patterns other than those already observed in the main effects.

Finally, as noted in Sec. I, Greenberg and Zee (1979) showed that the dynamic character of the pitch contour of a tone cannot be perceived with a duration shorter than 130 ms. Since no stimulus in the current study exceeded 49 ms in duration, it is unlikely the dynamic F0 contrasts among the four tones were perceptible even if there were F0 contour differences. This is additional evidence that the signal processing effectively neutralized the F0 contour difference among the four tones.

3. Amplitude

The average rms amplitude of the six glottal periods was obtained and the normalized rms amplitude adjusted by subtracting the frication noise amplitude was calculated. As noted, amplitude was calculated on values ranging from −1 to 1. Since the two measures essentially showed the same pattern, Fig. 3 shows only the rms amplitude. None of the ANOVAs on both measures showed an effect of tone or gender, indicating that no amplitude difference exists among the four tones or between the female and male stimuli. As noted, prior to the silencing procedure, peak amplitude had been equated across all intact syllables for the purpose of normalizing across speakers. Consequently, potential amplitude differences among the four tones, if any, may have been neutralized. Although Whalen and Xu (1992) showed that amplitude contour could be a cue for Mandarin tone identification, their results were based on signal-correlated noise stimuli of a syllable length when no F0 information was available. Whether amplitude could be of use of identification of brief stimuli with F0 information awaits further research.

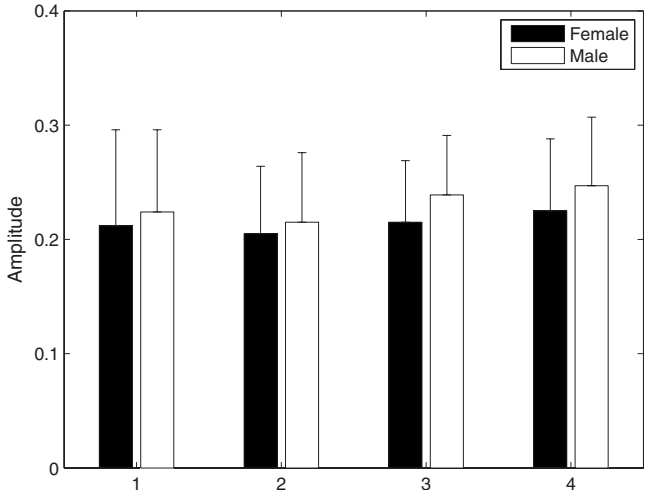


FIG. 3. Average rms amplitude (+1SD) of the six glottal periods in the tone stimuli.

4. Voice quality

Figure 4 shows the results of the three voice quality measures. For H1-H2, a measure of glottal open quotient, the ANOVA revealed no main effects but a significant gender $\times$tone interaction, $F(3,90)=3.89$, $p<0.05$. Overall, H1-H2 values for the tones were more distinct in the female stimuli than in the male stimuli. Specifically, the female data showed that Tones 1 and 4 have higher H1-H2 values than Tones 2 and 3. That is, the proportion of closure during a glottal cycle is higher for Tones 2 and 3 (the low-onset tones) and lower for Tones 1 and 4 (the high-onset tones).

For H1-A1, a measure of F1 bandwidth, the ANOVA showed significant main effects of gender, $F(1,30)=5.88$, $p<0.05$, and tone, $F(3,90)=8.63$, $p<0.0001$. There is also a significant tone$\times$gender interaction, $F(3,90)=5.64$, $p<0.005$. Overall, the female stimuli have higher H1-A1 values than the male stimuli, indicating wider F1 bandwidth and thus greater acoustic loss at the glottis. For the tone effect, all pairwise comparisons were significant except for the Tone 1–Tone 4 comparison and the Tone 2–Tone 3 comparison. That is, the F1 bandwidth is greater for Tones 1 and 4 than for Tones 2 and 3, indicating greater acoustic loss at the glottis for Tones 1 and 4. The interaction shows that H1-A1 values were more distinct among the tones in the female stimuli than in the male stimuli.

For H1-A3, an index of spectral tilt, the ANOVA revealed significant main effects of gender, $F(1,30)=5.87$, $p<0.05$, and tone, $F(3,90)=6.37$, $p<0.001$. Overall, the female stimuli have higher H1-A3 values than the male stimuli, indicating larger spectral tilt and thus greater acoustic loss at the glottis for the female speech. For the tone effect, all pairwise comparisons were significant except for the Tone 1–Tone 4 comparison and the Tone 2–Tone 3 comparison. That is, spectral tilt is greater for Tones 1 and 4 than for Tones 2 and 3, again indicating greater acoustic loss at the glottis for the high-onset tones.

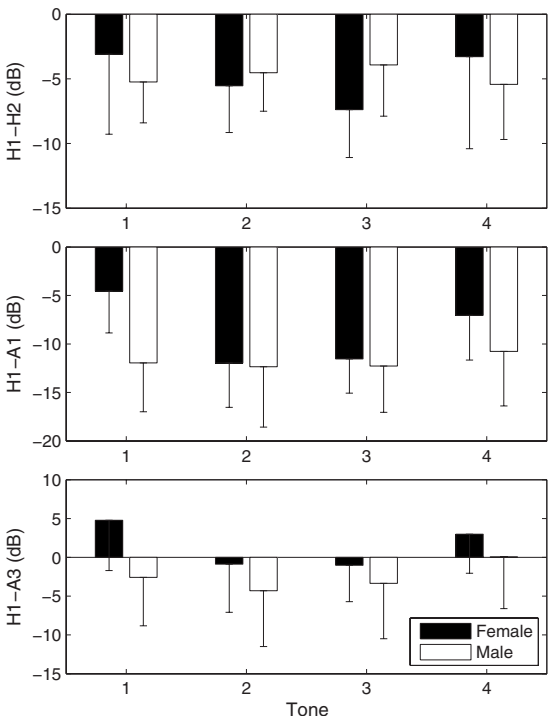


FIG. 4. Voice quality measures (−1SD) in the tone stimuli including open quotient (H1-H2), first formant bandwidth (H1-A1), and spectral tilt (H1-A3).

C. Relating acoustic and perceptual measures

The acoustic analyses showed that F0, duration, and two voice quality measures (F1 bandwidth and spectral tilt) were distinct between the high- and low-onset tones and between the female and male stimuli. These acoustic differences could be the basis for the findings from the identification experiment that listeners could identify the high- and low-onset tones above chance. In addition, the tone$\times$gender interaction found in duration, open quotient, and F1 bandwidth showed that the tonal contrasts were more distinct in the female stimuli. This result is also consistent with the finding that tone identification from the female stimuli was more accurate.

To evaluate the acoustic-perceptual relationship quantitatively, Pearson's correlation coefficients were derived between the seven acoustic measures (F0, duration, amplitude, normalized amplitude, open quotient, F1 bandwidth, and spectral tilt) and the two perceptual measures (accuracy of tone identification and accuracy of high-low F0 identification). Since the acoustic measures were obtained for each stimulus item, the accuracy of tone/F0 height identification was defined as the number of listeners who correctly identified the tone/F0 height of a stimulus. That is, the correlation analyses were conducted by items. Among the acoustic measures, a significant correlation was expected between F0 and duration because frequency and period are inversely related. Significant correlations were also expected among the three voice quality measures because they all involve estimation of the degree of acoustic loss at the glottis.

Table II shows a correlation matrix with all the acoustic and perceptual measures. Fisher's r to z transformation was carried out on the correlation coefficients to determine if the coefficients are significantly different from zero. The expected correlations between F0 and duration and among the

TABLE II. Correlations between acoustic and perceptual measures of tone identification. Statistically significant correlations are indicated by an asterisk (*). A total of 128 observations were used in the computation.

	Accuracy by F0 height	F0	Duration	Amplitude	Normalized amplitude	Open quotient	F1 bandwidth	Spectral tilt
Accuracy by tone	0.649*	0.225*	−0.085	−0.004	−0.015	−0.003	0.168	0.121
Accuracy by F0 height		0.431*	−0.306*	0.045	0.055	0.049	0.321*	0.216*
F0			−0.934*	−0.031	−0.014	0.083	0.451*	0.391*
Duration				0.055	0.032	−0.024	−0.396*	−0.358*
Amplitude					0.969*	−0.07	−0.079	0.038
Normalized amplitude						−0.116	−0.151	−0.005
Open quotient							0.492*	0.356*
F1 bandwidth								0.646*

three voice quality measures did turn out significant. In addition, F0 is significantly correlated with F1 bandwidth and spectral tilt. Duration is likewise correlated with the same voice quality measures. Taken together, duration, F1 bandwidth, and spectral tilt covary with F0, suggesting the listeners could have taken advantage of this covariation for F0 height estimation. Indeed, the accuracy of F0 height identification is significantly correlated with exactly these four acoustic measures.

Regression analyses were conducted to evaluate how well the acoustic measures could predict tone identification accuracy. To ensure an adequate sample size (Tabachnick and Fidell, 2001), the analyses were conducted across all four tones such that all 128 observations could be used. Since the ANOVAs on the acoustic measures reported earlier showed that the two amplitude measures did not distinguish the four tones, they are not likely to contribute to the tonal distinctions perceptually. Therefore, the two amplitude measures were not included in the regression models. Furthermore, since the correlation analyses noted earlier showed high correlation between F0 and duration, using both F0 and duration as predictors would introduce multicollinearity, which usually results in low tolerances and misleading beta weights (e.g., Cohen, 1996). It is more reasonable to enter F0 instead of duration based on the assumption that the duration difference may not be perceptible.

Two regression models were built: one with tone identification accuracy as the dependent variable and the other with tone classification based on F0 height as the dependent variable (high-onset tones include Tones 1 and 4; low-onset tones include Tones 2 and 3). In both models, the predictors included F0 and the three voice quality measures (open quotient, F1 bandwidth, and spectral tilt). In the first model, the four predictors accounted only for 6.1% of the variance and none of the regression coefficients turned out significant. In the second model, the predictors accounted for 21.1% of the variance. Specifically, F0 was the best predictor ($\beta=0.351$, $p<0.0005$). H1-A1 was the second best predictor even though the coefficient was not significant ($\beta=0.226$, $p=0.06$). These results indicate that F0 is the only statistically significant predictor of tone classification based on F0 height, suggesting that the listeners were able to use F0 height information for the classification. This observation is also consistent with the hypothesis that listeners were using the covariation among F0 and two of the voice quality measures (F1 bandwidth and spectral tilt) found in the correlation analyses for F0 height estimation.

IV. GENERAL DISCUSSION

The major finding from the present study is that tone identification responses to isolated, multispeaker Mandarin stimuli with six glottal periods were contingent on the tones of the stimuli, indicating that the tone responses were not random. Analyses by individual tones revealed the same result (except for Tone 2 stimuli in male speech), further suggesting that tone identification from the brief stimuli exceeded chance. Acoustic analyses showed that dynamic F0 contrasts among the tones were neutralized. Consequently, dynamic F0 information could not have been useful. No context was given; therefore the listeners could not have benefited from any syllable-external information. Familiarity with the speakers' voices is unlikely to have helped because each stimulus was presented only once and the listeners heard each speaker only four times.

The acoustic analyses showed F0 height contrasts between the high-onset tones (Tones 1 and 4) and the low-onset tones (Tones 2 and 3). This finding is consistent with the perceptual results that tones sharing a similar F0 height tend to be confused, and that F0 height identification accuracy exceeded chance. The issue remained, however, of how a particular F0 can be judged as high or low given the speaker variability. The acoustic analyses showed duration, F1 bandwidth (H1-A1), and spectral tilt (H1-A3) were all distinct between the high- and low-onset tones, suggesting that these three acoustic measures covary with F0 to provide information about F0 height. This observation was verified by the correlation analyses. The use of this covariation for perception was further supported by the significant correlation between F0 height identification accuracy and the four acoustic measures (F0, duration, F1 bandwidth, and spectral tilt), even though F0 itself was the only statistically significant predictor of tone classification by F0 height in the regression analyses. Nonetheless, these results suggest that the covariation among F0, duration, F1 bandwidth and spectral tilt was exploited for F0 height estimation, which served as the basis

for tone identification from these brief stimuli without context, dynamic F0 information, or prior exposure to the speakers' voices.

The speaker gender effect found in the perceptual accuracy measure and in the acoustic measures of F0, duration, F1 bandwidth, and spectral tilt also suggest that gender detection may be implicated in the tone identification process. The differences between female and male speakers in F0 (Peterson and Barney, 1952) and voice quality measures (Hanson, 1997; Hanson and Chuang, 1999) have been reported in the literature. However, F0 height information alone will not be useful because an F0 value could indicate a low tone for a female speaker or a high tone for a male speaker, as the current acoustic results showed (Fig. 2). In contrast, the voice quality measures consistently distinguished the female from the male voice. Since gender differences have been shown with these measures (Hanson, 1997; Hanson and Chuang, 1999), the listeners could have used these voice quality differences to determine whether a speaker was female or male, and then evaluated F0 height based on the gender information.

Honorof and Whalen (2005) offered a similar, gender-based interpretation of their finding that English-speaking listeners were able to locate an F0 reliably within a speaker's F0 range without context or prior exposure to a speaker's voice. In particular, F0 location judgments were positively correlated with the female and male F0 range values. This result suggests that F0 height may be estimated based on templates of F0 ranges for female and male speakers ("population tessitures") stored in a listener's memory. As noted in the introduction, a similar idea has also been proposed by Deutsch and co-workers (Deutsch, 1991; Deutsch, *et al.*, 1999, 2004; Deutsch, *et al.*, 1990). In particular, listeners acquire pitch class templates of prevalent speaking F0s from experience with the speakers in their linguistic community, and these templates are used for both speech production and speech perception.

For this gender-based strategy to work, the speaking F0 range in the linguistic community has to be fairly constrained. The F0 difference among the tones must also exceed the F0 variability among the speakers of a particular gender. Dolson (1994) reported that the average speaking F0 is within plus/minus three semitones for either gender within a linguistic community. The acoustic data from the current study also provide some support for this strategy. In particular, despite the considerable number of speakers used (16 females and 16 males), the error bars in Fig. 2 show that the F0 height for a particular tone does not vary too much within a gender group, particularly for the female speakers. In addition, the high-low tonal distinction is fairly well maintained within either gender group, although more prominently for the female speakers. With these reasonably constrained and distinct F0 ranges for both gender groups, gender detection may be used as a first pass before F0 height can be evaluated based on the gender-specific F0 templates.

However, this account does not specify how gender is actually detected. The acoustic data in this study suggest voice quality measures such as F1 bandwidth and spectral tilt may be implicated (Hanson, 1997; Hanson and Chuang, 1999). The covariation between these measures and F0 as well as the significant correlations between these voice quality measures and F0 height detection accuracy also support the perceptual role of the voice quality measures in gender detection. However, since the listeners in the current study were not asked to identify the gender of the speakers, it is not known if they were indeed able to identify gender from the stimuli. Another potential source of gender identification is formant frequency cues to vocal tract length. Bachorowski and Owren (1999) showed speaker gender can be classified acoustically with 92% accuracy from a database of a vowel produced by 125 speakers. Ingemann (1968) also showed that speaker gender can be identified with 75% accuracy from isolated [s] tokens produced by 14 speakers. The stimuli used in the current study happened to include both the fricative and some vowel information. The listeners could have used both the glottal source characteristics (indicated by the voice quality measures) and vocal tract filter information (represented by resonant frequencies from the fricative or vowel) to come up with a gender decision before using the gender information to make F0 height judgments. Further research is needed to test this hypothesis, but the available evidence seems to suggest that gender detection is involved.

The findings from the current study can also be compared to the literature on tone identification from incomplete acoustic input. The stimuli in the current study are generally shorter than the shortest fragment used in previous gating studies (Tseng, 1981; Whalen and Xu, 1992; Wu and Shu, 2003; Yang, 1992). Inspection of Whalen and Xu's (1992) data showed Tones 1 and 4 can be identified with 75% or higher accuracy with 80 ms of input, but Tones 2 and 3 can only be identified at 50% and 40% correct with 100 ms of input. However, if the high- versus low-onset classification is applied instead of specific tones, the data indicate that the high-low distinction can be perceived right at the first gate (40 ms). Wu and Shu (2003) found that a greater amount of acoustic input is needed for Tone 2 identification compared to the other tones (Tone 1: 156 ms; Tone 2: 178 ms; Tone 3: 151 ms; Tone 4: 148 ms). When the absolute values were converted to a percentage of the entire stimulus, however, Tone 2 no longer needed the longest input for identification. Rather, there was no difference among the four tones except for Tone 3 (Tone 1: 56%; Tone 2: 59%; Tone 3: 42%; Tone 4: 60%). These percentages are consistent with Tseng (1981) and Yang (1992), who showed that isolated Mandarin tones can be identified on the basis of the first half of a vowel/syllable. By most accounts (e.g., Whalen and Xu, 1992), half a vowel/syllable is the point where the F0 contours of the four tones become acoustically distinct. The conclusion to be drawn from existing evidence, then, is that the identification of specific tones requires F0 contour information, which requires about half of a vowel/syllable. Tone classification based on F0 height, in contrast, needs very little input. The current study further showed that F0 height can be detected from multispeaker stimuli when no context, dynamic F0, or prior exposure to speaker voice is available.

More broadly, these findings have implications for the efficiency of processing prosodic information in spoken lan-

guage comprehension. In particular, the processing of tonal information is normally considered to take place later than the processing of segmental information due to the temporally distributed nature of tones (Cutler and Chen, 1997). Previous studies showed tones presented in context can be identified quite early (Lee, 2000; Xu, 1994, 2004). The current findings further suggest that multispeaker, isolated tones can be processed in terms of their F0 height information fairly early as well. These findings are consistent with gating studies showing that lexical stress and lexical pitch accent can be identified with the input of one syllable (Cutler amd Otake, 1999; van Heuven, 1988, cited in Cutler and Donselaar, 2001; Cutler, *et al.*, 2007). Since stress and pitch accent contrasts are relative and presumably involve the comparison of at least two syllables, it seems counterintuitive to be able to detect stress or pitch accent based only on one syllable. Nonetheless, the stimuli in these studies were recorded by one speaker and were presented with a carrier phrase. Familiarity with speaker voice and the presence of context could have contributed to the evaluation of the acoustic cues available in the stimuli. It will be of interest to evaluate whether listeners can identify these prosodic contrasts in isolated stimuli produced by multiple speakers in the same way as they processed brief lexical tone stimuli in the current study.

ACKNOWLEDGMENTS

I am grateful to Allard Jongman and two anonymous reviewers for their helpful comments. I also thank Z. S. Bond for discussions, Ya-Ting Shih for assistance in administering the perception experiment, and Ning Zhou, Alex Sergeev, Anne Marie Christy, and Gayatri Ram for assistance in data processing. This research was supported in part by a faculty development fund from the School of Hearing, Speech and Language Sciences at Ohio University.

Abel, S. M. (**1972**). "Duration discrimination of noise and tone bursts," J. Acoust. Soc. Am. **51**, 1219–1223.
Abramson, A. S. (**1972**). "Tonal experiments with whispered Thai," in *Papers in Linguistics and Phonetics to the Memory of Pierre Delattre*, edited by A. Valdman (Mouton, The Hague), pp. 31–44.
Abramson, A. S. (**1978**). "Static and dynamic acoustics in distinctive tones," Lang Speech **21**, 319–325.
Andruski, J. E. (**2006**). "Tone clarity in mixed pitch/phonation-type tones," J. Phonetics **34**, 388–404.
Andruski, J. E., and Ratliff, M. (**2000**). "Phonation types in production of phonological tones: The case of Green Mong," J. Int. Phonetic Assoc. **30**, 37–61.
Bachorowski, J.-A., and Owren, M. J. (**1999**). "Acoustic correlates of talker sex and individual talker identity are present in a short vowel segment produced in running speech," J. Acoust. Soc. Am. **106**, 1054–1063.
Blicher, D. L., Diehl, R. L., and Cohen, L. B. (**1990**). "Effects of syllable duration on the perception of the Mandarin tone 2/tone 3 distinction: Evidence of auditory enhancement," J. Phonetics **18**, 37–49.
Childers, D. G., and Wu, K. (**1991**). "Gender recognition from speech. Part II: Fine analysis," J. Acoust. Soc. Am. **90**, 1841–1856.
Cohen, B. H. (**1996**). *Explaining Psychological Statistics* (Brooks/Cole, Pacific Grove).
Cutler, A., and Chen, H.-C. (**1997**). "Lexical tone in Cantonese spoken word processing," Percept. Psychophys. **59**, 165–179.
Cutler, A., and Donselaar, W. (**2001**). "Voornaam is not a homophone: Lexical prosody and lexical access in Dutch," Lang Speech **44**, 171–195.
Cutler, A., and Otake, T. (**1999**). "Pitch accent in spoken-word recognition in Japanese," J. Acoust. Soc. Am. **105**, 1977–1988.
Cutler, A., Wales, R., Cooper, N., and Janssen, J. (**2007**). "Dutch listeners' use of suprasegmental cues to English stress," Proceedings of the 16th International Congress of Phonetic Sciences, pp. 1913–1916.
Deutsch, D. (**1991**). "The tritone paradox: An influence of language on music perception," Music Percept. **8**, 335–347.
Deutsch, D., Henthorn, T., and Dolson, M. (**1999**). "Absolute pitch is demonstrated in speakers of tone languages," J. Acoust. Soc. Am. **106**, 2267.
Deutsch, D., Henthorn, T., and Dolson, M. (**2004**). "Absolute pitch, speech, and tone language: Some experiments and a proposed framework," Music Percept. **21**, 339–356.
Deutsch, D., North, T., and Ray, L. (**1990**). "The tritone paradox: Correlate with the listener's vocal range for speech," Music Percept. **7**, 371–384.
Dolson, M. (**1994**). "The pitch of speech as a function of linguistic community," Music Percept. **11**, 321–331.
Dooley, G. J., and Moore, B. C. J. (**1988**). "Duration discrimination of steady and gliding tones: A new method for estimating sensitivity to rate of change," J. Acoust. Soc. Am. **84**, 1332–1337.
Fox, R. A., and Qi, Y.-Y. (**1990**). "Context effects in the perception of lexical tones," J. Chin. Linguist. **18**, 261–284.
Fox, R. A., and Unkefer, J. (**1985**). "The effect of lexical status on the perception of tone," J. Chin. Linguist. **13**, 69–90.
Gandour, J. (**1983**). "Tone perception in Far Eastern languages," J. Phonetics **11**, 149–175.
Gandour, J. T., and Harshman, R. A. (**1978**). "Crosslanguage differences in tone perception: A multidimensional scaling investigation," Lang Speech **22**, 1–33.
Gelfand, S. A. (**1998**). *Hearing: An Introduction to Psychological and Physiological Acoustics* (Dekker, New York).
Gottfried, T. L., and Suiter, T. L. (**1997**). "Effects of linguistic experience on the identification of Mandarin Chinese vowels and tones," J. Phonetics **25**, 207–231.
Greenberg, S., and Zee, E. (**1979**). "On the perception of contour tones," UCLA Working Papers in Phonetics **45**, 150–164.
Grosjean, F. (**1996**). "Gating," Lang. Cognit. Processes **11**, 597–604.
Hanson, H. M. (**1997**). "Glottal characteristics of female speakers: Acoustic correlates," J. Acoust. Soc. Am. **101**, 466–481.
Hanson, H. M., and Chuang, E. S. (**1999**). "Glottal characteristics of male speakers: Acoustic correlates and comparison with female data," J. Acoust. Soc. Am. **106**, 1064–1077.
Honorof, D. N., and Whalen, D. H. (**2005**). "Perception of pitch location within a speaker's F0 range," J. Acoust. Soc. Am. **117**, 2193–2200.
Howell, D. C. (**1999**). *Fundamental Statistics for the Behavioral Sciences* (Brooks/Cole, Pacific Grove, CA).
Howie, J. M. (**1976**). *Acoustical Studies of Mandarin Vowels and Tones* (Cambridge University Press, Cambridge).
Huffman, M. (**1987**). "Measures of phonation type in Hmong," J. Acoust. Soc. Am. **81**, 495.
Ingemann, F. (**1968**). "Identification of the speaker's sex from voiceless fricatives," J. Acoust. Soc. Am. **44**, 1142–1144.
Johnson, K. A. (**2005**). "Speaker normalization in speech perception," *The Handbook of Speech Perception*, edited by D. B. Pisoni and R. E. Remez (Blackwell, Malden, MA), pp. 363–389.
Ladefoged, P. (**1996**). *The Sounds of the World's Languages* (Blackwell, Malden, MA).
Lass, N. J., Mertz, P. J., and Kimmel, K. L. (**1978**). "Effect of temporal speech alterations on speaker race and sex identifications," Lang Speech **21**, 279–290.
Leather, J. (**1983**). "Speaker normalization in perception of lexical tone," J. Phonetics **11**, 373–382.
Lee, C.-Y. (**2000**). "Lexical tone in spoken word recognition: A view from Mandarin Chinese," thesis, Brown University, Providence, RI.
Lee, C.-Y., Tao, L., and Bond, Z. S. (**2008**). "Identification of acoustically modified Mandarin tones by native listeners," J. Phonetics **36**, 537–563.
Lin, T., and Wang, W. S.-Y. (**1984**). ""Shengdiao ganzhi wenti" (The issue of tone perception)," Zhongguo Yuyan Xuebao (Bull. Chinese Linguistics) **2**, 59–69.
Liu, S., and Samuel, A. G. (**2004**). "Perception of Mandarin lexical tones when F0 information is neutralized," Lang Speech **47**, 109–138.
Massaro, D. W., Cohen, M. M., and Tseng, C.-Y. (**1985**). "The evaluation and integration of pitch height and pitch contour in lexical tone perception in Mandarin Chinese," J. Chin. Linguist. **13**, 267–289.
Mertus, J. A. (**2000**). *The Brown Lab Interactive Speech System* (Brown University, Providence, RI).
Moore, C. B., and Jongman, A. (**1997**). "Speaker normalization in the perception of Mandarin Chinese tones," J. Acoust. Soc. Am. **102**, 1864–1877.
Nygaard, L. C., and Pisoni, D. B. (**1998**). "Talker-specific learning in speech

perception," Percept. Psychophys. **60**, 355–376.
Palmeri, T. J., Goldinger, S. D., and Pisoni, D. B. (**1993**). "Episodic encoding of voice attributes and recognition memory for spoken words," J. Exp. Psychol. Learn. Mem. Cogn. **19**, 309–328.
Peterson, G. E., and Barney, H. L. (**1952**). "Control methods used in a study of vowels," J. Acoust. Soc. Am. **24**, 175–184.
Sinnott, J. M., Owren, M. J., and Peterson, M. R. (**1987**). "Auditory duration discrimination in Old World monkeys (*Macaca*. *Cercopithecus*) and humans," J. Acoust. Soc. Am. **82**, 465–470.
Stevens, K. N. (**1998**). *Acoustic Phonetics* (MIT, Cambridge, MA).
Stevens, K. N., Li, Z., Lee, C.-Y., and Keyser, J. (**2004**). "A note on Mandarin fricatives and enhancement," in *From Traditional Phonology to Modern Speech Processing*, edited by G. Fant, H. Fujisaki, J. Cao, and Y. Xu (Foreign Language Teaching and Research, Beijing).
Swerts, M., and Veldhuis, R. (**2001**). "The effect of speech melody on voice quality," Speech Commun. **22**, 297–303.
Tabachnick, B. G., and Fidell, L. S. (**2001**). *Using Multivariate Statistics*, 4th ed. (Allyn and Bacon, Boston).
Tseng, C.-Y. (**1981**). "An acoustic phonetic study on tones in Mandarin Chinese," Ph.D. thesis, Brown University, Providence, RI.
Heuven, V. J. (**1988**). "Effects of stress and accent on the human recognition of word fragments in spoken context: Gating and shadowing," *Proceedings of Speech '88, seventh FASE Symposium*, Edingburgh, pp. 811–818.
Wang, H. (**1986**). *A Frequency Dictionary of Modern Chinese* (Beijing Language Institute, Beijing).
Whalen, D. H., and Xu, Y. (**1992**). "Information for Mandarin tones in the amplitude contour and in brief segments," Phonetica **49**, 25–47.
Wong, P. C. M., and Diehl, R. L. (**2003**). "Perceptual normalization for inter- and intra-talker variation in Cantonese level tones," J. Speech Lang. Hear. Res. **46**, 413–421.
Wu, N., and Shu, H. (**2003**). "The gating paradigm and spoken word recognition of Chinese," Acta Psychologica Sinica **35**, 582–590.
Xu, Y. (**1994**). "Production and perception of coarticulated tones," J. Acoust. Soc. Am. **95**, 2240–2253.
Xu, Y. (**1997**). "Contextual tonal variations in Mandarin," J. Phonetics **25**, 61–83.
Xu, Y. (**2004**). "Understanding tone from the perspective of production and perception," Language and Linguistics **5**, 757–797.
Yang, S. (**1992**). "A preliminary study on the perceptual center of tones in Standard Chinese," Acta Psychologica Sinica **3**, 247–253.
Yost, W. A. (**2007**). *Fundamentals of Hearing: An Introduction* Elsevier, Burlington, MA.

Language experience and consonantal context effects on perceptual assimilation of French vowels by American-English learners of French[a)]

Erika S. Levy[b)]
Program in Speech and Language Pathology, Department of Biobehavioral Sciences, Teachers College, Columbia University, 525 West 120th Street, Box 180, New York, New York 10027

(Received 29 June 2007; revised 18 August 2008; accepted 14 November 2008)

Recent research has called for an examination of perceptual assimilation patterns in second-language speech learning. This study examined the effects of language learning and consonantal context on perceptual assimilation of Parisian French (PF) front rounded vowels /y/ and /œ/ by American English (AE) learners of French. AE listeners differing in their French language experience (no experience, formal instruction, formal-plus-immersion experience) performed an assimilation task involving PF /y, œ, u, o, i, ɛ, a/ in bilabial /rabVp/ and alveolar /radVt/ contexts, presented in phrases. PF front rounded vowels were assimilated overwhelmingly to back AE vowels. For PF /œ/, assimilation patterns differed as a function of language experience and consonantal context. However, PF /y/ revealed no experience effect in alveolar context. In bilabial context, listeners with extensive experience assimilated PF /y/ to /ʲu/ less often than listeners with no or only formal experience, a pattern predicting the poorest /u-y/ discrimination for the most experienced group. An "internal consistency" analysis indicated that responses were most consistent with extensive language experience and in bilabial context. Acoustical analysis revealed that acoustical similarities among PF vowels alone cannot explain context-specific assimilation patterns. Instead it is suggested that native-language allophonic variation influences context-specific perceptual patterns in second-language learning. © *2009 Acoustical Society of America.* [DOI: 10.1121/1.3050256]

PACS number(s): 43.71.Hw, 43.71.An, 43.71.Es, 43.70.Kv [PEI] Pages: 1138–1152

I. INTRODUCTION

The present study investigated the effects of second-language (L2) experience and consonantal context on the perceptual assimilation of Parisian French (PF) front rounded vowels by American English (AE) L2-learners of French. The class of vowels investigated in this study exemplifies the difficulty individuals may encounter upon learning L2 segments. Front rounded vowels, such as PF /y/ (vu /vy/ "seen") and /œ/ (in vœu /vœ/[1] "wish"), are produced with rounded lips, but unlike AE rounded vowels, the tongue body is described as being forward in the oral cavity (Tranel, 1987). The second and third formant frequencies (F2 and F3) of front rounded vowels are lower than those of front unrounded vowels, primarily because the length of the oral cavity is increased through lip rounding. Front rounded vowels are arguably nonexistent in AE and certainly not phonemic in the language (Gottfried, 1984). Findings conflict regarding whether AE listeners perceive (and produce) front rounded vowels in a nativelike manner (e.g., Best *et al.*, 1996; Flege, 1987; Flege and Hillenbrand, 1984; Gottfried, 1984; Levy and Strange, 2008; Polka, 1995; Rochet, 1995; Stevens *et al.*, 1969; Strange *et al.*, 2005), as will be discussed below.

A. Models of non-native and second-language speech perception

Two predominant models of cross-language speech perception, the Perceptual Assimilation Model for naïve listeners (PAM) and L2-learners (PAM-L2) (Best, 1995; Best and Tyler, 2007) and the Speech Learning Model (SLM) (Flege, 1995), posit that the perceived similarity of non-native segments to native categories is crucial to determining the difficulties listeners will encounter in their non-native language. In proposing the PAM-L2, Best and Tyler (2007) explored whether the principles involved in naïve learners' patterns of perceptual assimilation may be extended to perceptual patterns in L2-learning. Unlike the SLM, which focuses more on the L2 perception-production link, the PAM makes testable predictions in the realm of perception, specifically the relationship between assimilation and discrimination of non-native sounds (Harnsberger, 2001). Because the present study drew on findings from identification and discrimination studies to create testable predictions regarding assimilation in L2-learners, the PAM-L2 framework is discussed in greater depth here than the SLM.

The PAM (Best, 1995) posits that novel segments may be perceptually assimilated to native categories as "good" to "poor" instances along a continuum. In single-category assimilation, for example, contrastive non-native segments are both assimilated as good instances of the same category in the native language. In the category-goodness type, two non-native segments are assimilated into the same native cat-

[a)]This work was presented at the ASA Conference held in June 2006 in Providence, Rhode Island, and was previously entitled "Effects of language experience and consonantal context on perception of French front, rounded vowels by adult American English learners of French."

[b)]Electronic mail: elevy@tc.columbia.edu

egory, but one is a "better instance" than the other. If each non-native segment is assimilated to a different native category, this constitutes two-category assimilation. On the other hand, a segment may be "uncategorizable," that is, within the phonological space of the native language but outside any actual native category, whereas another may be similar to an AE category (uncategorized-categorized). And finally, both segments may be uncategorizable within the native-language inventory.

Perceptual assimilation patterns predict naïve listeners' ability to differentiate foreign speech sounds according to the PAM (Best, 1995). In presenting the PAM-L2, Best and Tyler (2007) demonstrated how predictions in the PAM framework may also predict success in L2 perceptual learning. According to the PAM, two non-native segments assimilating to a single native category are expected to be the most difficult contrasts to discriminate, especially if they are equally good instances of the category. Segments assimilating to separate native categories or as better and worse exemplars or as uncategorizable and categorizable exemplars will yield more accurate discrimination than would single-category assimilation. The PAM-L2 posits that in category-goodness assimilation, it is unlikely that a new category would be learned for the less deviant L2 phone. If both L2 phones are uncategorizable, learning patterns would depend, in part, on whether the L2 phones are perceived as similar to L1 phones that approximate each other in phonological space. Experiments inspired by the PAM have focused mostly on naïve listeners (e.g., Best *et al.*, 1988, 1996; Strange *et al.*, 2001). Few studies (e.g., Guion *et al.*, 2000, on consonant perception) have examined L2-learners' assimilation patterns.

Unlike the PAM (Best, 1995) for naïve listeners, Flege's (1995) SLM was designed specifically to explain the difficulties more experienced language learners face when learning L2 contrasts, with emphasis on the problem of inaccurate production (i.e., accentedness) by L2-learners and changes in production with exposure to L2. The PAM and the SLM concur that when non-native speech sounds are identified with one particular native category, discrimination difficulties ensue. However, the SLM specifies that it is at an allophonic level that L1 and L2 sounds relate, although the consequences for cross-language speech perception are not defined (Harnsberger, 2001). According to the SLM, the more perceptually dissimilar an L2 segment is from its closest L1 segment, the greater the likelihood that their phonetic differences will be discerned. New L2 categories may be established if at least some of the phonetic differences between L1 and the L2 speech sounds are discerned.

B. Previous perceptual research on French front rounded vowels

Several studies have demonstrated the effects of language background on perception of French rounded vowels. Rochet (1995) asked Canadian French, native Canadian English, and Brazilian Portuguese listeners to identify steady-state synthetic vowels on a high vowel continuum in which the second-formant (F2) frequency varied between 2500 and 500 Hz, as /i/ or /u/ (or /i-y-u/ for Canadian French listeners). Stimuli identified as /y/ by Canadian French listeners were most frequently identified (and produced) as /u/ by Canadian English listeners and as /i/ by Brazilian Portuguese listeners. Clearly, the participants' language background influenced to which native-language categories (/i/ or /u/) the segment (/y/) assimilated.

Using a perceptual assimilation task, Best *et al.* (1996) found that 8 of 13 naïve AE listeners assimilated Bretagne French front rounded vowels in /sœ-sy/ in a two-category pattern, 3 in an uncategorizable-categorizable pattern, and 2 in a category-goodness pattern. In a categorial AXB discrimination task, the listeners discriminated multiple tokens of the French syllables with fewer than 5% errors, consistent with the PAM's predictions of very good to excellent discrimination for 11 of the 13 listeners' assimilation patterns revealed. Consistent with Best *et al.*'s (1996) finding of little difficulty for naïve AE listeners on contrasts involving /y/, Flege and Hillenbrand (1984) reported high accuracy (90% correct) in identifying native French /ty/ by experienced AE listeners in a paired-comparison (/tu/ and /ty/) task, suggesting that the French /y-u/ contrast can be differentiated accurately by AE speakers, at least when presented in paired syllables. Flege and Hillenbrand (1984) referred to the French front rounded vowel /y/ as an example of a "new" L2 phone to American learners of French, as listeners are able to distinguish it from French /u/ in perception and production. However, Flege (1987) stated that in certain phonetic contexts, AE /u/ has an /y/-like phonetic quality, a topic he suggested ought to be explored. According to Flege (1987), AE learners of French may initially classify /y/ as /u/, but with experience, they will recognize that /y/ is not a realization of English /u/.

C. Context-dependent perception

Vowels are produced differently as a function of consonants surrounding them (Hillenbrand *et al.*, 2001) and coarticulatory variation differs from language to language (Strange *et al.*, 2007). It follows that cross-language perceptual patterns may also vary as a function of phonetic context (Bohn and Steinlen, 2003). Consonantal context (i.e., bilabial, alveolar, or velar) affects naïve Japanese listeners' assimilation of non-native vowels (Strange *et al.*, 2001). Moreover, naïve AE listeners' assimilation of North German front rounded vowels varies as a function of prosodic and consonantal context (Strange *et al.*, 2004a, 2007).

If learning an L2 involves learning to produce and perceive coarticulatory variation in the language, the ability to perceive L2 segments may also change with L2 experience and vary as a function of phonetic context. A seminal study by Gottfried (1984) examined the effect of syllabic context on perception of French vowels by AE listeners with and without L2 experience. AE listeners who spoke French discriminated vowels in /tVt/ context more accurately than those who spoke no French; the groups did not perform differently for vowels in isolation. Vowel perception may thus vary as a function of experience and syllabic context.

Levy and Strange (2008) extended Gottfried's (1984) study, focusing on the consonantal context effects on the perception of PF /y/, /œ/, /u/, and /i/. In this (cross-speaker[2])

categorial discrimination experiment, PF vowels were presented in /rabVp/ and /radVt/ bisyllables embedded in AXB triads of the phrase "neuf /raCVC/ à des amis" ("nine /raCVC/ to some friends"). Experience and context affected AE listeners' perceptual accuracy. Experienced listeners made fewer errors than the inexperienced group for three vowel pairs (/i-y/, /u-œ/, and /y-œ/). For /u-y/, no significant difference was found between the naïve group (24% errors) and the experienced group (30% errors) in their discrimination, despite the experienced group's many years of French instruction and immersion. The inexperienced group made more errors on the /u-y/ contrast in alveolar context than in bilabial context, but more errors on the /i-y/ contrast in bilabial context than in alveolar. The experienced group, on the other hand, did not reveal a significant context effect on discrimination of the /u-y/ contrast. For all contrasts except /i-y/ (where the reverse was true), for the inexperienced group, discrimination scores were higher in bilabial than in alveolar context. No significant context effect was revealed for the experienced group, although they showed a trend toward more errors in alveolar context. Thus, with language experience, L2-learners may begin to perceive segments on a more abstract level, less affected by acoustic variation.

Discrimination results from Levy and Strange (2008) suggested that L2 vowels may be perceptually assimilated to native vowel categories in a consonantal-context-dependent manner that changes with language experience. The present study explored the effects of consonantal context on *perceptual assimilation* patterns as listeners have more extensive L2 experience, thereby examining the extension of the PAM's (Best, 1995) theoretical domain to L2-learning. Perceptual assimilation of PF front rounded vowels by three groups of AE listeners was investigated: listeners with no French experience (NoExp), listeners with formal (i.e., classroom) French training (ModExp), and listeners with several years of formal French training and immersion experience (HiExp). The ModExp group was included to represent L2 vowel perception by the majority of students enrolled in United States schools, who begin studying a foreign language at an average age of 12 (Pufahl *et al.*, 2001) and who do not have immersion experience. Phrases were used rather than vowels or words in isolation in order to tap into linguistic categorization processes employed in the perception of continuous speech.

The following questions were asked, and predictions were made:

(1) Are PF front rounded vowels perceptually assimilated to AE front, unrounded or back, rounded vowels by L2-learners of French? Based on the PAM-L2's (Best and Tyler, 2007) claim that discrimination is typically poor when two L2 phones assimilate to a single L1 category, the expectation here was that, overall, the L2-learners would assimilate PF front rounded vowels to AE back rounded vowels more often than to AE front unrounded vowels.
(2) Does perceptual assimilation of PF front rounded vowels by L2-learners vary as a function of language experience? It was predicted that assimilation patterns for PF /œ/ would generally differ with language experience for these L2-learners but that the PF /y/ would be perceived similarly across groups. Goodness ratings were expected to decrease with experience based on the PAM-L2's (Best and Tyler, 2007) claim that L2-learning involves, to the extent possible, refining the learner's perception of higher order invariants in the L2. As no previous study had compared assimilation of PF vowels by AE listeners with formal French experience versus formal training plus immersion experience, it remained to be seen whether the ModExp group would demonstrate assimilation patterns more like NoExp or HiExp listeners.
(3) Does perceptual assimilation of PF front rounded vowels by L2-learners vary as a function of consonantal context and (4) are there interactions among perceptual assimilation of particular vowels, language experience, and context? Based on the PAM-L2's (Best and Tyler, 2007) claim of an association between perceptual assimilation and discrimination patterns in L2-learners and on Levy and Strange's (2008) observation that only beginning L2-learners showed context-specific patterns in discrimination, more assimilation of PF front rounded vowels to back vowels in alveolar context (than in bilabial context) was predicted for naïve listeners, with less of a context effect expected with more extensive language experience.

II. METHOD

A. Participants

A total of 43 native AE speakers volunteered as listeners for the tests. Participants were recruited from Columbia University, the Alliance française, a French conversation club, and the web.[3] Participants passed a bilateral hearing screening at 20 dB at 500, 1000, 2000, and 4000 Hz. Data from four of these participants were discarded for the following reasons: One participant failed the hearing screening, and two exceeded the criteria for errors in the familiarization task. Another participant revealed a language history that did not meet the inclusion criteria for the "moderate experience" group (of no more than two years of French in high school and of minimal French immersion) or for the "high experience" group (of having spent more than one year in a French-speaking country). Thus, data from 39 AE participants (13 in each language group) were analyzed.

All participants were born and raised in English-speaking households in the United States. None had had more than a year of instruction in any language with front rounded vowels aside from French. Of the 39 AE listeners, 13 were native AE speakers living in New York City, ages 20–40 years, with minimal French experience (NoExp group), i.e., no French instruction and little interaction with French speakers.

The second group, listeners with moderate French experience (ModExp group), consisted of 13 native speakers of AE living in New York City, ages 22–37 years, who had had formal French training (i.e., they had attended French classes) but minimal immersion in French. They had begun learning French in school no earlier than age 12 years

(mean=16.1, SD=2.8) and had received a mean of 3 years of instruction in French (range=2–4 years, SD=0.8) in high school and college. Their instruction occurred from .5 to 6 years before testing (mean=3.3 years, SD=1.6). They had spent no more than 5 months in a French-speaking country and had not been speaking French around the time of testing.

The third group, AE listeners with extensive French experience (HiExp), consisted of 13 native speakers of AE, ages 20–61 years,[4] who had extensive formal training and immersion experience in French and were using French regularly at the time of testing (range=2 h/week, 100% of the time, median=15 h/week). They had had a mean of 8 years of French instruction (range=5–13 years, SD=2.4), which began no earlier than age 12 (mean age of beginning learning: 14 years, SD=1.6). They had spent at least a year in a French-speaking country as adults (range=1 year to 16 years, median=1.4 years), had spoken French regularly in their professions (e.g., as teachers of French, translators, and foreign business consultants), and were currently using French. Three were living in France at the time of testing and were visiting New York. One had lived in France until 3 weeks before testing.

B. Stimulus materials

1. Recording, editing, and verification procedures

Recordings of three female adult native PF speakers were made in an IAC chamber. These speakers had resided in the U.S. for less than a year. They were instructed to read a list consisting of nine PF vowels, blocked by /rabVp/ or /radVt/ context, in the sentence: "J'ai dit neuf /raCVC/ à des amis" (I said nine /raCVC/ to some friends). They read four repetitions of each list and were asked to produce the sentences as if conversing with a native speaker.

The experimenter (a native speaker of French and English) conversed in French with the speakers. A Shure microphone fed signals through an Earthworks microphone preamp to a Soundblaster Live Wave sound card of a Dell Dimension XPS B800 computer. The stimuli were digitized with a sample rate of 22,050 Hz, 16 bit resolution, on a mono channel, using SoundForge™ software. The experimenter chose the three "best" instances of each vowel. A native French speaker judged whether the tokens were typical exemplars of the target vowel. Two tokens were replaced because the intonation on these stimuli did not match the others. The digital files were edited so that only the phrases "neuf /rabVp/ à des amis" and "neuf /radVt/ à des amis" remained, with the target front rounded vowels /y, œ/ and the vowels /i, u, ɛ, o, a/ for comparison. For stimulus verification, three monolingual native speakers of PF who had been in the United States for less than a month identified each stimulus. They made 0 errors and rated the stimuli a median of 9 on a 1–9 scale from foreign sounding (1) to native French sounding (9), indicating that these were good tokens of the intended categories. In order to examine the acoustic differences that might affect vowel perception, an acoustical analysis of the French vowel stimuli was performed.

2. Acoustic analysis

Acoustic analysis was performed by means of customized software in MATLAB (CVCZ by Valeriy Shafiro). First, the onset and offset of the syllable containing the target vowel (i.e., /bVp/ or /dVt/) were determined on the basis of the following operational definitions: Onset was defined as a change in amplitude, indicating release of the preceding consonantal occlusion. Offset was defined as the decrease in periodic energy indicated in F2 and F3 on a spectrogram coinciding with decreased amplitude on the waveform, indicating the beginning of closure of the following consonant. The program then calculated the temporal midpoint between onset and offset of the syllable and derived the first three formant frequencies for a 25 ms window centered around that point using linear predictive coding analysis (24 coefficients).

The top scatter plot in Fig. 1 represents all vowel stimuli (i.e., three tokens per speaker) uttered by the three PF speakers in bilabial /rabVp/ context with the F2 frequency along the *X*-axis and the F1 along the *Y*-axis (in barks).[5] For purposes of comparison, these are superimposed onto average values for seven (long) AE vowels (in symbols connected by dashed lines) spoken in bilabial context in nonsense trisyllables in sentences by three female AE speakers (Strange *et al.*, 2007). The AE vowel space appears to be shifted down (higher F1 values) relative to PF vowels; however, the relative distance of PF and AE vowels on the front-back dimension (F2 values) can be compared meaningfully. In bilabial context, PF front rounded /y/ is nearest to AE and PF front unrounded /i/, actually overlapping with tokens of PF /i/ across speakers of the same gender. Crucially, in this context, PF /y/ is far closer to /i/ than to /u/ in both languages. PF /œ/ is intermediate between front and back AE vowels.

The bottom plot represents the PF vowel stimuli in alveolar /radVt/ context, superimposed on average values for AE vowels spoken in alveolar context (Strange *et al.*, 2007). It should be noted that the front unrounded vowels, especially /i/, in both languages are little affected by changes in consonantal context. Front rounded PF /y/ remains at the "front" of the vowel space, some tokens overlapping with PF /i/ across speakers. However, the vowel spaces for both languages are generally more constricted along the front-back dimension because both PF and AE back vowels /u/ and /o/ (and AE /ʊ/, not shown here) are "fronted" (produced with higher F2 frequencies) in alveolar context relative to bilabial context. However, all tokens of PF /y/ remain closer to PF /i/ than to PF /u/ in alveolar context, whereas they are spectrally intermediate between AE /u/ and /i/. PF /œ/ is only slightly fronted. PF /o/ approximates both AE and PF /u/ more in bilabial than in alveolar context.[6] PF /a/ tokens are generally produced with higher F2 values in alveolar context than in bilabial context.

The acoustic analysis of the stimuli demonstrates that in bilabial context, PF /y/ was far closer to /i/ than to /u/ in both languages; thus, if the naïve participants and L2-learners perceptually assimilated /y/ to AE back vowels in bilabial context, acoustics alone would not explain their assimilation patterns. Furthermore, the relationship of the PF stimuli to AE vowels described above suggests that if the participants per-

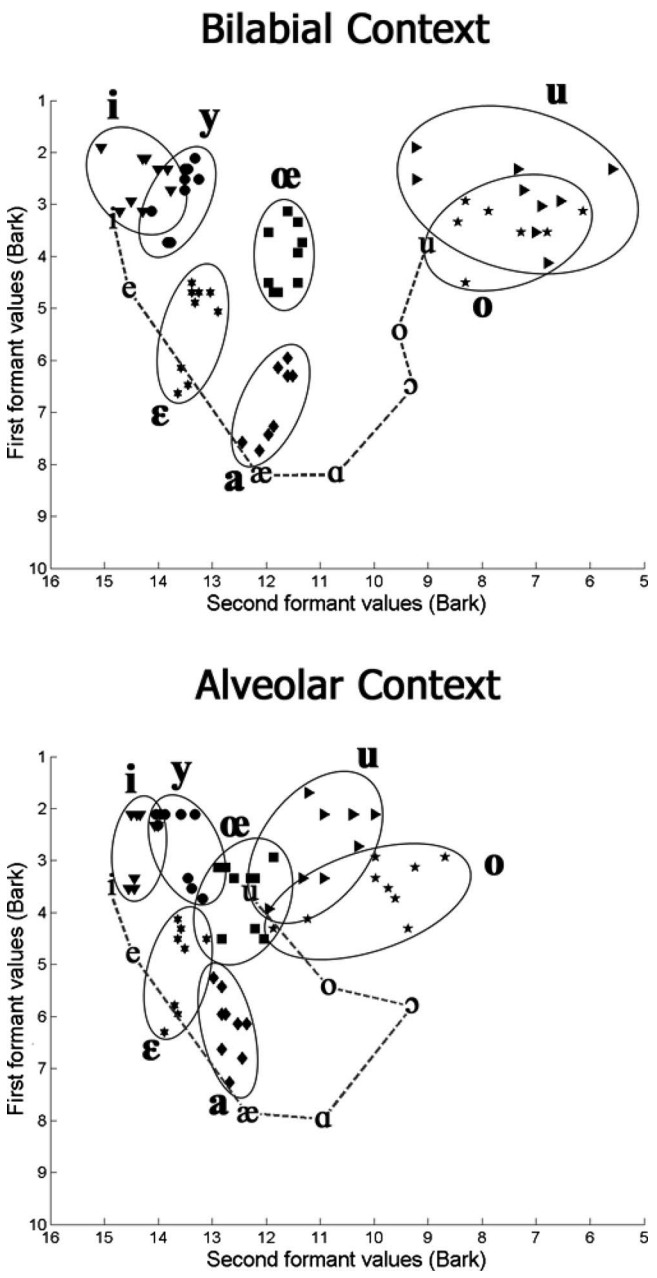


FIG. 1. Formant 1/formant 2 vowel spaces for bilabial /rab Vp/ stimuli (top) and alveolar /rad Vt/ stimuli (bottom) uttered in carrier phrases by three native speakers of PF. For comparison purposes, averages (in symbols connected by dashed lines) of four tokens from three monolingual female speakers of AE in bilabial /gəbVpə/ context (top) and alveolar /gədVtə/ context (bottom) in phrases from the production study of Strange *et al.* (2007) are provided.

ceptually assimilated front rounded vowels to back vowels more often in alveolar context, this may be more a function of the fronting of AE back vowels, i.e., of their native vowel space, than of the PF vowel space.

C. Procedure

The stimulus files were transferred via a zip disk to a Dell Dimension XPS B800 computer. Participants listened to stimuli presented via STAX Professional SR Lamda headphones connected to a STAX Professional SRM-1/MK-2 amplifier, receiving the signal from the computer. Sessions took place in the sound-attenuated chamber. Stimuli were entered into a customized software program designed to execute the perceptual assimilation task.

Prior to the experimental trials, keyword, task, and stimulus familiarization procedures trained listeners to choose appropriate responses and become familiar with the stimuli. In the keyword familiarization task, the experimenter asked the participant to read the 13 (AE) keywords ("heed, hid, hayed, head, had, hod, hawed, hud, hoed, hood, who'd, hued, herd") aloud to her. In the present experiment, the word "hued" was included as an AE keyword response option based on the finding that AE learners of French frequently produce /ʲu/ in French conversation when targeting the production of /y/ (Levy and Law II, 2008).

Task familiarization began with the presentation of AE phrases "five /gəCVCə/ this time" with the AE vowels (/i/, /ɪ/, /e/, /ɛ/, /æ/, /ɑ/, /ɔ/, /ʌ/, /o/, /ʊ/, /u/, /ʲu/, /ɝ/) in randomly presented /bVp/ and /dVt/ contexts recorded by a female native AE speaker, with keyword response alternatives and feedback. The listener was instructed to select the AE keyword that contained the vowel most similar to the second vowel of the nonsense word and then to rate the vowel from 1–9 according to how nativelike (9) or foreign sounding (1) the vowel was. On the last block (block 4), no feedback was provided, and listeners needed to achieve a criterion of no more than one error in identifying a particular AE vowel token and no more than three errors altogether in order to proceed to the stimulus familiarization.

The overall structure of the remaining blocks was the following: Stimuli were blocked by context. Thus, half of the listeners were presented all of the French stimuli in bilabial context before hearing the stimuli in alveolar context, and the other half were presented stimuli in alveolar context before bilabial. Stimulus familiarization consisted of a block of one token each of the seven PF vowels in the stimulus phrases and did not include feedback.

The perceptual assimilation experiment began with instructions presented to participants on the computer screen. Participants were instructed to listen to each phrase, paying attention to the second vowel (e.g., /radVt/) in the target nonsense word of each phrase. They were asked to focus on the target vowel and to try to ignore other aspects of the phrases (e.g., consonants or intonation) that might be distracting or sound different from English. The stimulus was presented once. The listeners saw the 13 AE keywords and chose the one that contained the vowel that was most similar to the target vowel. They heard the stimulus again and rated the French vowel on a scale of 1–9. A rating of 1 indicated "most foreign sounding," 9 indicated "most English sounding," and participants were encouraged to rate the stimulus as any number in between using the whole spectrum of the scale.

The test blocks contained all three tokens of each of seven French vowels by each speaker. Blocks were presented twice, with stimuli randomized within blocks. The presentation order of speakers within each context was determined by a Latin square. Each vowel was presented twice in each con-

sonantal context, in phrases. Each participant completed six judgments for each speaker's three vowel tokens, totaling 18 judgments per vowel per context.

III. RESULTS

A. Data analysis

The frequencies of selecting a particular response category in the perceptual assimilation task were tallied for each French vowel across all listeners within each language group within each consonantal context (summing over 18 judgment opportunities for each listener in each context). Frequencies of the modal response and the other chosen responses for each PF vowel stimulus were converted to percentages of total trials. The median goodness ratings for each AE response category were then computed. Because the range of median ratings was not large (mdn=3–7 for most vowels), the focus of the analysis is on response percentages as that appears to provide more telling information regarding perceived similarity.

To assess the effects of language experience and consonantal context on the perceptual assimilation patterns, a multinomial baseline-category logistic regression model (Agresti, 2007) was fit to the data. The most often chosen response category for the HiExp group was used (arbitrarily) as the baseline category. Standard errors and confidence intervals for the estimated odds ratios (ORs) were analyzed to judge whether the outcome was statistically significant, which meant here that the ORs were judged to be significantly different from 1.

In order to sort out inter- and intraparticipant variability, an "internal consistency" analysis of perceptual assimilation results was performed. Internal consistency was determined by examining each listener's modal responses, i.e., for each PF vowel, what percentage of trials each listener assimilated the vowel to his or her AE modal category, regardless of what that category was. The mean consistency score for each group was calculated. Internal consistency scores across all PF vowels and for each front rounded vowel are reported. A high internal consistency score indicates that individual listeners consistently gave the same response to a particular stimulus, whereas a low score suggests that they gave several different responses.

B. Overview of results

Table I displays the modal AE categories chosen by the NoExp group, the ModExp group, and the HiExp group for all PF stimuli presented in bilabial context (above) and in alveolar context (below). Within each language group, the left-hand column lists the PF stimuli. The second column represents the AE responses chosen. The "mode percent chosen" indicates the percentage of trials that the most frequent AE responses were chosen by the group. The median rating indicates the median of goodness ratings from 1 (most foreign sounding) to 9 (most AE sounding) for the trials in which each AE response category was selected. Only AE vowel responses chosen on at least 10% of trials (i.e., 23 responses out of 234 possible) in at least one consonantal context by at least one group are listed and analyzed.

The multinomial linguistic regression analysis revealed language experience effects for some comparisons on all PF vowels except /i, u/, for which the vast majority of responses were the AE /i/ and /u/, respectively. Significant consonantal context effects were found for /a, y, o, æ/, but not for /i, ɛ, u/. For some PF vowel categories, language experience effects were significant. Thus, the comparisons for each vowel were analyzed individually. (The Appendix lists the comparisons performed using the regression analysis and their resulting statistical significance or nonsignificance.) These results are discussed in detail below.

Overall, median goodness ratings were lower in the NoExp group than in the HiExp group for modal responses to /i/, /y/, /u/, /ɛ/, and /a/ but not to /œ/. The front rounded vowel /y/ received the lowest goodness ratings (median =6,3,3 by NoExp, ModExp, and HiExp, respectively, in bilabial context; median=4,3,2 by NoExp, ModExp, and HiExp, respectively, in alveolar context).

A three-factor (language experience × vowel × context) analysis of variance was performed to examine the internal consistency of perceptual assimilation of all vowels. The dependent variable in this analysis was internal consistency (calculation described above), and the independent variables were level of language experience, consonantal context, and vowel. A significant language experience effect was revealed [$F(2,545)=33.28$, $p<0.0001$], suggesting that with language experience, individual listeners selected their modal response more often. Controlling for consonantal context and vowel, individuals with moderate language experience were not significantly more consistent in their responses than those with no language experience [$F(1,545)=0.62$, $p=0.4314$], but the HiExp group responded more consistently than the ModExp group [$F(1,545)=44.08$, $p<0.0001$] and the NoExp group [$F(1,545)=55.15$, $p<0.0001$]. Consonantal context was revealed as a significant factor [$F(1,545)=9.82$, $p=0.0018$] in response consistency, with more consistency found in bilabial (mean=86.829, SD=17.08) than in alveolar context (mean=82.89, SD=18.45) context. A significant vowel effect was found [$(F(6,545)=23.31$, $p<0.0001$], suggesting that some vowels were assimilated in a more consistent manner than others. For example, responses to /y/ (mean=86.69, SD=13.41) were less consistent than to /i/ (mean=98.23, SD=3.56) but more consistent than to /œ/ (mean=76.46, SD=19.02).

In the following sections, the AE naïve listeners and L2-learners' perceptual assimilation patterns for PF front unrounded and low vowels (PF /i, ɛ, a/) will be presented to provide a basis for comparison with the PF other vowels, followed by their perceptual assimilation of PF back rounded vowels (/o, u/) and of front rounded vowels (/y, œ/).

C. Perceptual assimilation of front unrounded and low PF vowels /i/, /ɛ/, and /ɑ/

PF /i/ was perceptually assimilated to AE /i/ on 99% of responses and received the highest goodness ratings (7–8) of all vowels, suggesting that this was an excellent fit to the AE /i/ category for naïve and experienced listeners alike and that listeners were on task.

TABLE I. Perceptual assimilation of PF vowels /i, ɛ, a, o, u, y, œ/ in bilabial /rabVp/ and alveolar /radVt/ contexts by AE listeners with no French experience (No Exp), moderate French experience (Mod Exp), and extensive formal+immersion French experience (Hi Exp): Percent chosen for each response and median goodness ratings (in parentheses) are presented for each vowel. For each AE vowel stimulus, only responses chosen on at least 10% of the trials (i.e., 23 responses out of 234 possible) in at least one consonantal context by at least one experience group are listed. [Total number of responses per vowel =234 (i.e., 18 judgments per vowel per context by 13 participants in each group).]

	No Exp			Mod Exp			Hi Exp		
PF	AE	%	Med	AE	%	Med	AE	%	Med
				Bilabial context					
i	i	98	8	i	99	7	i	100	7
ɛ	ɛ	79	7	ɛ	85	6	ɛ	89	6
	æ	12	6	e	11	5	e	10	6
a	ɑ	58	6	ɑ	52	5	ɑ	56	4
	æ	36	6	æ	34	6	æ	43	6
o	u	62	7	o	71	6	o	99	6
	o	24	7	u	18	4	u	1	4
u	u	90	7	u	85	4	u	78	6
	ʲu	6	4				o	13	4
y	ʲu	80	6	ʲu	85	3	ʲu	72	3
	i	12	4	u	13	3	u	28	3
œ	ʊ	38	6	ʊ	65	4	ʊ	60	6
	u	34	6	ʌ	12	4	ɝ	27	5
	ɝ	1	7	u	11	3	ʌ	8	4
				Alveolar context					
i	i	97	7	i	98	7	i	99	7
ɛ	ɛ	76	6	ɛ	90	6	ɛ	95	6
	e	11	6	e	10	6	e	4	5
	æ	10	6						
a	æ	57	7	æ	48	6	æ	62	6
	ɑ	29	7	ɑ	30	5	ɑ	35	6
	ɛ	12	6	ɛ	18	6			
o	u	43	6	o	69	5	o	98	6
	o	41	6	u	19	4			
u	u	84	6	u	75	4	u	91	6
	ʲu	9	4	ʲu	15	3			
y	ʲu	65	4	ʲu	71	3	ʲu	61	2
	u	31	6	u	24	4	u	34	4
œ	u	59	6	ʊ	40	3	ʊ	61	5
	o	17	6	o	25	4	ɝ	20	6
	ʊ	17	5	u	23	4	u	9	4

Results for PF /ɛ/ indicate significant differences in perceptual assimilation patterns as a function of language experience [$\chi^2(6)=49.56$, $p=0.0001$] but not of consonantal context [$\chi^2(3)=0.47$, $p<0.3250$]. Listeners with extensive language experience perceptually assimilated PF /ɛ/ more often to AE /ɛ/ (as opposed to AE /e/ or AE /æ/). However, the language experience effect was not statistically significant for all comparisons. For example, no significant language experience differences were found in the estimated odds of perceptual assimilation of PF vowel /ɛ/ to AE /e/ versus that to the reference category AE /ɛ/ (NoExp versus ModExp OR=0.8704, $p=0.5466$; ModExp versus HiExp OR=0.657, $p=0.0757$; NoExp versus HiExp OR=1.32, $p=0.2606$). However, for the AE /æ/ responses versus /ɛ/ responses, the NoExp listeners differed significantly in their estimated odds of selecting /ɛ/ over /æ/ from the estimated odds of the ModExp listeners selecting /ɛ/ over /æ/ (OR =8.39, $p<0.0001$) and from the estimated odds of the HiExp selecting /ɛ/ over /æ/ (OR=30.7, $p<0.0001$). In contrast, no significant differences were found between the odds of a listener in the ModExp group and a listener in the HiExp group selecting /ɛ/ over /æ/ (OR=3.662, $p=0.1068$).

Results for the perceptual assimilation of vowel /a/ provide evidence of a significant language experience effect

$[\chi^2(4)=36.77,\ p<0.0001]$ and a consonantal context effect $[\chi^2(2)=82.65,\ p<0.0001]$. AE /æ/ was the overall modal choice in alveolar context (56%), and AE /ɑ/ was second-most chosen (31%), whereas /ɑ/ was the modal choice (55%) in bilabial context and /æ/ was the second-most selected response (38%). For the regression analysis, the overall modal category /æ/ was selected as the baseline category for comparison purposes.

For /ɑ/ versus /æ/ comparisons of responses to PF /a/, no statistically significant language differences were revealed (NoExp versus ModExp OR=0.92, $p=0.61$, ModExp versus HiExp OR=1.15, $p=0.34$, NoExp versus HiExp OR=1.06, $p=0.66$). However, for the OR of /ɛ/ versus /æ/ assimilation, a significant language experience was revealed (NoExp versus ModExp OR=0.555,$p=0.0096$; ModExp versus HiExp OR=13.286, $p<0.0001$; No Exp versus HiExp OR=7.371, $p<0.0001$). Results suggest a significant effect of context on participants' choosing /ɑ/ over /æ/ in assimilating PF /a/ (OR=2.58, $p<0.0001$). That is, the estimated odds of choosing /ɑ/ over /æ/ in bilabial context were 2.58 times the odds of choosing /ɑ/ over /æ/ in alveolar context. No significant context effect was found for choosing /ɛ/ over /æ/ (OR =0.654, $p>0.0626$).

D. Perceptual assimilation of back PF vowels /o/ and /u/

PF /o/ revealed a significant effect of language experience in perceptual assimilation $[\chi^2(4)=47.16,\ p<0.0001]$ in both comparisons between NoExp and ModExp groups and between ModExp and HiExp groups. A significant consonantal context effect $[\chi^2(4)=18.36,\ p=0.0001]$ was revealed, as well as an interaction $[\chi^2(2)=11.98,\ p=0.0175]$ between language experience and consonantal context. Although it was expected that this vowel would be perceived as most similar to AE /o/ by all groups, the NoExp group assimilated PF /o/ more often to AE /u/ than to AE /o/. The modal category and the second most frequently chosen response reversed themselves with moderate experience. In the HiExp group, listeners had learned to assimilate PF /o/ as most similar to AE /o/. Using AE /o/ as the reference group, NoExp participants were significantly more likely to choose AE /u/ over AE /o/ when compared to the ModExp group (bilabial context: OR=10.07, $p<0.0001$; alveolar context: OR=3.82, $p<0.0001$). Similarly, the ModExp group compared to the HiExp group was significantly more likely to choose /u/ versus /o/ (bilabial context: OR=29.66, $p<0.0001$; alveolar context: $p<0.0001$[7]), as were the NoExp group versus the HiExp group (bilabial context: OR=298, $p<0.0001$; alveolar context: $p<0.0001$. The median ratings for modal responses to PF /o/ were 6 for all three language groups. A context effect was shown by the NoExp group, who assimilated PF /o/ to AE /u/ more often in bilabial context than alveolar context $[\chi^2(1)=17.83,\ p<0.0001]$ but not for the ModExp group $[\chi^2(1)=0.0998,\ p=0.7520]$. The HiExp group assimilated 98% of the PF /o/ stimuli to AE /o/, with no PF /u/ responses in alveolar context; thus no regression analysis was performed for a context effect for the HiExp group.

PF /u/ was perceptually assimilated primarily to AE /u/ and to AE /ʲu/ on a minority of trials (see Table I). Although significant effects of language experience $[\chi^2(4)=27.01,\ p<0.0001]$ and consonantal context PF /u/ $[\chi^2(2)=7.53,\ p<0.0232]$ were shown for PF /u/, statistical analyses of all context and language group comparisons were not significant (for /u/ versus /ʲu/ for NoExp versus ModExp $[\chi^2(1)=0.0002,\ p=0.99]$, ModExp versus HiExp $[\chi^2(1)=0.0000,\ p=1]$ and NoExp versus HiExp $[\chi^2(1)=0.0000,\ p=1]$), probably due to the low rate of /ʲu/ responses. Eight of the 13 HiExp listeners perceptually assimilated PF /u/ to AE /u/ more often in alveolar context than in bilabial, and five assimilated PF /u/ to AE /u/ less often in alveolar context than in bilabial context. The median ratings for PF /u/ as an exemplar of the AE response /u/ were a mean of 7 for NoExp, decreasing to 4 for ModExp and increasing to 6 for HiExp.

E. Perceptual assimilation of front rounded PF vowels /y/ and /œ/

To determine whether PF front rounded vowels were assimilated to AE front or back vowels (Question 1), responses to stimuli containing PF /y/ and /œ/ were collapsed. AE front unrounded vowel responses (/i/, /ɪ/) were combined, central vowels (/ʌ/, /ɝ/) were combined, and back rounded vowels (/u/, /ʲu/, /ʊ/, /o/) were combined. When presented with PF front rounded vowels, 4% of the NoExp group's responses were front AE vowels, 1% were central, whereas 95% were back vowels. The 4% of front vowels chosen by NoExp were primarily due to one listener assimilating all /y/ tokens in bilabial context to /i/ and by one to three front vowel responses (out of 18) by a minority of listeners. Similarly, the ModExp group assimilated 1% of the front rounded vowels to front vowels, 1% to mid, and 98% to back vowels. The HiExp group categorized no front rounded vowels as AE front vowels and 88% as back vowels. Central vowels, primarily rhotacized /ɝ/, accounted for 12% of HiExp responses. Thus, the large majority of listeners perceived front rounded vowels as most similar to AE back vowels.[8] Hence, the analysis will focus on assimilation of PF front versus back rounded vowels.

An overview of perceptual assimilation results in each consonantal context is provided in Tables II and III, which display a matrix of the PF front and back rounded vowels presented to the AE listeners (left column in each context) and the AE vowels to which they were perceptually assimilated (/ʲu/, /u/, /i/, /ʊ/, /o/, /ʌ, /ɝ/) for at least 10% of responses by at least one experience group in one consonantal context. Within each consonantal context (bilabial in Table II, and alveolar in Table III), the second column lists the language experience groups. The figures indicate the percentage of trials (out of 234 opportunities per context) that a particular AE response was chosen by the group. The median rating, in parentheses below the response percentage, indicates the median of goodness ratings from 1 (most foreign sounding) to 9 (most AE sounding) for the trials in which the modal response category was selected.

For the front rounded PF /y/, an overall language experience effect $[\chi^2(6)=43.43,\ p<0.0001]$ and an overall con-

TABLE II. Matrix of PF vowels, /y, œ, u, o/ perceptually assimilated to AE vowels by AE listeners with no French experience (NoExp), moderate French experience (ModExp), and extensive formal+immersion French experience (HiExp) in bilabial context: Percent chosen for each response and median goodness ratings (in parentheses) are presented for each vowel. For each AE vowel stimulus, only responses chosen on at least 10% of the trials (i.e., 23 responses out of 234 possible) in at least one consonantal context by at least one experience group are listed. [Total number of responses per vowel=234 (i.e., 18 judgments per vowel per context by 13 participants in each group).]

PF vowels, bilabial context	Language experience	American English vowels						
		ʲu	u	i	ʊ	o	ʌ	ɝ
y	NoExp	80	7	12				
		(6)	(5)	(4)				
	ModExp	85	13	0.4				
		(3)	(3)	(2)				
	HiExp	72	28					
		(3)	(3)					
œ	NoExp		34		38	8	8	1
			(6)		(6)	(6)	(5)	(7)
	ModExp		11		65	6	12	
			(3)		(4)	(2)	(4)	
	HighExp		3		60	2	8	27
			(4)		(6)	(4)	(4)	(5)
u	NoExp	6	90					
		(4)	(7)					
	ModExp	3	85			4		
		(3)	(4)			(3)		
	HiExp	1	78			13		
		(5)	(6)			(4)		
o	NoExp		62			24		
			(7)			(7)		
	ModExp		18			71		
			(4)			(6)		
	HiExp		1			99		
			(4)			(6)		

text effect [$\chi^2(6)=42.37$, $p<0.0001$] were revealed.[9] As indicated in Table I, significantly more /ʲu/ responses were selected by the ModExp group than by the NoExp group [$\chi^2(1)=23.83$, $p<0.0001$] and by the HiExp group [$\chi^2(1)=15.26$, $p<0.0001$]. The significant experience by context interaction [$\chi^2(6)=22.50$, $p<0.0001$] in the perceptual assimilation of /y/ reflected a significant language experience effect in bilabial context but not in alveolar context. In bilabial context, the estimated OR of selecting AE /u/ versus /ʲu/ was statistically significant for the participants with ModExp when compared to the estimated OR of participants with NoExp selecting AE /u/ versus AE /ʲu/ (OR=1.76, $p=0.08$). Similarly, the estimated OR of selecting AE /u/ versus AE /ʲu/ was significant for the participants with HiExp when compared to the estimated OR of participants with ModExp selecting AE /u/ over /ʲu/ (OR=2.59, $p<0.0001$). Moreover, in bilabial context, participants in the HiExp group chose AE /u/ over /ʲu/ significantly *more* often than participants in the NoExp group OR=4.59, $p<0.0001$). In alveolar context no significant language experience effect was revealed for /y/ (NoExp versus ModExp OR=0.71, $p=0.11$; ModExp versus HiExp OR=1.67, $p=0.95$; NoExp versus HiExp OR=1.18, $p=0.40$).

The consonantal context effect [$\chi^2(3)=45.58$, $p<0.0001$] was evident in that listeners selected /ʲu/ more often in bilabial context than in alveolar. Conversely, AE /u/ was selected in response to PF /y/ more often in alveolar context than in bilabial. The context effect for PF /y/ was smaller for the more experienced groups; differences between alveolar and bilabial contexts for PF /y/ assimilated to AE /u/ were 24%, 9%, and 6% for NoExp, ModExp, and HiExp, respectively, and 15%, 14%, and 11% for PF /y/ assimilated to AE /ʲu/. When experience was treated as an ordinal rather than a nominal variable, it was estimated that the OR comparing alveolar to bilabial context decreased by 52% with each level of experience ($p<0.0001$).

In response to the Levy and Strange (2008) finding that NoExp listeners confused /i-y/ in bilabial context, the AE /i/ responses to PF /y/ were examined relative to the modal AE /ʲu/ response. Individual data suggest that the higher odds of selecting /i/ in bilabial context than alveolar was primarily due to one NoExp participant's assimilation performance. This participant assimilated /y/ to /i/ 100% in bilabial context and 0% in alveolar context. All other NoExp participants' modal responses to /y/ were back vowels. However, their nonmodal responses revealed that PF /y/ was perceptually

TABLE III. Matrix of PF vowels, /y, œ u, o/ perceptually assimilated to AE vowels by AE listeners with no French experience (NoExp), moderate French experience (ModExp), and extensive formal+immersion French experience (HiExp) in alveolar context: Percent chosen for each response and median goodness ratings (in parentheses) are presented for each vowel. For each AE vowel stimulus, only responses chosen on at least 10% of the trials (i.e., 23 responses out of 234 possible) in at least one consonantal context by at least one experience group are listed. [Total number of responses per vowel=234 (i.e., 18 judgments per vowel per context by 13 participants in each group).]

PF vowels, alveolar context	Language experience	American English vowels						
		ʲu	u	i	ʊ	o	ʌ	ɜ˞
y	NoExp	65	31	1				
		(4)	(6)	(7)				
	ModExp	71	24	1				
		(3)	(4)	(2)				
	HighExp	61	34					
		(2)	(4)					
œ	NoExp		59		17	17		
			(6)		(5)	(6)		
	ModExp		23		40	25	5	1
			(4)		(3)	(4)	(2)	(4)
	HighExp		9		61	2	8	20
			(4)		(5)	(4)	(4)	(6)
u	NoExp	9	84			1		
		(4)	(6)			(6)		
	ModExp	15	75			2		
		(3)	(4)			(4)		
	HighExp	5	91					
		(3)	(6)					
o	NoExp		43			41		
			(6)			(6)		
	ModExp		19			69		
			(4)			(5)		
	HighExp					98		
						(6)		

assimilated to AE /i/ in bilabial context on 17% of trials by two participants and on 6% of trials by two other participants. In alveolar context, two participants perceptually assimilated PF /y/ to AE /i/ in 6% of the trials. The NoExp group's responses to /y/ were not significantly more likely to be /i/ (as opposed to /ʲu/) than those of the ModExp Group, for whom only 7% of responses to /y/ were /i/ (OR=1.37, $p=0.084$). Because none of the HiExp selected /i/ as a response choice, the regression model could not examine whether the odds were higher of an individual in the NoExp or in the ModExp group to assimilate /y/ to /i/ than those of an individual in the HiExp group to do so. Thus, a chi-square analysis was performed, comparing experience and choosing /i/ (versus "not /i/") as a response. The statistically significant $[\chi^2(1)=467.00, p<0.0001]$ results suggest a dependence between the level of language experience and choosing AE /i/ as the response to PF vowel /y/.

Internal consistency of /y/ was not significantly different $[F(1,545)=0.28, p=0.5996]$ between listeners in the NoExp group (87% in bilabial, 76% in alveolar) and ModExp group (91% for bilabial and 73% in alveolar). However, consistency was significantly higher with extensive experience compared to moderate experience $[F(1,545)=5.05, p=0.0277]$ and significantly different $[F(1,545)=7.70, p=0.007]$ between NoExp and HiExp groups (95% in bilabial and 78% in alveolar). This suggests that with extensive experience, the listeners had generally selected a category to which they assimilated most of the /y/ tokens. The consonantal context effect was not found to be significant for internal consistency $[F(1,545)=0.03, p=0.8593]$, suggesting that individuals did not respond in a more or less consistent manner as a function of whether PF /y/ appeared in bilabial versus alveolar context.

For PF /œ/, a significant overall language experience effect was revealed $[\chi^2(12)=307.32, p<0.0001]$, with comparisons of estimated odds of response choices between the NoExp and ModExp groups, as well as those between the ModExp and HiExp groups, reaching statistical significance. PF /œ/ was perceptually assimilated to more AE categories than was /y/. For the NoExp group, in fact, when consonantal contexts were combined, no AE category was selected more than 50% of the time in response to PF /œ/; thus comparisons of modal choice AE /ʊ/ versus AE /u/, AE /ʊ/ versus AE /o/, and AE /ʊ/ versus AE /ɜ˞/ were performed.

In the /u/ versus /ʊ/ comparison, the odds of a participant in the ModExp group choosing AE /u/ over AE /ʊ/ in response to PF /œ/ were estimated to be 17% of the odds of an individual in the NoExp group choosing AE /u/ over AE /ʊ/ (OR=0.17, $p<0.0001$). For the ModExp group, the estimated odds of choosing AE /u/ rather than AE /ʊ/ were 392% of the odds of an individual in the HiExp group choosing /u/ rather than /ʊ/ (OR=3.92, $p<0.0001$). Similarly, in a comparison of /o/ selection as opposed to /ʊ/, the estimated odds of a participant with NoExp selecting /o/ over /ʊ/ were 1.56 times the odds of a participant with ModExp selecting /o/ over /ʊ/ (OR=1.56, $p=0.0089$), and the odds of a participant with ModExp choosing /o/ over /ʊ/ were estimated to be 10.87 times the odds of a participant with HiExp choosing /o/ over /ʊ/ (OR=10.87, $p<0.0001$). And finally, for the group with NoExp, the odds of choosing /o/ over /ʊ/ were 19 times the odds of a participant with HiExp choosing /o/ over /ʊ/ (OR=0.19, $p<0.0001$). For the HiExp group, the second most frequent response category was the rhotacized vowel /ɜ˞/ as in "herd" (24%), a response selected virtually only by this group. The estimated OR of a participant in the NoExp group selecting /ɜ˞/ over /ʊ/ versus the odds of a participant in the ModExp group choosing /ɜ˞/ over /ʊ/ were not statistically significant (OR=3.8, $p<0.27$). However, the estimated odds that an individual in the ModExp Group versus an individual in the HiExp group would select /ɜ˞/ rather than /ʊ/ were statistically significant (OR=0.01, $p<0.0001$), as they were only 1% as large as the estimated odds that an individual in the HiExp Group would select /ɜ˞/ over /ʊ/. Similarly, for an individual in the NoExp Group, the estimated odds of choosing /ɜ˞/ rather than /ʊ/ were only 3.8% as high as the estimated odds that an individual in the HiExp group would choose /ɜ˞/ over /ʊ/ (OR=0.038, $p<0.0001$). In the comparison of AE /o/ versus the reference category AE /ʊ/ responses to PF /œ/ for the NoExp group, the estimated odds of choosing AE /o/ rather than AE /ʊ/ were 19 times (1900% of) the odds of a participant with HiExp choosing /o/ rather than /ʊ/ (OR=19.5, $p<0.0001$). For the ModExp group, the estimated odds of choosing AE /o/ rather than AE /ʊ/ were 11 times the odds of an individual in the HiExp group choosing AE /o/ rather than AE /ʊ/ (OR=11.05, $p<0.0001$). The median goodness rating of the modal responses to PF /œ/ was 5 for the HiExp group, 4 for the ModExp group, and relatively high (6) for the NoExp group.

Responses to PF /œ/ also varied significantly as a function of consonantal context [$\chi^2(6)=113.9$, $p<0.0001$]. The AE /u/ response was less frequently chosen in the bilabial condition than in alveolar by all groups. Conversely, more AE /u/ responses were selected by all groups in bilabial than in alveolar context. The odds of choosing AE /u/ rather than /ʊ/ in bilabial context were 26.5% of the odds of choosing AE /u/ rather than /u/ in alveolar context (OR=0.265, $p<0.0001$). The odds of assimilating PF /œ/ to AE /o/ rather than to /ʊ/ in bilabial context were 0.19 times (i.e., 19%) ($p<0.0001$) the odds of choosing AE /o/ over AE /ʊ/ in alveolar context. Thus, the odds of an individual with HiExp choosing AE /o/ rather than AE /ʊ/ were approximately 5.3 (530%) times larger than the odds in alveolar context. However, the odds of choosing AE /ɜ˞/ rather than /ʊ/ were not significantly different (OR=1.36, $p>0.05$) in bilabial versus alveolar context. The experience-context interaction was not significant for /œ/ [$\chi^2(12)=6.94$, $p=0.862$]. A trend toward a decreased context effect with experience was noted (e.g., the differences in response percentages to /œ/ in bilabial versus alveolar context for AE /ʊ/ responses were 21%, 25%, and 1% and for /ʊ/ were 25%, 12%, and 6% for the NoExp, ModExp, and HiExp groups, respectively); however, these decreases were not statistically significant ($p=0.20$).

Internal consistency in perceptual assimilation of /œ/ revealed a significant effect of language experience [$F(1,545)=8.24$, $p=0.0006$.], but no consonantal context effect [$F(1,545)=2.29$; $p=0.1348$.]. The NoExp and ModExp groups demonstrated the lowest consistency for any vowel (NoExp=58% in bilabial and 76% in alveolar; ModExp =75% in bilabial and 57% in alveolar), suggesting that individual listeners selected several categories in response to /œ/ stimuli. The HiExp group, on the other hand, demonstrated high internal consistency (90% in bilabial and 82% in alveolar), suggesting that most /œ/ stimuli were assimilated to the majority of listeners' particular modal category. Planned comparisons indicated no significant difference between the NoExp and ModExp groups [$F(1,545)=0.00$, $p=0.9938$] but a significant difference in internal consistency between ModExp and HiExp [$F(1,545)=12.33$, $p=0.0008$].

IV. DISCUSSION

The main findings for perceptual assimilation of PF /y/ by AE naïve listeners and L2 French learners can be summarized as follows: (a) For all language groups, PF /y/ was perceived as most similar to the AE palatalized vowel /ʲu/ or to /u/. A language experience effect/interaction was found: In bilabial context, PF /y/ was heard most often as relatively similar to AE /ʲu/ by L2-learners with formal education, less often so by naïve listeners, and least often so by listeners with formal + immersion experience. In alveolar context, listeners perceptually assimilated PF /y/ to AE /ʲu/ or /u/ similarly, despite their different language backgrounds. (b) A significant consonantal context effect was revealed with more /ʲu/ responses in bilabial context than in alveolar context. In bilabial context, /ʲu/ responses were greatest for naïve listeners, smaller for listeners with just formal instruction, and smallest for listeners with extensive experience. In alveolar context, a trend in the same direction was noted but did not reach statistical significance. (c) Naïve listeners and listeners with formal experience were relatively consistent in selecting their modal response to PF /y/. With extensive formal and immersion experience, listeners had more or less settled on their particular response.

The main findings regarding perceptual assimilation of the mid front rounded PF vowel /œ/ were the following: (a) PF /œ/ was perceived most often as similar to AE /u/ by naïve listeners. Listeners with formal French experience perceived /œ/ as similar to AE /ʊ/, and individuals with extensive formal and immersion French experience perceived PF /œ/ as most similar to AE /ʊ/ or /ɜ˞/. (b) A consonantal context effect was evident with listeners in all groups perceiving PF /œ/ as similar to AE /ʊ/ more often in bilabial context and

as similar to AE /u/ more often in alveolar context. (c) Individual listeners gave scattered responses to PF /œ/, perceiving this vowel as most similar to several different AE vowel categories (e.g., AE /ʊ, u, o, ʌ, ɝ/. Assimilation patterns were scattered within and across naïve listeners, but with extensive experience, listeners settled on the particular vowel to which they assimilated PF /œ/. In addition, for all vowels combined, assimilation was most internally consistent with extensive language experience and in bilabial context.

The overall language experience effect supports the general finding that adults are capable of continued L2 perceptual learning of vowels in adulthood (e.g., Flege and Hillenbrand, 1984; Gottfried, 1984; Levy and Strange, 2008). L2 phonological learning occurs both in the classroom and with extensive classroom and immersion experience, but some vowels are less learnable than others. The finding that consonantal context affects perceptual assimilation patterns in L2 vowel learning is generally consistent with previous discrimination studies (Gottfried, 1984; Levy and Strange, 2008) but has not been explored in previous perceptual assimilation studies and thus requires replication and further discussion.

A. Effects of language experience on PF front rounded vowels

Returning to Research Question 1, L2-learners' perceptual assimilation of PF front rounded vowels to back AE vowels, despite the PF vowels being acoustically more similar to AE front vowels, replicates work on German vowels (Polka, 1995; Strange *et al.*, 2007) and was predicted from discrimination difficulties with the PF /u-y/ contrast in earlier studies (Gottfried, 1984; Levy and Strange, 2008). In response to Questions 2 and 3, an effect of language experience had not been predicted for PF /y/ but had been predicted for PF /œ/, and more assimilation to PF /u/ had been predicted in alveolar than in bilabial context. The expectation of /œ/ revealing a language experience and context effect was generally borne out. The expectation of PF /y/ not revealing a language experience effect was found to be true in alveolar context but not in bilabial. In bilabial context, /y/ is not an appropriate variant of the AE /u/. Despite this, approximately one-third of the listeners with extensive experience had "phonologized" /y/ to their AE /u/ category, perceiving /y/ consistently as most similar to /u/.

The PAM-L2 (Best and Tyler, 2007) predicts that perceiving PF /y/ as most similar to AE /ʲu/ rather than to AE /u/ will be more advantageous to an L2-learner for discriminating the /u-y/ contrast in that pairing a /ʲu/-like PF /y/ with a /u/-like PF /u/ would result in two-category or category-goodness assimilation rather than single-category or less differentiated category-goodness assimilation. Thus, the language experience effect found in bilabial context for very experienced listeners, with fewer AE /ʲu/ responses and more AE /u/ responses to PF /y/, predicts more discrimination errors than for listeners with less experience or naïve listeners.

Levy and Strange (2008) found less accurate discrimination by individuals with immersion experience than by listeners with no experience. Similarly, in the present study, the HiExp group assimilated /y/ less often to /ʲu/ than did the NoExp group and more often to /u/ than the NoExp group, predictive of more discrimination difficulties. Levy and Strange (2008) did not test listeners with just formal experience, but the present finding of more /ʲu/ responses by these listeners would predict that listeners with only formal experience would fare better in discrimination than listeners with no or extensive immersion experience.

One could speculate that the higher number of /ʲu/ responses by the ModExp group may be a reflection of the orthographic mapping learned in introductory French classes, heightening these learners' awareness of the differences between /u-y/ (Burnham and Mattock, 2007). On the other hand, orthography could also contribute to the /u-y/ confusions, as PF /y/ is spelled as *u* (whereas PF /u/ is spelled as *ou*). One might further hypothesize that with extensive exposure to the PF /u-y/ contrast, listeners may cease to try to distinguish this difficult contrast. That is, the fewest AE /ʲu/ responses for PF /u/ by listeners with extensive immersion experience may point to a sort of "learned helplessness" (Seligman *et al.*, 1968) for the difficult /u-y/ pair, the development of a sense of failure that may be detrimental to language learning (Williams *et al.*, 2004). The absence of a language experience effect in alveolar context, characterized by no overall increase in /ʲu/ responses with experience, is consistent with Gottfried's (1984) and Levy and Strange's (2008) findings of poor PF /u-y/ discrimination in alveolar context, even with French immersion experience, but is not consistent with Flege and Hillenbrand's (1984) report of excellent discrimination of PF /u-y/ in alveolar context.

Assimilation patterns for PF /œ/ hold more promise for perceptual learning with French experience than do patterns for /y/. With formal instruction (ModExp), overall modal responses (/u/ and /ʊ/) reversed, with /ʊ/ being most frequently chosen. For HiExp, rhotacized vowel /ɝ/ was the second most frequent response. The difference in assimilation patterns for PF /œ/ across NoExp and ModExp groups suggests that classroom language experience may be associated with more accurate discrimination for contrasts involving PF /œ/. All comparisons involving HiExp listeners demonstrate perceptual assimilation differences from the less experienced groups, suggesting a difference in phonetic representation with L2 immersion. Consistent with this finding, Levy and Strange (2008) demonstrated that discrimination of PF /œ-u/ was more accurate for L2-learners with immersion experience. The PAM-L2 might attribute the learning of this contrast in part to the high frequency of /œ-u/ minimal pairs (e.g., deux /dœ/ "two" versus doux /du/ "soft/sweet") in dense phonological neighborhoods that contribute to the communicative relevance of learning the contrast. Further studies comparing discrimination by individuals with formal versus immersion experience should shed light on the sequence of discrimination changes in the course of L2-learning (see Levy, 2004).

B. Effects of consonantal context on PF front rounded vowels

As had been predicted, the consonants surrounding front rounded vowels affected assimilation patterns for L2-learners, with more tokens assimilating to /u/ in alveolar con-

text than in bilabial context. In the present study, both PF /y/ and /œ/ perceptually assimilated to AE /u/ almost twice as often in alveolar context than in bilabial context. This may be attributed, in part, to the greater frequency and possibly the more phonemic status of the palatalized /ju/ following bilabials (e.g., "beauty") than following alveolars in AE dialects. In alveolar context, e.g., in "tune," most American dialects have lost the glide preceding the /u/, although exceptions remain, primarily in southern dialects (Phillips, 1981).

Clearly, AE perceptual assimilation of PF vowels cannot be predicted entirely on the basis of the acoustic similarities of French vowels as measured by midsyllable formant frequencies. As shown in Fig. 1, PF /y/ is much closer in acoustic space to PF /i/ than to PF /u/. Yet, even in bilabial context, AE listeners overwhelmingly perceived PF /y/ as most similar to back rounded or palatalized back AE vowels. Instead, explanations must rely on AE listeners' perceptual categorization of French vowels based on their AE phonological system. A developmental linguistic explanation might be the redundancy of the "roundedness" feature in AE. From infancy, AE monolinguals learn to equate roundedness with "backness," as every AE rounded vowel is a "back" vowel (at least at an abstract level of representation) and mid to high back vowels are always rounded. This explanation alone, however, does not account for why the "redundancy effect" would be stronger in alveolar context than in bilabial.

A second explanation involves the allophonic variation of back vowels in AE. The bottom graph in Fig. 1 shows that high to mid back AE vowels (/u, o/ but also /ʊ/, not shown on graph) are fronted in alveolar context, i.e., produced with the tongue further forward in the oral cavity (Hillenbrand *et al.*, 2001; Strange *et al.*, 2007). Thus, in English, the /u/ is produced as back rounded /u/ in nonalveolar contexts (e.g., "boom"), but in alveolar context (e.g., "dude"), AE /u/ is a fronted vowel. AE listeners, then, may be more likely to categorize front rounded vowels as AE /u, ʊ, o/ when these vowels are surrounded by alveolar consonants than when they are in other contexts. This explanation would account for the increased number of AE /u/ responses to PF /y/ and /œ/ in alveolar context—the context in which fronting is most extreme.

The bottom graph in Fig. 1 also reveals PF /œ/ as only slightly fronted in alveolar context relative to its position in bilabial context (top graph), but the greater fronting of the AE back vowels causes the PF /œ/ to overlap with the AE /u/ vowel category, while remaining distinct from AE front vowels. Similarly, discriminant analysis by Strange *et al.* (2007) found PF /œ/ more spectrally similar to the (fronted) AE /u/ and /ʊ/ vowels in alveolar context than to AE /u/ and /ʊ/ in bilabial context in sentence materials. Thus, more assimilation of PF /œ/ to AE /u/ in alveolar context than in bilabial context can be understood based on these acoustical data, as can AE listeners' increased discrimination difficulty for PF /œ/ paired with back vowels in alveolar context (Levy and Strange, 2008). With more French experience, listeners tended to assimilate the PF vowel /œ/ primarily to AE /ʊ/ and /ɝ/ in both consonantal contexts. The selection of the rhotacized vowel /ɝ/ response may be due to the lip rounding and raising of the tongue (as occurs for front rounded vowels) that accompanies AE rhotacization.

C. Theoretical implications

Results from this study contribute to the literature indicating that variations in the acoustic realization of vowels as a function of consonantal context affect their perception in an unfamiliar language (Strange *et al.*, 2001; Strange *et al.*, 2004a) and now also in an L2 (e.g., Gottfried, 1984; Levy and Strange, 2008). In the terminology of the PAM (Best, 1995), the consonants surrounding vowels affect to which native vowel a non-native vowel will be assimilated and the goodness of fit to that category. For example, AE listeners perceptually assimilate PF /œ/ to AE /u/ more often in alveolar context than in bilabial context. The PAM, factoring in context, would predict that the /œ/ would be more difficult to differentiate from /u/ in alveolar context than in bilabial, a prediction borne out in Levy and Strange's (2008) study.

The relatively high goodness rating for PF /u/ by the HiExp group in the present study supports Flege's (1987) claim that PF /u/ is perceived as a good instance of AE /u/ and thus might be produced in an equivalent, i.e., accented, manner. In terms of the SLM (Flege, 1995), the data are consistent with category formation in bilabial context but not in alveolar context. The context-specific categorization revealed here suggests an allophonic level of representation (as discussed by Flege, 1995) operating in equivalence classification, such that listeners perceive vowels as "new" or "similar," depending on their phonetic environment. Thus, /y/ could be "equivalent" to AE /u/ in alveolar context and "new" in bilabial. However, the relatively low goodness ratings for /y/ in both contexts for the HiExp group suggest that as L2-learners became familiar with French, they began to discern differences between /y/ and AE vowels.

Evidence suggests that listeners do not encode contrasts in terms of context-independent phonemic units (Pisoni *et al.*, 1994). Such abstract units could not explain the context effects found in the present study and in numerous other studies (e.g., Gottfried, 1984; Levy and Strange, 2008; Strange *et al.*, 2001; 2004a). Clearly, neither an abstract, phonemic analysis nor an acoustic description of L2 segments will adequately predict the way in which an L2 segment is learned. For now, it may be concluded that L2 segments are initially perceived on an intermediate, context-sensitive allophonic level. Learning an L2 ideally involves the formation of new phonological categories, including knowledge about the systematic variations that exist within each L2 category (Flege, 1995). Although perception and production of segments might improve with experience (Flege, 1987; Gottfried, 1984; Levy and Strange, 2008) and with perceptual training (Bradlow *et al.*, 1997; Iverson *et al.*, 2005; Rvachew, 1994), even the most experienced late learners' native phonological knowledge (including language-specific allophonic rules) may continue to influence their L2 perception.

TABLE IV. Statistical significance on comparisons performed for regression analysis (*=significance at the $p=0.05$ level; nonsignificance=ns at the $p=0.05$ level; n/a=not applicable because few responses in any category were other than modal).

Vowel	Comparison	Language experience (in at least one context): NoExp vs ModExp	Language experience (in at least one context): ModExp vs HiExp	Language experience (in at least one context): NoExp vs HiExp	Consonantal context	Interaction
i	n/a	n/a	n/a	n/a	n/a	n/a
ɛ	ɛ vs e	ns	ns	ns	ns	ns
	ɛ vs æ	*	ns	*	ns	ns
a	æ vs ɑ	ns	ns	ns	*	ns
	æ vs ɛ	*	*	*	ns	ns
o	o vs u	*	*	*	*	*
u	ʊ vs ^{j}u	n/a	n/a	n/a	n/a	n/a
y	^{j}u vs u	*	*	*	*	*
œ	ʊ vs u	*	*	*	*	ns
	ʊ vs o	*	*	*	*	ns
	ʊ vs ɜ˞	ns	*	*	ns	ns

ACKNOWLEDGMENTS

This research was funded by a grant to the author (NIH-NIDCD Grant No. 1F31DC006530-01). Special thanks are due Winifred Strange. The author also gratefully acknowledges Loraine Obler, Martin Gitterman, Catherine Best, James Jenkins, Kanae Nishi, Valeriy Shafiro, Gary Chant, Miwako Hisagi, Franzo Law II, Bruno Tagliaferri, Natalia Martínez, and the Teachers College Speech Production and Perception Laboratory for their contributions to all aspects of this project.

APPENDIX: REGRESSION COMPARISONS AND STATISTICAL SIGNIFICANCE

Table IV shows the comparisons using the regression analysis and their statistical significance or nonsignificance.

[1]The front rounded vowels /ø/ and /œ/ are almost never contrastive in PF. For the present purposes, /œ/ will represent the midfront rounded vowel.

[2]In cross-speaker tasks, the speaker differs across A, X, and B tokens; thus, listeners must make judgments on the basis of speaker-independent categorical representations of the stimuli.

[3]One of the PF speakers' stimuli had also been used in Levy and Strange (2008), but an additional token of each of her vowels plus three tokens each of /o/ and /ɛ/ were used here. Of the 39 listeners in this study, one HiExp participant had also participated in Levy and Strange. However, three years had passed, thus learning effects were expected to be minimized.

[4]Although listeners in the HiExp group were generally of a more advanced age than those in the other groups, it is not expected that this affected the results, as the HiExp group performed similarly to the other groups on the "control" vowel /i/ and on other vowel stimuli.

[5]To determine whether F3, which is lowered with lip rounding, might have contributed to the perceptual patterns, F1 values were subtracted from F2 values and F2 values from F3 values for /y/, /u/, and /i/ in each context (e.g., F3−F2=1.1 bark for /y/, 2.3 bark for /i/, and 7.7 bark for /u/ in bilabial context). Results indicated that even taking F3 into consideration, /y/ is acoustically more similar (i.e., closer in vowel space) to /i/ than to /u/ in PF.

[6]The acoustical data from four tokens of vowels produced by three PF and three AE female speakers in the study of Strange *et al.* (2007) reveal more separation between PF /u/ and PF /o/ vowels and less of a shift in F1 between PF and AE vowels than in the present study; thus the differences seen here in F1 may be a function of individual differences related to the vocal tract size. In the present study, the HiExp listeners assimilated PF /o/ to AE /o/ on 98% of trials (as opposed to naïve listeners, who assimilated PF /o/ to AE /o/ on 32% of trials and to AE /u/ on 53% of trials), suggesting that with experience, listeners had become skilled at sorting out differences between PF /o/ and /u/, perhaps by reference to point vowels, which also shifted in vowel space.

[7]Fisher's exact test (appropriate for zero-cell counts) was used for comparisons with HiExp groups in alveolar context for PF /o/, as no participant in the HiExp group selected AE /u/ in response to PF /o/ in alveolar context; thus the regression analysis could not be performed.

[8]An argument could be made that if /^{j}u/ had not been a response choice, more AE /i/ responses would have been chosen. Studies that have not used /^{j}u/ as a response alternative in perceptual assimilation have found assimilation of PF /y/ primarily to AE back vowels. For example, Strange *et al.* (2004b) found that naïve listeners assimilated PF /y/ in sentence context to back vowels on 84% of responses. This suggests that without the inclusion of a /^{j}u/ response option, back-vowel responses would have been similar or slightly fewer in number.

[9]When /^{j}u/ and PF /u/ responses were combined and compared to the resulting second-most selected choice, AE /ʊ/, the overall language experience effect was statistically significant [$\chi^2(4)=15.64$, $p<0.0035$], as was the context effect [$\chi^2(2)=20.92$, $p<0.0001$]. However, as /ʊ/ was chosen by fewer than 10% of listeners, comparisons with /ʊ/ may not be meaningful.

Agresti, A. (**2007**). *An Introduction to Categorical Data Analysis*, 2nd ed. (Wiley-Interscience, New York).

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., Faber, A., and Levitt, A. (**1996**). "Assimilation of non-native vowel contrasts to the American English vowel system," J. Acoust. Soc. Am. **99**, 2602.

Best, C. T., McRoberts, G. W., and Sithole, N. M. (**1988**). "Examination of perceptual reorganization for non-native speech contrasts: Zulu click discrimination by English-speaking adults and infants," J. Exp. Psychol. **14**, 345–360.

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 (Benjamin, Amsterdam), pp. 13–34.

Bohn, O.-S., and Steinlen, A. K. (**2003**). "Consonantal context affects cross-language perception of vowels," Proceedings of the 15th International Congress of Phonetic Sciences, pp. 2289–2292.

Bradlow, A. R., Pisoni, D. B., Akahane-Yamada, R., and Tohkura, Y. (**1997**). "Training Japanese listeners to identify English /r/ and /l/: Some effects of perceptual learning on speech production," J. Acoust. Soc. Am. **101**, 2299–2310.

Burnham, D., and Mattock, K. (**2007**). "The perception of tones and phones," 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. 259–280.

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.

Flege, J. E. and Hillenbrand, J. (**1984**). "Limits on phonetic accuracy in foreign language speech production," J. Acoust. Soc. Am. **76**, 708–721.

Gottfried, T. L. (**1984**). "Effects of consonant context on the perception of French vowels," J. Phonetics **12**, 91–114.

Guion, S. G., Flege, J. E., Akahane-Yamada, R., and Pruitt, J. C. (**2000**). "An investigation of current models of second language speech perception: The case of Japanese adults' perception of English consonants," J. Acoust. Soc. Am. **107**, 2711–2724.

Harnsberger, J. D. (**2001**). "On the relationship between identification and discrimination of non-native nasal consonants," J. Acoust. Soc. Am. **110**, 489–503.

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.

Iverson, P., Hazan, V., and Bannister, K. (**2005**). "Phonetic training with acoustic cue manipulations: A comparison of methods for teaching /r/-/l/ to Japanese adults," J. Acoust. Soc. Am. **118**, 3267–3278.

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, Graduate School and University Center, City University of New York.

Levy, E. S., and Law, F., II (**2008**). "Production of Parisian French front rounded vowels by second-language learners," J. Acoust. Soc. Am. **123**, 3078.

Levy, E. S., and Strange, W. (**2008**). "Perception of French vowels by American English adults with and without French language experience," J. Phonetics **36**, 141–157.

Phillips, B. S. (**1981**). "Lexical diffusion and southern tune, duke, news," Am. Speech **56**, 72–78.

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 Words*, edited by H. C. Nusbaum and J. Goodman (MIT, Cambridge), pp. 121–166.

Polka, L. (**1995**). "Linguistic influences in adult perception of non-native vowel contrasts," J. Acoust. Soc. Am. **97**, 1286–1296.

Pufahl, I., Rhodes, N. C., and Christian, D. (**2001**). "What we can learn from foreign language teaching in other countries," ERIC Clearinghouse on Languages and Linguistics, Center for Applied Linguistics, Washington, DC., pp. 1–9.

Rochet, B. L. (**1995**). "Perception and production of second-language speech sounds by adults," in *Speech Perception and Linguistic Experience: Issues in Cross-Language Research*, edited by W. Strange (York, Timonium, MD), pp. 379–410.

Rvachew, S. (**1994**). "Speech perception training can facilitate sound production learning," J. Speech Hear. Res. **37**, 347–57.

Seligman, M., Maier, S. F., and Geer, J. (**1968**). "The alleviation of learned helplessness in dogs," J. Abnorm. Psychol. **73**, 256–262.

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., Akahane-Yamada, R., Kubo, R., Trent, S. A., and Nishi, K. (**2001**). "Effects of consonantal context on perceptual assimilation of American English vowels by Japanese listeners," J. Acoust. Soc. Am. **109**, 1691–1704.

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., 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., Levy, E. S., and Lehnhoff, R., Jr. (**2004b**). "Perceptual assimilation of French and German vowels by American English listeners: Acoustic similarity does not predict perceptual similarity," J. Acoust. Soc. Am. **115**, 2606.

Strange, W., Weber, A., Levy, E. S., Shafiro, V., Hisagi, M., and Nishi, K. (**2007**). "Acoustic variability within and across German, French and American English vowels: Phonetic context effects," J. Acoust. Soc. Am. **122**, 1111–1129.

Tranel, B. (**1987**). *The Sounds of French* (Cambridge University Press, New York).

Williams, M., Burden, R., Poulet, G., and Maun, I. (**2004**). "Learners' perceptions of their successes and failures in foreign language learning," Lang. Learn. **30**, 19–29.

Intelligibility of interrupted sentences at subsegmental levels in young normal-hearing and elderly hearing-impaired listeners[a)]

Jae Hee Lee[b)] and Diane Kewley-Port[c)]
Department of Speech and Hearing Sciences, Indiana University, Bloomington, Indiana 47405

(Received 16 December 2007; revised 8 October 2008; accepted 13 October 2008)

Although listeners can partially understand sentences interrupted by silence or noise, and their performance depends on the characteristics of the glimpses, few studies have examined effects of the types of segmental and subsegmental information on sentence intelligibility. Given the finding of twice better intelligibility from vowel-only glimpses than from consonants [Kewley-Port *et al.* (2007). "Contribution of consonant versus vowel information to sentence intelligibility for young normal-hearing and elderly hearing-impaired listeners," J. Acoust. Soc. Am. **122**, 2365–2375], this study examined young normal-hearing and elderly hearing-impaired (EHI) listeners' intelligibility of interrupted sentences that preserved four different types of subsegmental cues (steady-states at centers or transitions at margins; vowel onset or offset transitions). Forty-two interrupted sentences from TIMIT were presented twice at 95 dB SPL, first with 50% and second with 70% of sentence duration. Compared to high sentence intelligibility for uninterrupted sentences, interrupted sentences had significant decreases in performance for all listeners, with a larger decrease for EHI listeners. Scores for both groups were significantly better for 70% duration than for 50% but were not significantly different for the type of subsegmental information. Performance by EHI listeners was associated with their high-frequency hearing thresholds rather than with age. Together with previous results using segmental interruption, preservation of vowels in interrupted sentences provides greater benefit to sentence intelligibility compared to consonants or subsegmental cues.
 [DOI: 10.1121/1.3021304]

PACS number(s): 43.71.Ky, 43.71.Es, 43.66.Sr [MSS] Pages: 1153–1163

I. INTRODUCTION

In everyday listening situations, both the target speech and background noise continuously fluctuate, especially when the noise consists of competing speech or dynamic environmental noise. Because these fluctuations are usually independent, the speech signal may be only partially audible or even completely inaudible depending on the relative levels of the target and the noise. Listeners must then integrate individual glimpses of target information within the dips or valleys of fluctuating noise in order to understand the target's entire meaning. Previous studies of temporal interruption demonstrated that when young normal-hearing (YNH) listeners glimpse the full spectrum of speech, their performance is relatively high for 8–10 Hz of silence or noise interruption at a 50% duty cycle, regardless of the type of stimulus materials such as monosyllabic words (Miller and Licklider, 1950; Kirikae *et al.*, 1964), sentences (Bergman *et al.*, 1976; Bergman, 1980; Nelson and Jin, 2004; Iyer *et al.*, 2007; Li and Loizou, 2007), and connected discourse passage (Powers and Speaks, 1973).

Two studies have examined the ability of elderly hearing-impaired (EHI) listeners to integrate glimpses in target sentences that were either periodically interrupted (Gordon-Salant and Fitzgibbons, 1993) or segmentally (consonants versus vowels) interrupted (Kewley-Port *et al.*, 2007). As expected, both studies have found that EHI listeners showed poorer recognition of interrupted sentences than did YNH listeners. Although the overall presentation level of the speech portions glimpsed was above the hearing thresholds of EHI listeners up to 4 kHz (i.e., 85–90 dB SPL), the results of both studies also indicated that variability in intelligibility scores was better accounted for by high pure-tone threshold average (PTA) (averaged hearing thresholds at 1, 2, and 4 kHz) than by age.

Kewley-Port *et al.* (2007) reported that intelligibility for both YNH and EHI listeners was two times better when only vowels remained in sentences compared to when only consonants remained (this study is referred to as KBL07 in the rest of this article). This finding replicated previous results for YNH listeners by Cole *et al.* (1996) of a ratio of 2:1 for intelligibility in vowel-only sentences versus consonant-only sentences for various types of interruption noises (no noise, harmonic complexes, and white noise). Together, these results suggest that as long as sentences are reasonably audible, intelligibility of segmentally interrupted speech for both YNH and EHI listeners depends strongly on the type of segmental information glimpsed. Because the importance of high-frequency information for consonants has been long

[a)]Portions of the data were presented at the 151th Meeting of the Acoustical Society of America and at the fourth Joint Meeting of the Acoustical Society of America and the Acoustical Society of Japan.
[b)]Author to whom correspondence should be addressed. Present address: Department of Audiology, Institute of Audiology, Hallym Institute of Advanced International Studies, Korea. Electronic mail: leejaehee@hallym.ac.kr
[c)]Electronic mail: kewley@indiana.edu

emphasized for the clinical amplification in the hearing aids, these findings of the greater contribution of vowels compared to consonants for sentence intelligibility are noteworthy.

The motivation for the present study was to examine contributions of glimpsing vowels versus consonants from somewhat different but theoretically based definitions of vowel and consonant information. Traditionally steady-state acoustic cues have relatively less spectral change over time and specify vowel information (Peterson and Barney, 1952), whereas dynamic transition cues have more spectral changes over time and specify consonant information (Liberman *et al.*, 1957). However, another theoretical point of view (dynamic specification theory) (Strange and Bohn, 1998) documented in a series of studies by Strange and colleagues (Strange *et al.*, 1976; Strange, 1989; Jenkins *et al.*, 1994) stresses the importance of dynamic transition cues in vowel perception. Their studies concentrated primarily on the contribution of dynamic transition information at CVC margins to vowel identification. Thus different theories of what specifies vowel versus consonant information suggest that there exist regions of subsegmental information in speech that contribute differentially to intelligibility. Four subsegmental regions, steady-state centers versus dynamic margins, and onset versus offset transitions were the focus of this study. The outcome should demonstrate whether there are more important information-rich subsegmental regions of speech that result in better sentence intelligibility. If this was found for EHI listeners, then those regions should be considered to be preserved or enhanced in the future design of speech processors for hearing assistive devices for EHI listeners.

Few studies have been conducted on the ability of older listeners to use dynamic cues glimpsed from nonsense syllables, and the results were not in agreement with each other. Fox *et al.* (1992) reported an age-related decrement in the ability to use dynamic cues in CVC margins for vowel and consonant perception among various age groups with relatively normal hearing, supporting age-related deficit hypothesis. Ohde and Abou-Khalil (2001), however, found the similar abilities among young, middle-aged, and older adults who had near-normal hearing for their age (i.e., less than 40 dB HL at 4000 Hz in older adults) to use these dynamic formant transition cues for vowel and consonant perception, thereby not supporting age-related deficits in using dynamic cues. Dorman *et al.* (1985) reported that differences in performance among YNH, elderly normal-hearing (ENH), and EHI listener groups were not consistent across various phonetic identification tasks, but rather were varied depending on the type of vowels and consonants contrasted. We note, however, that it is not clear how these previous, somewhat conflicting, results for older listeners' use of dynamic versus static cues in CVC syllables for vowel identification can be generalized to sentence intelligibility for EHI listeners.

The present study employs the same TIMIT sentences (Garofolo *et al.*, 1990) used by KBL07 and examines how four regions of subsegmental glimpsing cues might differentially contribute to sentence intelligibility for both YNH and EHI listeners. Centers or margins within each segment were selected as the first and the second subsegmental target regions to focus on the contributions of quasi-steady-state versus transition information to sentence intelligibility and were primarily motivated by Strange's theory. Full vowel onset or offset transitions were selected as the third and the fourth target regions based on long-standing results that acoustic cues for consonants in CVC syllables have been found to be more salient at the onsets of syllables compared to the offsets (Redford and Diehl, 1999). Two different durations of glimpsing (50% and 70% durations of each segment) were employed.

The objectives of the present study were to examine the following questions: (1) would EHI listeners show significantly reduced ability to integrate subsegmental cues glimpsed from interrupted but audible (95 dB SPL) sentences than YNH listeners; (2) would dynamic transition cues result in equivalent intelligibility in sentence recognition in listeners compared to quasi-steady-state cues, similar to the role of dynamic cues shown for vowel identification; (3) what impact would different subsegmental cues (i.e., phoneme steady-state at centers versus transitions at margins; vowel onset versus offset transition regions) have on the ability of listeners to recognize interrupted sentences; and (4) how would performance improve as the duration of speech portions glimpsed increases from 50% to 70%. In addition, correlational analyses examined the relation between individual differences in the performance of EHI listeners with the variables of hearing loss or age.

II. GENERAL METHODS

A. Overview of experimental design

To investigate the effect of four regions of subsegmental cues on sentence intelligibility for YNH and EHI listeners, a mixed design was developed with three variables: two between-subject variables (two listener groups and four stimulus conditions) and one within-subject variable (two durations of glimpsing, 50% and 70%). The 42 TIMIT test sentences (Texas Instruments/Massachusetts Institute of Technology) (Garofolo *et al.*, 1990) that were used in KBL07 were employed in this study as test materials. These 42 sentences were interrupted with the same speech-shaped noise (SSN) in KBL07, but in four different ways to preserve four different subsegmental cues depending on the regions within each segment. The four conditions (see Fig. 2) preserved four different subsegmental regions focusing on either phoneme steady-state or three transition cues as follows: (1) the center region of vowel and consonant segments that are generally quasi-steady-states in each segment (CENTER), (2) the two margin regions of each vowel and consonant segments where generally the formant transitions in each segment are found (MARGIN), (3) the final portion of a consonant and the initial portion of following vowel incorporating the vowel onset transitions (ONSET), and (4) the final portion of a vowel and the initial portion of the following consonant incorporating the vowel offset transitions (OFFSET). Each test sentence was presented twice with two durations allowing glimpses of the target speech, 50% (50% speech-on and 50% noise-on, alternately) and 70% (70% speech-on and 30% noise-on, alternately) of the duration of

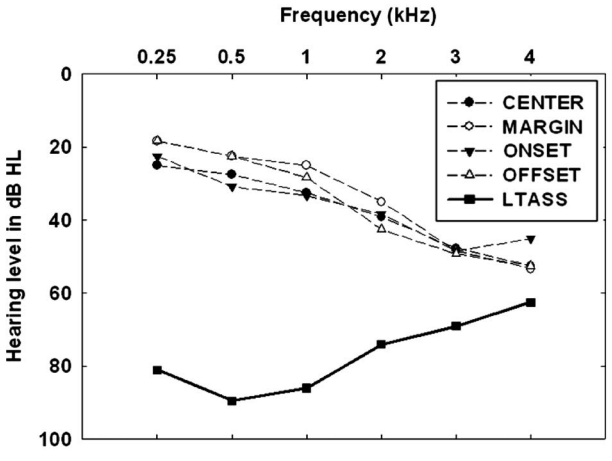


FIG. 1. Average pure-tone thresholds for the test ear of EHI listeners across CENTER, MARGIN, ONSET, and OFFSET conditions ($N=6$ per condition) displayed by broken lines in dB HL (ANSI, 1996). Solid line displays the LTASS in dB HL when calibrated at the 95 dB SPL level.

each segment. In a pilot study, the 50% duration yielded near zero performance for some EHI listeners. Therefore, the 70% duration of the target sentence was added as a second presentation to both listener groups. Details of the methods follow.

B. Participants

Twenty-four YNH and 24 EHI listeners were paid to participate. All listeners were native American-English speakers. Each listener group was gender balanced (12 males and 12 females). YNH and EHI listeners were recruited from Indiana University and from the Indiana University Hearing Clinic, respectively. To participate, all listeners were required to have a passing score (>27/30) on the Mini-Mental Status Examination (MMSE) (Folstein *et al.*, 1975) for cognitive status, a score of 5 or greater on the auditory forward digit span test, and a score of 4 or greater on the auditory backward digit span test for normal short-term memory. YNH listeners ranged in age from 20 to 35 years ($M=27$ years), and had pure-tone thresholds no greater than 20 dB HL at octave intervals from 250 to 8000 Hz (ANSI, 1996). EHI listeners ranged from 65 to 80 years of age ($M=74$ years). The hearing criteria for EHI listeners were normal middle ear status, and postlingual, bilateral, high-frequency, mild-to-moderate hearing loss of cochlear origin with pure-tone thresholds less than 60 dB HL from 2000 to 4000 Hz. With a quasi random assignment, participants listened to the 42 test sentences in one of the four conditions. To reduce individual variability in performance, age and hearing loss were matched carefully across the four conditions. In Fig. 1, each of the dashed lines shows the average air-conduction pure-tone thresholds for the tested ear in EHI listeners for each of the four conditions (CENTER, MARGIN, ONSET, and OFFSET). The solid line displays the level of the long-term-averaged speech spectrum (LTASS) of target speech in dB HL. As shown, target speech was presented to EHI listeners above pure-tone thresholds from 250 to 4000 Hz (e.g., at least 9 dB above pure-tone threshold at 4000 Hz). Results of Levene's test for equality of variances and a one-way analysis of variance (ANOVA) revealed that there were homogeneous variances as well as no significant differences in both age and pure-tone thresholds for six EHI listeners in each of the four conditions.

C. Stimuli

KBL07 used 42 sentences (21 male and 21 female speakers) from the TIMIT corpus as test material. Speakers were from the North Midland dialect region that matched the catchment area of the participants in this study, namely, Indianapolis, IN, and further north.

In the TIMIT database, segmental boundaries and phonetic transcriptions were established by expert phoneticians. KBL07 verified the segmental boundaries provided by the TIMIT corpus. KBL07 added three minor rules appropriate for identifying the vowels and consonants in sentences: (1) stop closure symbols were combined with the following stop and treated as a single consonant; (2) syllable V+[*r*] was considered as a single rhotocized vowel; and (3) the glottal stop [q] occurred between two vowels such that; [VqV] was treated as a vowel. These three rules were also used in the present study. As in the previous study, in order for all the TIMIT sentences needed to be sufficiently audible for both listener groups, 95 dB SPL was used as the signal level to present the sentences after digitally scaling them to a constant rms (root-mean-square) value (for more details, see Sec. II E).

D. Processing for interrupted sentences

1. Speech information in interrupted sentences

The test sentences were interrupted at specified subsegmental intervals with low-level SSN in four different conditions. Each condition presented one of four different regions of subsegmental glimpsing (CENTER, MARGIN, ONSET, and OFFSET) with either a 50% or 70% duration. Figure 2 shows temporal waveforms of an example CVC word "mean" extracted from a test sentence. The top waveform of "mean" has no interruption. The remaining waveforms of "mean" show four regions of subsegmental glimpsing in which 50% of glimpsing duration was applied.

As shown in Fig. 2, the 50% duration yields two pairs of conditions, CENTER/MARGIN (second and third waveforms) and ONSET/OFFSET (fourth and fifth waveforms) that present complementary acoustic pieces of the subsegmental intervals. The CENTER regions preserved 50% of the center portions of vowel and consonant segments, which represented subsegmental information containing the quasi-steady-state parts of vowels and consonants. As a complementary condition, the MARGIN regions preserved 25%of each of the two margin portions of the vowel and consonant segments, which contained mostly the spectral transitions of the vowels and consonants, similar to stimuli used by Strange *et al.* (1976). The ONSET and OFFSET conditions focused on different types of vowel transitional information. The ONSET condition preserved vowel onset information by capturing transitional information from the last 50% of a consonant preceding a vowel and the initial 50% of the following vowel. The complementary OFFSET preserved

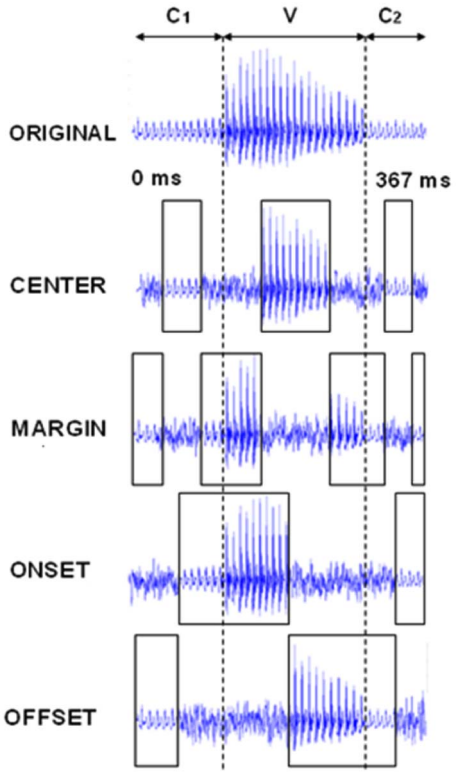


FIG. 2. (Color online) Temporal waveforms of the C_1VC_2 syllable "mean" (/min/) extracted from a test sentence, "What did you mean by that rattlesnake gag?". Top waveform of "mean" is from the original sentence with no interruption. Portions inside the boxes from second to fifth waveforms display four different subsegmental cues preserved in each condition (CENTER, MARGIN, ONSET, and OFFSET from second to fifth waveforms) for the 50% proportion of duration, while the regions outside the boxes are replaced with SSN.

vowel offsets by capturing the last 50% of a vowel and the initial 50% of the following consonant. As expected, the TIMIT sentences include not only CVCs shown in Fig. 2 but also many consonant and vowel clusters such as CCVC, CVCC, and CVVC. The clusters were considered as a single syllable (for example, CC→C, VV→V).

The purpose of the 70% of glimpsing duration was to reduce the amount of interruption in sentences such that this duration would elevate the near floor performance of some EHI listeners in the 50% duration observed in pilot testing. 10% more information on either side of the 50% duration was added to each preserved subsegmental unit to comprise the 70% glimpsing duration of target speech. MATLAB scripts were used in conjunction with the TIMIT boundaries to calculate the glimpsing duration and insert the noise for all four conditions.

2. Noise in interrupted sentences

The SSN was generated by MATLAB and was used to replace parts of the sentences. The SSN shape was based on a standard LTASS (ANSI, 1969) that had a flat shape of 0–500 Hz and a −9 dB/octave roll-off above 500 Hz. The present study attempted to set the level of SSN to be low relative to the vowel, yet be audible to EHI listeners with mild-to-moderate hearing loss. Presumably the low-level noise would smooth out the somewhat choppy sentences, reduce boundary transients, and encourage phoneme restoration (Warren, 1970). After informal listening in a pilot test, 75 dB SPL (i.e., 20 dB lower than the average level of 95 dB SPL) was chosen to be the level of the SSN.

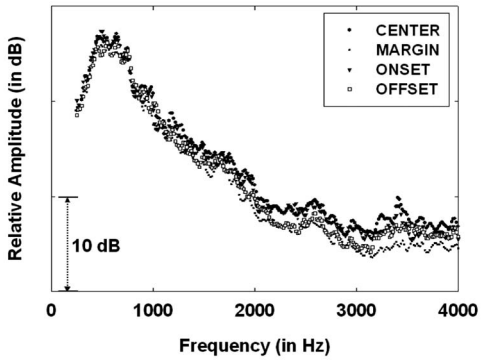


FIG. 3. Root-mean-square amplitude spectra of the concatenated 42 sentences processed in the four conditions (CENTER, MARGIN, ONSET, and OFFSET) as a function of frequency. Note that all sentences used in the current study were low-pass filtered at 4400 Hz.

E. Calibration

Similar calibration procedures from KBL07 were administered to verify signal levels. First, scripts were written in MATLAB to verify that all the 42 test sentences had similar average rms levels (i.e., within ±2 dB). Second, a MATLAB script was used to find the most intense vowel across all sentences and then iterated that vowel to produce a calibration vowel of 4 s. To avoid effects of hearing loss beyond 4000 Hz in EHI listeners, all sentences were filtered by a low-pass finite impulse response filter that was flat to 4000 Hz with a 3 dB cutoff at 4400 Hz and a 200 dB/octave steep slope in a Tucker Davis Technologies (TDT) PF1. The sound level for the calibration vowel was also low-pass filtered, and its sound level was set to 100 dB SPL through ER-3A insert earphones in a HA-2.2 cm^3 coupler using a Larson Davis model 2800 sound level meter with linear weighting. Relative to the loudest calibration vowel, the mean of the distribution of the loudest vowels in the other sentences was 95 dB SPL, being the nominal level referenced to this study. An additional low-level background noise was continuously presented during testing. The purpose of this noise was to reduce transients between speech and noise. This noise was generated by the TDT WG2 and was also low-pass filtered at 4400 Hz. The level of this noise was reduced by more than 50 dB compared to the calibration vowel (100 dB SPL) measured using the same equipment described above.

Spectral analyses were used to verify that the long-term spectra for each of the conditions were similar. All 42 test sentences were concatenated together after eliminating pauses. Long-term spectra were calculated with a Hanning window and the Baum–Welch algorithm in MATLAB. We confirmed very similar spectral envelopes (i.e., within ±3 dB) across the frequency range of 0–4000 Hz for each of the four conditions (CENTER, MARGIN, ONSET, and OFFSET), as shown in Fig. 3.

F. Test procedures

1. Stimulus presentation

All screening tests for hearing and cognitive functions were administered before testing. Each listener was instructed about the tasks using written and verbal instructions. Test stimuli were controlled by TDT system II hardware connected to a personal computer and were presented through ER-3A insert earphones to listeners in a sound-treated booth. Test sentences were presented to the better ear of the hearing-impaired listeners (generally the right ear) and to the right ear for the normal-hearing listeners. Six familiarization sentences consisted of two unprocessed sentences (i.e., no interruption) and four processed sentences. The four processed sentences were presented two times, with 50% and 70% durations of glimpsing corresponding to each experimental condition, and then the unprocessed sentence was given as feedback.

The 42 test sentences were randomized and then presented in one fixed order. Listeners heard the 42 sentences twice, first with the 50% duration and second with the 70% duration. After each presentation listeners were asked to respond by repeating verbally any words they thought they heard from the test sentence. All responses were recorded by a digital recorder. Listeners were encouraged to guess, regardless of whether the responded words or partial words made sensible sentences. No feedback was provided. The correctly identified words were scored by the experimenter during testing and then rechecked by a linguist from the recorded responses later. Experimental testing lasted 1 h for YNH listeners and 1.5 h for EHI listeners.

2. Scoring and data analysis

The number of correctly identified words was counted and scored as percentage of correct words relative to the total number of words in the sentences. All the words were scored as correct only when they exactly matched with the target words (i.e., incorrect for morphological variants). The purpose of this word scoring was to compare overall ability to understand interrupted sentences between YNH and EHI listeners. All the correct words were scored from the recorded responses by a second scorer who was a native American-English listener and a linguistics doctoral student. All discrepancies for the scoring of correct words were resolved by consensus between experimenter and the second scorer using the recorded responses. Word scores in percentage were transformed into rationalized arcsine units (RAUs) (Studebaker, 1985) for all statistical analyses. Statistical tests were based on a general linear model ANOVA with repeated-measures in SPSS statistical software (version 14.1; SPSS Inc., Chicago, IL)

III. RESULTS

A. Intelligibility for interrupted sentences

The scores of all listeners were obtained for the two uninterrupted (100% duration) sentences from the familiarization task. The averaged score of these two familiarization sentences presented at 95 dB SPL was 100% for YNH and 92% for EHI listeners. This accuracy was comparable to the results of KBL07 that reported a range of 97–100% for YNH and 88–99% for EHI listeners for the 14 uninterrupted sentences at 95 dB SPL. Thus full sentences at 95 dB SPL were reasonably audible up to 4000 Hz for both YNH and EHI listeners. Figure 4 shows the mean percentage of words correct in sentences across four conditions. The bars from left to right display scores obtained from YNH and EHI listeners with the 50% duration (YNH-50 and EHI-50) and then with the 70% duration (YNH-70 and EHI-70).

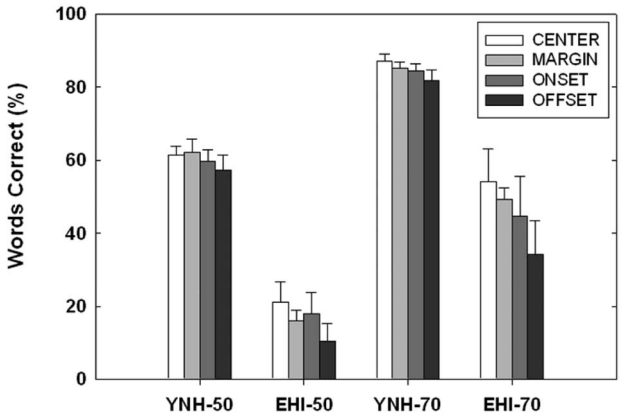


FIG. 4. Mean number of words correct in percentage in each of four conditions for YNH and EHI listeners with 50% duration (YNH-50 and EHI-50) and for YNH and EHI listeners with 70% duration (YNH-70 and EHI-70). Error bars indicate standard errors.

In general, EHI listeners performed poorer than YNH listeners, regardless of duration. For the 70% duration across all conditions, YNH listeners identified words in sentences about 85% correct, whereas EHI listeners identified words only about 46% correct. Compared to scores for uninterrupted sentences, 30% noise interruption resulted in a 15% decrease in scores (from 100% to 85% correct) for YNH but a 46% decrease (from 92% to 46% correct) for EHI listeners. For the 50% duration across all conditions, YNH listeners identified words in sentences about 60% correct, whereas EHI listeners identified words only about 16% correct. Compared to uninterrupted sentences, a 50% interruption yielded a 40% decrease (from 100% to 60% correct) for YNH but a 76% decrease (from 92% to 16% correct) for EHI listeners.

An ANOVA with repeated-measures was calculated for two between-subject variables (two groups × four conditions) and one within-subject repeated-variable (two durations) with the dependent variable of words correct in RAU. Results showed a significant ($p<0.05$) main effect of listener group [$F(1,40)=112.5$] and a significant main effect of duration [$F(1,40)=1321.9$] but no significant main effect of condition [$F(3,40)=1.5$, $p=0.23$]. As expected, results indicate that YNH outperformed EHI overall, and performance with the 70% duration was better than that with 50% duration. Unexpectedly performance was similar across four different subsegmental conditions. Although scores showed that EHI performed the best in CENTER and the worst in OFFSET regardless of the duration presented, a significant effect of condition was not obtained due to large individual differences and the small sample size ($N=6$) of EHI listener as-

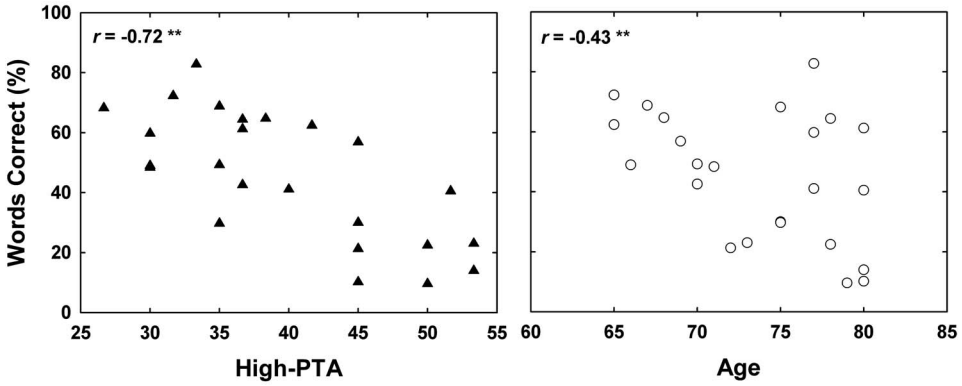


FIG. 5. Scatter plots of data for word scores by EHI-70 as a function of high-PTA (from 1, 2, and 4 kHz) in the left panel and as a function of age in the right panel.

signed to each of four conditions. There was one significant two-way interaction between group and duration [$F(1,40)=13.4$], indicating that the amount of benefit for the increased information in the 70% duration over the 50% duration was greater for EHI compared to YNH listeners.

B. Individual differences for EHI listeners

Not surprisingly, large individual variability was observed for temporally interrupted sentences for EHI listeners even though sentences were audible. Two previous studies in which interrupted sentences were presented to EHI listeners at high levels, 85 dB SPL (Gordon-Salant and Fitzgibbons, 1993) and 95 dB SPL (KBL07), reported stronger correlations between hearing loss and sentence intelligibility scores compared to that for age. The present study also examined the relation between age and hearing thresholds, either averaged or individual. First, correlational analysis results showed that age was not significantly correlated with high-frequency pure-tone thresholds (high-PTA) averaged at 1, 2, and 4 kHz ($p=0.06$, $r=0.39$). There was also no significant correlation between age and individual hearing thresholds from 0.25 to 8 kHz (with r ranging from 0.26 to 0.37 and p values from 0.07 to 0.67). With this in mind, the main correlational analyses investigated whether the large EHI individual differences in intelligibility are better predicted by high-PTA or age. Only scores for the 70% duration were analyzed because scores for the 50% duration were at floor for some listeners. Percentages of correct word scores of all EHI listeners ($N=24$) were averaged across the four subsegmental conditions (given no significant condition effect) and transformed to RAU.

High-PTA had a high negative correlation ($p<0.001$, $r=-0.72$), and age had a weaker, but significant, negative correlation ($p<0.004$, $r=-0.43$) with word scores. This relationship between hearing loss and word scores for EHI-70 is displayed in the left panel of Fig. 5, while the right panel plots relationship between age and performance. Forward stepwise regression analyses showed that high-PTA accounted for 51% of the variance in word scores [$F(1,22)=22.97$, $p<0.001$], while age was not a significant predictor ($p=0.14$). Apparently the large spread of word scores from the EHI listeners ranging 75–80 years, as displayed in Fig. 5, underlies weaker correlations obtained for age compared to high-frequency hearing loss. This dominant contribution of high-PTA rather than age suggests that factors associated with hearing loss contributed to decreased sentence intelligibility by EHI listeners. Thus, the higher-than-normal speech levels did not eliminate the negative effects of hearing loss on EHI performance.

IV. GENERAL DISCUSSION

A. Intelligibility with subsegmental versus segmental interruption

Forty-two TIMIT sentences were temporally interrupted allowing two types of speech portions to be glimpsed, either four subsegmental cues in this study or two segmental cues in a previous study [Kewley-Port *et al.*, 2007 (denoted as KBL07)]. In the present study, sentences were subsegmentally interrupted by SSN, preserving either quasi-steady-state versus three types of transition cues, with either a 50% or 70% of duration. In KBL07, segmental interruption replaced either consonants or vowels in the target sentences by SSN, resulting in vowel-in (VIN) and consonant-in (CIN) conditions. As expected, both studies reported a high performance for uninterrupted sentences for both YNH and EHI listener groups, but for interrupted sentences a significant reduction in sentence intelligibility was obtained for all listeners, with a larger drop for EHI. However, the contribution of subsegmental versus segmental cues to intelligibility of interrupted sentences was not the same. The contribution of four different subsegmental cues to sentence intelligibility was significantly affected by the duration of glimpsing but not by the regions of subsegmental information (steady-state or dynamic transitions). Unlike subsegmental cues, the types of segmental cues contributed differentially to sentence intelligibility (i.e., 2:1 better performance in VIN than in CIN conditions in KBL07). Below a direct comparison of the two studies was attempted because the stimuli and methods are similar (i.e., the same 42 sentences, the same overall presentation level, and similar criteria of hearing status and age for EHI listeners).

In order to compare the two studies, the approximate duration of vowels versus consonants preserved in sentences was calculated relative to total sentence duration using the segment boundaries in the TIMIT database. Specifically, the proportion of the sum of the duration of all vowels in the VIN condition relative to sentence duration was approximately 45%, while the proportion of consonant duration in the CIN condition was 55%. Note that this 55% of glimpsing

TABLE I. The first row shows the duration (%) in each condition across current and KBL07 studies. The score of words correct (%) with standard errors (SEs) is displayed in the second row for both YNH and EHI listeners.

	YNH			EHI		
Duration (%) (condition)	45 (VIN)	50 (Sub-50)	55 (CIN)	45 (CIN)	50 (VIN)	55 (Sub-50)
Words correct (%)	65.06	60.13	51.59	40.13	16.29	19.96
(SE)	(1.36)	(1.60)	(2.44)	(3.79)	(2.43)	(4.14)

duration for CIN compared to 45% of VIN was due to frequently occurring consonant clusters and while the average duration of individual consonants was actually less than that of the average vowel.

Table I shows the scores of words correct (%) for segmental conditions (VIN with 45% and CIN with 55% duration) and scores averaged across the four subsegmental conditions with 50% duration that were labeled as Sub-50. Although durations of 45%, 50%, and 55% are not particularly different in relation to the glimpsing opportunities of the target sentence, the highest word score was for VIN with the shortest, 45%, duration compared to the other conditions with longer durations. This pattern of greater performance in VIN than in others was more obvious in EHI listeners. This supports the previous finding (KBL07) that vowels contribute more to intelligibility of interrupted sentences than consonants, and relative to the current study more than any of four subsegmental cues.

To examine the differences shown in Table I in more detail, an additional one-way ANOVA measure was administered separately for YNH and EHI listener groups, with one between-subject variable (three conditions, VIN, Sub-50, and CIN) and the dependent variable of words correct in RAU. As expected from Table I, a significant ($p<0.05$) main effect of condition was found for YNH [$F(2,53)=12.1$] and for EHI [$F(2,53)=12.8$] listener groups. Results of a Bonferroni post-test indicated that differences in scores were significant between VIN and CIN and between Sub-50 and CIN but not significant between VIN and Sub-50 for YNH listeners. For EHI listeners, significant differences in performance were found between VIN and Sub-50 and between VIN and CIN, but not between Sub-50 and CIN conditions. These results confirm a strong benefit of vowels compared to other cues as the most important glimpsing source for EHI listeners. Note that a 10% difference in scores between Sub-50 ($N=24$) and CIN ($N=16$) reached significance in YNH listeners in Table I, although a 20% difference in scores between CENTER ($N=6$) and OFFSET ($N=6$) subsegmental conditions in EHI listeners (see Fig. 4) did not reach a statistical significance in the current study.

Power analysis was used to determine if the sample size ($N=6$) was too small between conditions to correctly accept the null hypothesis. A simple power analysis between the two most extreme conditions, CENTER ($M=54\%$) and OFFSET ($M=34\%$), revealed that although Cohen's d showed that effect size was large (1.05), power was only 0.446. To raise power to 0.80 (or 0.90) a sample size of $N=11$ (or $N=15$) would be required, approximately double the current sample size. Moreover, partial eta squared values observed from the SPSS ANOVA showed that the condition factor accounted for a very low, 13.1%, overall variance for EHI listeners, with even lower variance, 9.2%, for YNH listeners. These analyses suggest that our small sample size is not the reason underlying the negligible effect of the condition factor but rather that for both EHI and YNH listeners the manipulation of subsegmental information had no significant effect in these sentences.

Root-mean-square long-term amplitude spectra of the concatenated sentences in VIN, CIN, and Sub-50 were compared to investigate level differences across these conditions. Figure 6 shows that VIN sentences (displayed by unfilled circles) had overall 10 dB of level advantage than CIN sentences (displayed by unfilled squares), as reported in KBL07. The level of the concatenated sentences across the four subsegmental conditions (Sub-50, displayed by filled triangles) was very similar to the level of VIN sentences (i.e., within ± 3 dB). Considering this 6–7 dB of level advantage but the lower scores of EHI found in Sub-50 than in CIN, we speculate that the two times more frequent rate of interruption occurring in subsegmental condition than in segmental condition may be more challenging for EHI listeners to integrate the glimpses of the target speech. This hypothesis may be related to previous reports of auditory temporal processing deficits in older listeners [see reviews in Gordon-Salant (2005)].

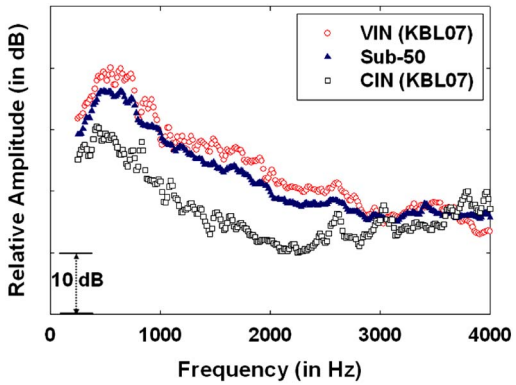


FIG. 6. (Color online) Spectra of the concatenated sentences used in VIN and CIN conditions (from KBL07) displayed as unfilled circles and unfilled squares, respectively. The filled triangles show the amplitude spectra of the concatenated sentences across the four subsegmental conditions with 50% duration (Sub-50).

B. Intelligibility of interrupted speech of young and older listeners

This section compares the present results with previous literature concerning intelligibility of interrupted sentences by younger and older listeners when 50% of duration was interrupted by either periodic or nonperiodic sound. For comparison purposes, we calculated from six randomly selected sentences that the approximate periodic rate of subsegmental interruption in our study for the 50% duration was about a 10 Hz rate, i.e., ten times of interruption per second. Because earlier reports using periodic interruption applied various interruption rates (0.1–10 000 Hz of interruption), glimpsing duration (6%–75% duration), and monaural versus binaural presentation, we selected only results that used a range of 8–10 Hz rate of periodic interruption, a 50% duration, and monaural presentation for comparison with our data.

60% to almost 100% correct intelligibility was found in young listeners with normal hearing when periodic interruption was applied to monosyllabic words (Miller and Licklider, 1950; Kirikae *et al.*, 1964), sentences (current data; Korsan-Bengtsen, 1973; Bergman *et al.*, 1976; Bergman, 1980; Nelson and Jin, 2004; Iyer *et al.*, 2007), and connected passages (Powers and Speaks, 1973). Findings from various studies suggest an apparent benefit for an approximately 10 Hz interruption rate compared to very low interruption rate (e.g., 1–2 Hz interruption). This 10 Hz periodic interruption rate might allow YNH listeners to have several glimpses at some essential segmental or subsegmental information regardless of the type of stimulus materials, whereas an entire syllable or word might be lost with a very low, 1–2 Hz, rate of interruption.

Only a few studies examined how well older listeners could integrate the glimpses of target sentences interrupted with the 8–10 Hz interruption rate using elderly listeners either with sensorineural hearing loss (current; Korsan-Bengtsen, 1973) or with near-normal hearing (Bergman *et al.*, 1976, Bergman, 1980). Similar to our study, elderly listeners in the studies above showed very low but consistent accuracy, 15%–16% correct, for identifying words in interrupted sentences. Given the high accuracy in performance by younger listeners across studies, we conclude that almost floor scores for interrupted sentences in older listeners indicate their general inability to successfully integrate information glimpsed from the target speech, regardless of types or difficulty of test materials.

An earlier study of Bergman *et al.* (1976) had 185 adult listeners ranging from 20 to 80 years of ages and showed a consistent drop in performance as a function of age (80%, 73%, 45%, 32%, 22% to 15% correct from 20 to 80 years of age). Based on this systematic drop as a function of age, they concluded that older listeners were less successful in integrating interrupted sentences due to age-related changes either in the auditory central nervous system or at a more cognitive level. Although this systematic drop in performance has significant implications, we note that Bergman *et al.* (1976) used a relatively lax hearing criteria for normal hearing (i.e., "35 dB at 0.5, 1, and 2 kHz and 40 dB at 4 kHz") and did not investigate hearing status as a factor.

It should be noted that our study did not include either young hearing-impaired listeners or ENH listeners to control for possible confounding contributions of aging and hearing loss. The correlational results in the current study, however, revealed that high-frequency hearing thresholds better predicted the intelligibility of interrupted sentences than did age, consistent with correlational results reported in previous findings (Gordon-Salant and Fitzgibbons, 1993; KBL07). This stronger role of peripheral high-frequency hearing loss rather than age as a predictor has been well documented in various measures of speech understanding. Humes (2007) compared results of various speech measures, showing strong correlation between high-PTA and simple level-raised speech performance. Therefore, it seems inconclusive whether the poorer performance of elderly listeners occurred mainly from age-related changes reported in earlier studies by Bergman. We note that while age-related cognitive deficits have been found with time-compressed speech (Wingfield *et al.*, 1985; Gordon-Salant and Fitzgibbons, 2004; Wingfield *et al.*, 2006), those results do not appear to make clear predictions for our temporal interruption studies at normal speech rates. While these issues are complex, performance in this study was primarily predicted by hearing loss. Thus our approach of using a high presentation level was worthwhile for demonstrating that signal level was not sufficient to eliminate the negative effects of high-frequency audibility when EHI participants listened to interrupted speech.

Based on the previous findings, we expected that YNH listeners' ability to integrate target sentences would be equivalent or even better from transition cues than from steady-state cues based on the importance of dynamic cues to vowel identification described by Strange [e.g., see Strange and Bohn (1998)]. On the other hand, EHI listeners were expected to benefit more from steady-state information compared to transition cues, based on age-related deficits in using dynamic cues discussed by Fox *et al.* (1992). No significant disadvantages to use four subsegmental cues was shown for our EHI participants, although we note that the high variability in EHI listener performance may have obscured possible subsegmental condition effects in this study. Our finding of no significant differences among steady-state and three transition regions of subsegmental cues to sentence intelligibility for both YNH and EHI listeners has some theoretical and clinical implications. Most of the previous studies on the role of dynamic transition cues primarily used monosyllabic words or nonsense words. Although this approach yielded well-controlled laboratory data that focused on the specific aims of each project, it was not known whether findings from word or syllable scores would generalize to sentence recognition because of various redundant cues preserved in sentences, as well as complex top-down processing used to comprehend incomplete sentences. Currently findings with syllables versus sentences do not correspond, suggesting that the contributions of transition cues found at the syllable or word level are different from contributions of transition cues to overall sentence recognition. Thus, future research on the importance of specific speech cues for EHI listeners should be established not only with words or syllables but also with

TABLE II. Examples of the four error types for incorrect words (phonetically matched word, MW; phonetically unmatched word, UW; phonetically matched pseudoword, MP; and phonetically unmatched pseudoword, UP).

Target sentence	Incorrect word response	Error type
Her study of history was persistently pursued.	*Or studies*… persistently pursued.	MW
But it was a hopeful *sign*, he *told* himself.	But it was a hopeful *time*, he *called* himself.	MW
No one will even suspect that it is your work.	*Tell them we* suspect that is your work.	UW
In a way, he couldn't blame her.	*There's no* way he couldn't blame her.	UW
Instead of that he was engulfed by bedlam.	*Insap*… that.	MP
What elements of our behavior are *decisive*?	Behavior… *side-sive*?	MP
But the problems cling to pools, as any pool owner *knows*.	Any pool owner *nevvins*.	UP
But that *explanation* is only partly true.	That *lup-nik* is……true.	UP

sentence materials, thereby avoiding overgeneralization of the role of speech cues at the word level to overall speech understanding by EHI listeners. A potential implication of our findings is that algorithms to enhance subsegmental cues focusing on transitions in hearing assistive devices for EHI listeners may be less beneficial than expected in clinical practice if evaluation materials exclude sentences.

C. Perceptual strategies in processing subsegmentally interrupted sentences

Two additional analyses were conducted, which detailed differences between the performance strategies of YNH and EHI listeners. Clearly, high-frequency hearing loss of EHI listeners causes a quantitative reduction in intelligibility of interrupted sentences. However, EHI listeners may also have used qualitatively different strategies to integrate information in interrupted sentences that might be related to age-related cognitive deficits. To determine if qualitatively different strategies were used, first the error distributions of incorrect words were compared. Second, the rankings of easier to harder sentences among the 42 test sentences across groups were examined for qualitative differences.

For the analysis of the distributions for incorrect words, we categorized incorrectly identified words into one of four error types. This analysis was based on a native American-English linguist's phonetic transcriptions of incorrect responses. The four types of errors for incorrect words were phonetically matched words (MWs), phonetically unmatched words (UWs), phonetically matched pseudowords (MPs), and phonetically unmatched pseudowords (UPs). The MW errors occurred when listeners' incorrect responses were apparently activated from meaningful words that sound similar to test words. For UW errors, listeners incorrectly responded with meaningful words; however their responses were phonetically dissimilar with the test words. MP and UP errors occurred when listeners' responses were meaningless pseudowords, as shown in Table II. If pseudoword errors were phonetically similar to the target, they were "matched" i.e., MP responses, while phonetically dissimilar responses, were "unmatched," i.e., UP. Table II displays examples of the incorrect word responses collected from listeners in the experiment corresponding to each of the four error types. Categorizing incorrect responses into meaningful words versus meaningless pseudowords was used to examine whether EHI listeners frequently guessed sounds they might have heard because the task was so hard for them. As shown in Table III, the order of the error distribution was similar across groups regardless of duration (i.e., MW > UW, MP > UP). Incorrect word responses for both listener groups occurred mostly from MWs, indicating that the strategies to recognize interrupted speech were very similar between YNH and EHI listeners as long as interrupted sentences were reasonably audible to listeners.

For the comparison of sentence rankings between groups, the 42 sentences were rank ordered based on correct words averaged, across conditions for each of the durations. Spearman rank correlation coefficients were computed within and across groups, as shown in Table IV. All sentence rankings were significantly and positively correlated between groups both across durations and within the same duration. Lower coefficients were found, as expected, when scores were either near floor for EHI-50 or near ceiling for YNH-70. However, a very high strength of correlation was found between 50% and 70% durations within each group ($r=0.80$, $p<0.01$ for YNH-50 and YNH-70; $r=0.94$, $p<0.01$ for EHI-50 and EHI-70). That is, the most understandable/difficult sentences with 50% duration were

TABLE III. Distribution (%) of four types of errors for the incorrect word responses pooled across conditions.

	Error type			
Group-duration (%)	MW (%)	UW (%)	MP (%)	UP (%)
YNH-50	71.2	25.5	2.8	0.5
EHI-50	67.3	31.6	0.9	0.2
YNH-70	75.0	19.8	4.8	0.4
EHI-70	70.2	27.1	2.0	0.7

TABLE IV. Spearman rank correlations (r) for the sentence ranking ($N=42$) between groups. The first column and row show the mean group performance [Mean=mean of words correct (%) across conditions and SE=standard errors]. Significant correlations are marked with **($p<0.01$).

Group-Duration (%)	YNH-50 (Mean=60%; SE=3.3)	YNH0–70 (Mean=85%; SE=2.1)
EHI-50 (Mean=16%; SE=4.8)	0.65**	0.43**
EHI-70 (Mean=46%; SE=8.2)	0.71**	0.54**

also the most understandable/difficult sentences with 70% duration for both YNH and EHI listeners. In addition, the correlations in Table IV reveal a strong relation (r=0.71) between sentence rankings for YNH-50 and EHI-70, indicating consistency of sentence rankings across two groups. Note that the significant but more moderate correlations in Table IV occurred where ceiling or floor performance was evident. Overall the Spearman correlations revealed that strategies for processing the interrupted sentences were quite similar across groups, regardless of hearing status.

To summarize, additional analyses revealed that YNH and EHI listener groups had similar error-type distributions across four error categories, as well as the positive relation between sentence rankings. This suggests that the negative effect of high-frequency hearing loss on the understanding of interrupted speech at a high presentation level resulted in quantitative, rather than qualitative, differences in performance between groups. These results further support that poorer performance of EHI listeners in processing interrupted sentences is unlikely to occur because of qualitatively different perceptual strategies that could be accompanied by age-related cognitive deficits but, rather, is largely caused by factors related to their peripheral hearing loss.

V. CONCLUSIONS

Following our previous study on sentence intelligibility with segmental interruption (Kewley-Port *et al.*, 2007), this study examined the ability of YNH and EHI listeners to comprehend interrupted sentences when different regions of subsegmental information (i.e., centers, margins, vowel onset and offset transition regions) preserved 50% or 70% of sentence duration. The major findings were as follows:

(i) Despite high intelligibility of sentences without interruption, EHI listeners had less successful auditory integration of interrupted speech signals than YNH listeners, regardless of the region and the duration of subsegmental cues glimpsed.

(ii) As expected, both groups performed better with a longer duration (70% versus 50% duration), resulting in improvement of 25% for YNH and of 30% for EHI listeners.

(iii) Different types of subsegmental information had similar effects on intelligibility of interrupted sentences for both YNH and EHI listeners. Thus, dynamic transition cues do not benefit YNH listeners more than quasi-steady-state cues do, at least when the task involves the use of these cues to identify sentences from partial information. Apparently EHI listeners are not substantially impaired in using dynamic transition information compared to steady-state information preserved in interrupted sentences.

(iv) Individual differences in word scores for EHI listeners were better predicted by their high-frequency hearing loss than by their age, despite the high presentation level of 95 dB SPL. This cautions us that a high presentation level does not ameliorate the negative effects of audibility in processing of interrupted sentences.

Additional analyses revealed similarity in error distributions of incorrect words and sentence rankings between groups. This indicates that reduced hearing at high frequencies may cause EHI listeners to have quantitatively worse performance than YNH but not qualitatively different perceptual strategies in processing interrupted sentences. Specifically, changes in perceptual strategies that might be attributed to age-related cognitive decline were not observed in EHI participants with our particularly challenging interrupted sentence task. Moreover, results from our two studies (current and KBL07) demonstrated that vowels contribute more to sentence intelligibility than did other cues for both YNH and EHI listener groups. Although reduced audibility of EHI listeners negatively affects their ability to integrate sentences interrupted with noise, preservation of vowel-only information compared to consonant-only or other subsegmental information has significant and substantial benefit for sentence understanding by EHI listeners. This motivates new ideas for the design of algorithms of speech processors for hearing aids, specifically to maximize the intelligibility of vowels. That is, algorithms for compensating hearing loss should preserve vowel information as much as possible in order to maximize possible resources that EHI listeners need when processing temporally interrupted speech information, a situation found in everyday listening environments.

ACKNOWLEDGMENTS

This research was supported by the National Institutes of Health Grant No. DC-02229 awarded to D.K.P. and the National Institute on Aging Grant No. AG022334 awarded to Dr. Larry E. Humes. The authors thank Brian W. Riordan for his assistance for the additional analysis of word scoring. They are also very appreciative of contributions made by Dr. Larry E. Humes to this project.

ANSI (**1969**). "American National Standard Methods for Calculation of the Articulation Index," ANSI S3.5-1969, American National Standards Institute, New York, NY.

ANSI (**1996**). "Specifications for audiometers," ANSI S3.6-1996, American National Standards Institute, New York, NY.

Bergman, M. (**1980**). "Effects of physical aspects of the message," in *Aging and the Perception of Speech* (University Park, Baltimore), pp. 69–78.

Bergman, M., Blumenfeld, V. G., Cascardo, D., Dash, B., Levitt, H., and Margulies, M. K. (**1976**). "Age-related decrement in hearing for speech: Sampling and longitudinal studies," J. Gerontol. **31**, 533–538.

Cole, R. A., Yan, Y. H., Mak, B., Fanty, M., and Bailey, T. (**1996**). "The contribution of consonants versus vowels to word recognition influent speech," in Proceedings of the ICASSP'96, pp. 853–856.

Dorman, M. F., Marton, K., Hannley, M. T., and Lindholm, J. M. (**1985**). "Phonetic identification by elderly normal and hearing-impaired listeners," J. Acoust. Soc. Am. **77**, 664–670.

Folstein, M. F., Folstein, S. E., and McHugh, P. R. (**1975**). "Mini-mental state: A practical method for grading the cognitive state of patients for the clinician," J. Psychiatr. Res. **12**, 189–198.

Fox, R. A., Wall, L. G., and Gokcen, J. (**1992**). "Age-related differences in processing dynamic information to identify vowel quality," J. Speech Hear. Res. **35**, 892–902.

Garofolo, J. S., Lamel, L. F., Fisher, W. M., Fiscus, J. G., Pallett, D. S., and Dahlgren, N. L. (**1990**). "DARPA TIMIT acoustic-phonetic continuous speech corpus CDROM," National Institute of Standards and Technology, NTIS Order No. PB91–505065.

Gordon-Salant, S. (**2005**). "Hearing loss and aging: New research findings and clinical implications," J. Rehabil. Res. Dev. Clin. Suppl. **42**, 9–24.

Gordon-Salant, S., and Fitzgibbons, P. J. (**1993**). "Temporal factors and speech recognition performance in young and elderly listeners," J. Speech

Hear. Res. **6**, 1276–1285.
Gordon-Salant, S., and Fitzgibbons, P. J. (**2004**). "Effects of stimulus and noise rate variability on speech perception by younger and older adults," J. Acoust. Soc. Am. **115**, 1808–1817.
Humes, L. E. (**2007**). "The contributions of audibility and cognitive factors to the benefit provided by amplified speech to older adults," J. Am. Acad. Audiol **18**, 590–603.
Iyer, N., Brungart, D. S., and Simpson, B. D. (**2007**). "Effects of periodic masker interruption on the intelligibility of interrupted speech," J. Acoust. Soc. Am. **122**, 1693–1701.
Jenkins, J. J., Strange, W., and Miranda, S. (**1994**). "Vowel identification in mixed-speaker silent-center syllables," J. Acoust. Soc. Am. **95**, 1030–1043.
Kewley-Port, D., Burkle, T. Z., and Lee, J. H. (**2007**). "Contribution of consonant versus vowel information to sentence intelligibility for young normal-hearing and elderly hearing-impaired listeners," J. Acoust. Soc. Am. **122**, 2365–2375.
Kirikae, I., Sato, T., and Shitara, T. (**1964**). "A study of hearing in advanced age," Laryngoscope **74**, 205–220.
Korsan-Bengtsen, M. (**1973**). "Distorted speech audiometry: A methodological and clinical study," Acta Oto-Laryngol., Suppl. **310**, 1–75.
Li, N., and Loizou, P. C. (**2007**). "Factors influencing glimpsing of speech in noise," J. Acoust. Soc. Am. **122**, 1165–1172.
Liberman, A. M., Harris, K. S., Hoffman, H. S., and Griffith, B. C. (**1957**). "The discrimination of speech sounds within and across phoneme boundaries," J. Exp. Psychol. **54**, 358–368.
Miller, G. A., and Licklider, J. C. R. (**1950**). "The intelligibility of interrupted speech," J. Acoust. Soc. Am. **22**, 167–173.
Nelson, P. B., and Jin, S. H. (**2004**). "Factors affecting speech understanding in gated interference: Cochlear implant users and normal-hearing listeners," J. Acoust. Soc. Am. **115**, 2286–2294.
Ohde, R. N., and Abou-Khalil, R. (**2001**). "Age differences for stop-consonant and vowel perception in adults," J. Acoust. Soc. Am. **110**, 2156–2166.
Peterson, G. E., and Barney, H. L. (**1952**). "Control methods used in a study of the vowels," J. Acoust. Soc. Am. **24**, 175–184.
Powers, G. L., and Speaks, C. (**1973**). "Intelligibility of temporally interrupted speech," J. Acoust. Soc. Am. **54**, 661–667.
Redford, M. A., and Diehl, R. L. (**1999**). "The relative perceptual distinctiveness of initial and final consonants in CVC syllables," J. Acoust. Soc. Am. **106**, 1555–1565.
Strange, W. (**1989**). "Dynamic specification of coarticulated vowels spoken in sentence context," J. Acoust. Soc. Am. **85**, 2135–2153.
Strange, W., and Bohn, O. S. (**1998**). "Dynamic specification of coarticulated German vowels: Perceptual and acoustical studies," J. Acoust. Soc. Am. **104**, 488–504.
Strange, W., Verbrugge, R. R., Shankweiler, D. P., and Edman, T. R. (**1976**). "Consonant environment specifies vowel identity," J. Acoust. Soc. Am. **60**, 213–224.
Studebaker, G. A. (**1985**). "A 'rationalized' arcsine transform," J. Speech Hear. Res. **28**, 455–462.
Warren, R. M. (**1970**). "Perceptual restoration of missing speech sounds," Science **23**, 392–393.
Wingfield, A., McCoy, S. L., Peelle, J. E., Tun, P. A., and Cox, L. C. (**2006**). "Effects of adult aging and hearing loss on comprehension of rapid speech varying in syntactic complexity," J. Am. Acad. Audiol **17**, 487–497.
Wingfield, A., Poon, L. W., Lombardi, L., and Lowe, D. (**1985**). "Speed of processing in normal aging: Effects of speech rate, linguistic structure, and processing time," J. Gerontol. **40**, 579–585.

Unsupervised joint prosody labeling and modeling for Mandarin speech

Chen-Yu Chiang[a] and Sin-Horng Chen[b]
Department of Communication Engineering, National Chiao Tung University, 1001 Ta-Hsueh Road, Hsinchu 300, Taiwan, Republic of China

Hsiu-Min Yu[c]
Language Center, Chung Hua University, 707, Sec. 2, Wu-Fu Road, Hsinchu 300, Taiwan, Republic of China

Yih-Ru Wang[d]
Department of Communication Engineering, National Chiao Tung University, 1001 Ta-Hsueh Road, Hsinchu 300, Taiwan, Republic of China

(Received 11 September 2007; revised 3 December 2008; accepted 3 December 2008)

An unsupervised joint prosody labeling and modeling method for Mandarin speech is proposed, a new scheme intended to construct statistical prosodic models and to label prosodic tags consistently for Mandarin speech. Two types of prosodic tags are determined by four prosodic models designed to illustrate the hierarchy of Mandarin prosody: the break of a syllable juncture to demarcate prosodic constituents and the prosodic state to represent any prosodic domain's pitch-level variation resulting from its upper-layered prosodic constituents' influences. The performance of the proposed method was evaluated using an unlabeled read-speech corpus articulated by an experienced female announcer. Experimental results showed that the estimated parameters of the four prosodic models were able to explore and describe the structures and patterns of Mandarin prosody. Besides, certain corresponding relationships between the break indices labeled and the associated words were found, and manifested the connections between prosodic and linguistic parameters, a finding further verifying the capability of the method presented. Finally, a quantitative comparison in labeling results between the proposed method and human labelers indicated that the former was more consistent and discriminative than the latter in prosodic feature distributions, a merit of the method developed here on the applications of prosody modeling.
 [DOI: 10.1121/1.3056559]

PACS number(s): 43.72.Ar [DOS] Pages: 1164–1183

I. INTRODUCTION

The term *prosody* refers to certain inherent suprasegmental properties that carry melodic, timing, and pragmatic information of continuous speech, encompassing accentuation, intonation, rhythm, speaking rate, prominences, pauses, and attitudes or emotions intended to express. Prosodic features are physically encoded in the variations in pitch contour, energy level, duration, and silence of spoken utterances. Prosodic studies have indicated that these prosodic features are not produced arbitrarily, but rather realized after a hierarchically organized structure which demarcates speech flows into domains of varying lengths by boundary or break cues such as pre- and postboundary lengthening, pitch and energy change, pauses, etc. Therefore, prosodic structure in English, for example, functions to set up syntagmatic contrasts to mark a prosodic word (PW), an intermediate phrase, or an intonational boundary.[1–3] On the other hand, the prosodic structure of Mandarin Chinese also parses continuous speech into different prosodic constituents by breaks that reflect different levels of Chinese linguistic processing: phonetic, lexical, syntactic, and pragmatic. As a result, successive words with related prosodic feature variations are aggregated to form prosodic phrases (PPhs), and contiguous PPhs are, in turn, integrated to form PPhs of a higher level. Consequently, deep exploration and an appropriate description of speech prosody are essential to the study of the speech processing of any language given. To provide a possible specification of prosodic features of utterances, a three-layer structure comprising PWs, intermediate phrases (or PPhs) and intonational phrases are commonly used, especially at the sentential level.[4–6] Some recent studies[7] proposed to integrate PPhs into PPh groups to interpret the contributions of higher-level discourse information to the wider-range and larger variations in the prosodic features of utterances of long texts. In the science of speech processing, to model prosody is to exploit a framework or a computational model to represent a hierarchy of PPhs of speech and to describe its relationship with the syntactic structure of the associated text.

In the past many prosody modeling methods have been proposed for various applications, including generation of prosodic information for text to speech (TTS),[8–10] segmenta-

[a]Electronic mail: gene.cm91g@nctu.edu.tw
[b]Electronic mail: schen@mail.nctu.edu.tw
[c]Electronic mail: kuo@chu.edu.tw
[d]Electronic mail: yrwang@cc.nctu.edu.tw

tion of untranscribed speech into sentences or topics,[11–13] generation of punctuations from speech,[14–16] detection of interrupt points in spontaneous speech,[11,17–19] automatic speech recognition (ASR),[20–26] and so forth. It can be found from those prosody modeling studies that four main issues have been intensively addressed. The first one is concerning representing a hierarchical PPh structure indirectly by tags marking important prosodic events. Among various prosodic events explored in the relevant literature,[27–32] break type and tone pattern are the most important ones: the break types of all word boundaries can determine the hierarchical PPh structure of an utterance, and the tonal patterns of all syllables/words can indicate the accented syllables/words of an utterance and may specify the pitch contour patterns of the prosodic constituents. Several prosody representation systems have been proposed in the past. They include tones and breaks indices (ToBI) (a standard prosody transcription system for American English utterances),[27] PROSPA,[29] INTSINT,[30] and TILT.[31] Among them, ToBI and its modifications to other languages, such as Pan-Mandarin ToBI (Ref. 32) and C-ToBI,[33] are most popular conventions for Mandarin Chinese prosodic tagging. The second main issue is about realizing the constituents of a hierarchical PPh structure by using prosodic feature patterns. This is mainly used in TTS for the generation of prosodic information from prosodic tags. A common approach is to use a multicomponent representation model to superimpose several prototypical contours of multilevel PPhs for each prosodic feature.[34–36] In Ref. 34, three components of sentence-specific contours, word-specific contours, and tone-specific contours are superimposed to form the synthesized contours of pitch and syllable duration for Mandarin TTS. The third main issue is related to exploring the relationship between prosodic tags (or boundary types) and the acoustic features surrounding the associated word juncture. Patterns of pause duration, pitch, and energy around word junctures are modeled for each prosodic tag or boundary type to help speech segmentation,[11–13] topic identification,[13] punctuation generation,[14–16] interrupt point detection,[11,17–19] and ASR (Refs. 20–26) based on word-based features. The last issue is upon modeling the relationship between prosodic structure and syntactic structure. It is known that prosodic structure is closely related to syntactic structure although they are not identical. Usually, only the relationship between a prosodic tag, such as break or prominence, and contextual linguistic features of syntactic structure is built. A good break-syntax model should be very useful in predicting breaks of various levels from input text for TTS. Main methods of building a break-syntax model for TTS are hierarchical stochastic model,[37,38] N-gram model,[39] classification and regressive tree (CART),[38,40–42] Markov model,[43] artificial neural networks,[44] maximum entropy model,[45–48] etc. In the popular Markov model-based approach, emission probabilities can be generated by CART (Ref. 42) or maximum entropy model.[48]

In all those studies, prosody modeling has been proved to be useful in above-mentioned applications, and the most commonly adopted approach by the previous studies is a supervised one to construct prosodic model from an annotated speech database with tags marking prosodic events being prelabeled manually. However, the supervised prosody modeling based on human labeling unavoidably arises such problems as diseconomy due to labeler training and manual labeling labor, and interlabelers' and intralabeler's inconsistency caused by individual subjectivity and fatigue during long time labeling, respectively. This inconsistency may mislead prosody modeling to obtain erroneous results, and hence lead to unwanted degradation of modeling performance. Even in the studies where prosody labeling can be automatically done by machine, their model is still trained with a manually annotated speech corpus,[26,28,49–54] so the performance of machine labeling is still subject to the quality of human prosody labeling.

To tackle the problems arising from the supervised prosody modeling with manual labeling, this work proposes a new unsupervised approach of prosody modeling to jointly perform prosody modeling and labeling for Mandarin speech based on an unlabeled speech database. The basic idea is to properly model data and then let the modeled data determine prosodic tags by themselves. The task is to automatically determine two types of prosodic tags for all utterances of a corpus and to build four prosodic models simultaneously. The two types of prosodic tags are (1) the break types of intersyllable locations (or syllable junctures) which can be used to demarcate the constituents of a hierarchy of Mandarin speech prosody and (2) the prosodic states of syllables which can be used to construct the pitch contour patterns of prosodic constituents. As will be discussed later, the prosodic state of a syllable is defined as a quantized and normalized pitch level affected by the current tone and the coarticulations from the two nearest neighboring syllables being properly eliminated. Since it mainly carries the information of PPhs, we therefore name it to refer to the state in a PPh. In Sec. IV, we will demonstrate its capability on realizing the pitch contour patterns of multilevel PPhs. It should be mentioned that in this study only pitch information is considered in the prosodic-state tag labeling. We will extend the study to consider the other two features of syllable duration and energy level in the future. The four prosodic models are introduced to describe the various relationships between the two types of prosodic tags and all available information sources including acoustic prosodic features and syntactic structure features. The first model, referred to as the syllable pitch contour model, describes the variations in syllable pitch contours controlled by several major affecting factors. The next one, referred to as the break-acoustics model, describes the relationship between the break type of a syllable juncture and nearby acoustic features. The third one describes the relationship between the break type of a syllable juncture and contextual linguistic features. It is referred to as the break-syntax model. Finally, the last model describes the relationship between the prosodic states of syllables and the break types of neighboring syllable junctures and is referred to as the prosodic-state model. A sequential optimization training algorithm is designed to iteratively estimate parameters of the four prosodic models and find all prosodic tags using an unlabeled speech corpus. Three advantages of the proposed method can be found. First, prosody modeling and labeling are accomplished jointly and automatically without using

human-labeled training corpus. Second, all information sources, including acoustic and linguistic features, are systematically used (via introducing the four prosodic models) in the prosody labeling. We therefore expect that the result of the prosodic labeling is more consistent than that done by human, which will in turn make the four prosodic models more accurate. Third, the four prosodic models constructed address all the four main issues of prosody modeling discussed above. So they are useful models and may be directly used or extended to be used in those applications mentioned above.

The remainder of this paper is organized as follows. Section II briefly describes the prosodic structure of Mandarin speech. Section III presents the proposed method. In Sec. IV experimental results are discussed, and in Sec. V some conclusions are drawn.

II. THE HIERARCHY OF MANDARIN SPEECH PROSODY

Much literature on Chinese prosody has shown that the prosody of Mandarin speech can be organized into hierarchical structures. A commonly agreed and used structure consists of four layers, including, from the lowest layer to the highest one, syllable layer, PW layer, PPh layer (or intermediate phrase), and intonation phrase.[38,41,42,44,45,48] As far as the major prosodic information relevant to each of the layers is concerned, given that Mandarin is a monosyllabic and tonal language, where each syllable with its inherent tone contains a lexical meaning, and each tone carries a lexically contrastive role, the features of every syllabic tone of an utterance are the most important prosodic information for the lowest layer; besides, tone along with syllable constituents affects syllable duration and energy level as well. As for the second prosodic layer, a PW refers to disyllabic and multisyllabic words or phrases composed of words syntactically and semantically closely related or most frequently collocated, so the words or phrases are uttered as a single unit as in hen "very"+*bu* 'not'+*zhuan-ye* "professional" (*not very professional*). As for the third prosodic layer, PPh is composed of one or several PWs and it usually ends with a perceptible but unobvious break. Finally, intonation phrase is at the top layer of the Mandarin prosodic structure. It determines the pitch contour of the intonation of a sentence containing one or several PPhs and it ends with an obvious break. Basically, the four-layer prosodic structure interprets the pitch and duration variations in syllable well for sentential utterances.

Recently, Tseng *et al.*[7] proposed to integrate contiguous PPhs into PPh groups to interpret the contributions of higher-level discourse information to the wider-range and larger variations in syllable pitch and duration of long utterances in paragraphs. Figure 1 displays the hierarchical prosodic phrase grouping (HPG) model of Mandarin speech proposed by Tseng *et al.* It is a five-layer structure. The first three layers in the hierarchy proposed by Tseng *et al.* are the same as those of the four-layer prosodic structure discussed above, which are referred to as syllable (SYL), PW, and PPh in the system of Tseng *et al.*, respectively. The fourth layer, breath group (BG), is formed by combining a sequence of PPhs,

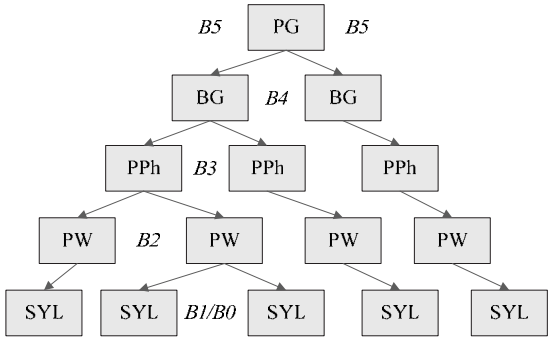


FIG. 1. A conceptual prosody hierarchy of Mandarin speech proposed by Tseng *et al.* in Ref. [7].

and a sequence of BGs, in turn, constitutes the fifth layer, prosodic phrase group (PG). The above five prosodic units are delimited by different types of the six breaks proposed by Tseng *et al.* First, $B0$ and $B1$ are defined for SYL boundaries within PW. Here, $B0$ represents reduced syllabic boundary and $B1$ represents normal syllabic boundary. Usually no identifiable pauses exist for both $B0$ and $B1$. Second, $B4$ and $B5$ are defined for BG and PG boundaries, respectively. $B4$ is a breathing pause and $B5$ is a complete speech paragraph end characterized by final lengthening coupled with weakening of speech sounds. Third, $B2$ and $B3$ are perceivable boundaries defined for PW and PPh boundaries, respectively.

In this study, we adopt the prosodic structure of Tseng *et al.* because our speech database also consists of long Mandarin utterances of paragraphs. However, we modify the break-type labeling scheme of HPG model by dividing $B2$ into two types, $B2$-1 and $B2$-2., and combining $B4$ and $B5$ into one denoted simply by $B4$. Here, $B2$-2 represents syllabic boundary of $B2$ perceived by pause, while $B2$-1 is $B2$ with $F0$ movement. The reason of dividing $B2$ into $B2$-1 and $B2$-2 is due to the difference of their acoustic cues to be modeled. On the contrary, the combination of $B4$ and $B5$ owes to the similarity of their acoustic characteristics. So, the break-type tags used is in $\Lambda=\{B0, B1, B2\text{-}1, B2\text{-}2, B3, B4\}$. These six break-type tags can be used to delimit four types of prosodic units: SYL, PW, PPh, and BG/PG. These four units are the constituents of our hierarchical prosodic structure.

To further specify the four-layer prosodic structure, a representation of its constituents using prosodic features is needed. Two main approaches of representation can be considered. One is direct representation approach to represent each individual prosodic constituent by multiple prototypical patterns for each prosodic feature of syllable pitch contour, duration, or energy level.[7–9,34–36] The other is indirect representation approach[55,56] by using some tags which carry the information of prosodic constituents and are treated as hidden. Due to the following two reasons, we do not adopt direct representation approach in the prosody modeling and labeling study. First, the technique of direct representation approach is still not mature enough to produce a good direct representation for the hierarchy of Mandarin speech prosody. The modeling errors, defined as the ratio of root mean square errors of direct representations to the standard deviations of the raw data, are still as high as about 30% for the multilayer representations of syllable duration, and energy using the

HPG model.[7] Second, a good direct representation is not easy to be realized for the case of joint prosody modeling and labeling using an unlabeled speech corpus in which the prosodic structures of all utterances are not well determined in advance. Degeneration may occur because break labeling errors may produce inaccurate representation patterns of prosodic constituents, which in turn may cause more break labeling errors to occur. Instead, we adopt an indirect representation approach to employ a new prosodic tag to represent the aggregative contributions of the constituents of the upper three layers on syllable pitch level. This tag is defined as a quantized and normalized syllable pitch level with the affections from the current tone and the two nearest neighboring tones being properly eliminated. So it carries mainly the pitch-level information of the upper three layers of the prosodic structure, i.e., PW, PPh, and BG/PG. We call it *prosodic state* to roughly mean the state in the pitch contour of a PPh (PW, PPh, or BG/PG). Two advantages of using the prosodic-state tag can be found. First, the tag is defined for each individual syllable so that the affection of a labeling error is limited to the current syllable only. No degeneration in the joint prosody modeling and labeling process will occur. Second, the tag carries the full information of pitch-level variation in the upper three layers of the prosodic structure. In Sec. IV, we will show the capability of the prosodic-state tag on constructing the pitch contour patterns of PW, PPh, and BG/PG. It is worthy to note that prosodic states of syllable duration and energy level can be similarly defined and added to the joint prosody labeling and modeling study. But for simplicity we only consider the prosodic state of syllable pitch level in this study.

III. THE PROPOSED METHOD

The proposed method first treats the problem as a model-based prosody labeling problem to define the four prosodic models to describe various relationships between the prosodic tags to be labeled and the available information sources of acoustic and syntactic features. It then extends the formulation for the joint prosody labeling and modeling problem and applies a sequential optimization procedure to jointly label prosodic tags and estimate the model parameters using an unlabeled speech corpus. We discuss these two parts in detail as follows.

A. The design of the four prosodic models

The prosody labeling problem can be generally formulated as a parametric optimization problem to find the best prosodic tag sequence $\mathbf{T}^*$ given with the acoustic feature sequence $\mathbf{A}$ of the input speech utterance and the linguistic feature sequence $\mathbf{L}$ of the associated text:

$$\mathbf{T}^* = \arg\max_{\mathbf{T}} P(\mathbf{T}|\mathbf{A},\mathbf{L}) = \arg\max_{\mathbf{T}} P(\mathbf{T},\mathbf{A}|\mathbf{L}). \quad (1)$$

Two types of prosodic tags which carry the information of prosodic structure of Mandarin speech are considered in this study. One is the break type of syllable juncture. A set of six break types, defined in Sec. II, is used. It is denoted as $\{B0, B1, B2\text{-}1, B2\text{-}2, B3, B4\}$. These six break types are used to define a hierarchy of speech prosody comprising four constituents of SYL, PW, PPh, and BG/PG. Another is the prosodic state of syllable defined as a quantized and normalized syllable pitch level with the affections of the current tone and the two nearest neighboring tones being properly eliminated. As discussed in Sec. II, it is an indirect representation of the prosodic constituents to carry the pitch-level information of PW, PPh, and BG/PG. So, $\mathbf{T}$ can be refined to comprise a break-type sequence $\mathbf{B}$ and a prosodic-state sequence $\mathbf{p}$.

Two types of acoustic features can be considered. One is the prosodic features which carry the information of prosodic constituents. Acoustic features of this type are assumed to be closely related to the prosodic-state tags and loosely related to or independent of the break-type tags. Primary features of this type include syllable pitch contour, syllable duration, and syllable energy level. For simplicity we only consider syllable pitch contour in this study and will extend the study to include the other two in the future. Another is the acoustic features used to specify the break type of syllable juncture. Acoustic features of this type are assumed to be closely related to the break-type tags and loosely related to or independent of the prosodic-state tags. Primary features of this type include pause duration and energy-dip level of syllable juncture, energy, and pitch jumps across syllable juncture, lengthening factor of syllable duration, etc. Among them, pitch jump has been implicitly considered via the use of prosodic-state tag, energy jump is somewhat a redundant feature as energy-dip level is used, and lengthening factor will be considered together with the syllable duration modeling in the future. We therefore only consider the two features of pause duration and energy-dip level in this study. From above discussions, $\mathbf{A}$ can be refined to comprise a syllable pitch contour sequence $\mathbf{sp}$, a pause duration sequence $\mathbf{pd}$, and an energy-dip level sequence $\mathbf{ed}$.

The linguistic features used span a wide range from syllable level, such as syllable tone and initial type; word level, such as syllable juncture type (intraword and interword), word length, part of speech (POS), and type of punctuation mark (PM); to syntactic tree level, such as size of syntactic phrase and syntactic juncture type (intraphrase and interphrase). Since syllable tone is an important linguistic feature and mainly used in the modeling of syllable pitch contour, we separate it from other linguistic features. So, $\mathbf{L}$ is refined to include a syllable tone sequence $\mathbf{t}$ and a reduced linguistic feature set $\mathbf{l}$.

Based on above discussions, we rewrite $P(\mathbf{T},\mathbf{A}|\mathbf{L})$ by

$$\begin{aligned} P(\mathbf{T},\mathbf{A}|\mathbf{L}) &= P(\mathbf{B},\mathbf{p},\mathbf{sp},\mathbf{pd},\mathbf{ed}|\mathbf{l},\mathbf{t}) \\ &= P(\mathbf{sp},\mathbf{pd},\mathbf{ed}|\mathbf{B},\mathbf{p},\mathbf{l},\mathbf{t})P(\mathbf{B},\mathbf{p}|\mathbf{l},\mathbf{t}), \end{aligned} \quad (2)$$

where $P(\mathbf{sp},\mathbf{pd},\mathbf{ed}|\mathbf{B},\mathbf{p},\mathbf{l},\mathbf{t})$ is a general prosodic feature model describing the variations in acoustic prosodic features $(\mathbf{sp},\mathbf{pd},\mathbf{ed})$ controlled by the prosodic tags $(\mathbf{B},\mathbf{p})$ representing the prosodic structure and the linguistic features $(\mathbf{l},\mathbf{t})$ representing the syntactic structure, and $P(\mathbf{B},\mathbf{p}|\mathbf{l},\mathbf{t})$ is a general prosody-syntax model which describes the relationship between $(\mathbf{B},\mathbf{p})$ and $(\mathbf{l},\mathbf{t})$.

Since the break-type tag sequence, $\mathbf{B}$, has already carried the prosodic cues related to syllable junctures, we there-

fore assume that the observed syllable-based acoustic feature, **sp**, and the juncture-based acoustic features, (**pd**,**ed**), are independent as **B** is given. So we split $P(\mathbf{sp},\mathbf{pd},\mathbf{ed}|\mathbf{B},\mathbf{p},\mathbf{l},\mathbf{t})$ into two terms:

$$P(\mathbf{sp},\mathbf{pd},\mathbf{ed}|\mathbf{B},\mathbf{p},\mathbf{l},\mathbf{t}) \approx P(\mathbf{sp}|\mathbf{B},\mathbf{p},\mathbf{l},\mathbf{t})P(\mathbf{pd},\mathbf{ed}|\mathbf{B},\mathbf{p},\mathbf{l},\mathbf{t}). \tag{3}$$

Here $P(\mathbf{sp}|\mathbf{B},\mathbf{p},\mathbf{l},\mathbf{t})$ is a syllable pitch contour model describing the variation in syllable pitch contour controlled by $(\mathbf{B},\mathbf{p},\mathbf{l},\mathbf{t})$ and $P(\mathbf{pd},\mathbf{ed}|\mathbf{B},\mathbf{p},\mathbf{l},\mathbf{t})$ is a break-acoustics model describing the acoustic cues of syllable junctures for different break types. In this study, the syllable pitch contour model is realized using a modified version of the syllable pitch contour model proposed previously.[55] It models the pitch contour of each syllable separately and considers four main affecting factors, including the current prosodic state p_n, the current tone t_n, and the coarticulations from the two nearest neighboring tones, t_{n-1} and t_{n+1}, conditioned, respectively, on the break types, B_{n-1} and B_n, of the syllable junctures on both sides. Specifically, the model is expressed by

$$P(\mathbf{sp}|\mathbf{B},\mathbf{p},\mathbf{l},\mathbf{t}) \approx P(\mathbf{sp}|\mathbf{B},\mathbf{p},\mathbf{t}) \approx \prod_{n=1}^{N} P(\mathbf{sp}_n|p_n,B_{n-1}^{n},t_{n-1}^{n+1}), \tag{4}$$

where

$$\mathbf{sp}_n = \mathbf{sp}_n^r + \boldsymbol{\beta}_{t_n} + \boldsymbol{\beta}_{p_n} + \boldsymbol{\beta}^f_{B_{n-1},tp_{n-1}} + \boldsymbol{\beta}^b_{B_n,tp_n} + \boldsymbol{\mu} \quad \text{for } 1 \le n \le N \tag{5}$$

is the observed pitch contour of nth syllable (referred to as *syllable n* hereafter) represented by the first four orthogonally transformed parameters of syllable log $F0$ contour;[57] $B_{n-1}^{n}=(B_{n-1},B_n)$, $t_{n-1}^{n+1}=(t_{n-1},t_n,t_{n+1})$, $\mathbf{sp}_n^r$ is the normalized (or residual) version of $\mathbf{sp}_n$, and $\boldsymbol{\beta}_x$ represents the affecting pattern (AP) of affecting factor x. Here AP means the effect of a factor on increase or decrease in the observed syllable pitch contour vector $\mathbf{sp}_n$. $\boldsymbol{\beta}_{t_n}$ and $\boldsymbol{\beta}_{p_n}$ are the APs of affecting factors t_n and p_n, respectively; tp_n is the tone pair $t_n^{n+1}=(t_n,t_{n+1})$; $\boldsymbol{\beta}^f_{B_{n-1},tp_{n-1}}$ and $\boldsymbol{\beta}^b_{B_n,tp_n}$ are the APs of forward and backward coarticulations contributed from *syllable* $n-1$ and *syllable* $n+1$, respectively; and $\boldsymbol{\mu}$ is the AP of global mean. For taking care of utterance boundaries, two special break types, B_b and B_e, are assigned to the two ending locations of all utterances, i.e., $B_0=B_b$ and $B_N=B_e$, and two special APs of coarticulation, $\boldsymbol{\beta}^f_{B_b,t_1}=\boldsymbol{\beta}^f_{B_0,tp_0}$ and $\boldsymbol{\beta}^b_{B_e,t_N}=\boldsymbol{\beta}^b_{B_N,tp_N}$, are accordingly adopted to represent the effects of utterance onset and offset, respectively. In this study, $\boldsymbol{\beta}_{p_n}$ is set to have nonzero value only in its first dimension in order to restrict the influence of prosodic state merely on the log $F0$ level of the current syllable. By assuming that $\mathbf{sp}_n^r$ is zero mean and normally distributed, i.e., $N(\mathbf{sp}_n^r;\mathbf{0},\mathbf{R})$, we have

$$P(\mathbf{sp}_n|p_n,B_{n-1}^{n},t_{n-1}^{n+1}) = N(\mathbf{sp}_n;\boldsymbol{\beta}_{t_n} + \boldsymbol{\beta}_{p_n} + \boldsymbol{\beta}^f_{B_{n-1},tp_{n-1}} + \boldsymbol{\beta}^b_{B_n,tp_n} + \boldsymbol{\mu},\mathbf{R}) \quad \text{for } 1 \le n \le N. \tag{6}$$

It is noted that the affection from **l** is assumed to be implicitly included in the affection of **p** and hence is neglected. We also note that the coarticulation effect is elegantly treated to consider different degrees of coupling between two neighboring syllables via letting it depend on the break type of the syllable juncture.

The break-acoustics model $P(\mathbf{pd},\mathbf{ed}|\mathbf{B},\mathbf{p},\mathbf{l},\mathbf{t})$ is further elaborated via assuming that $(\mathbf{pd},\mathbf{ed})$ is independent of $(\mathbf{p},\mathbf{t})$ which mainly carries information of prosodic constituents rather than that of syllable juncture. So we have

$$P(\mathbf{pd},\mathbf{ed}|\mathbf{B},\mathbf{p},\mathbf{l},\mathbf{t}) \approx P(\mathbf{pd},\mathbf{ed}|\mathbf{B},\mathbf{l}) \approx \prod_{n=1}^{N-1} P(\mathrm{pd}_n,\mathrm{ed}_n|B_n,\mathbf{l}_n), \tag{7}$$

where pd_n and ed_n are the pause duration and energy-dip level of the juncture following syllable n (referred to as *juncture n* hereafter) and $\mathbf{l}_n$ is the contextual linguistic feature vector around juncture n. For mathematically tractable, $P(\mathrm{pd}_n,\mathrm{ed}_n|B_n,\mathbf{l}_n)$ is further simplified and realized by the product of a gamma distribution for pause duration and a normal distribution for energy-dip level:

$$P(\mathrm{pd}_n,\mathrm{ed}_n|B_n,\mathbf{l}_n) = g(\mathrm{pd}_n;\alpha_{B_n,\mathbf{l}_n},\beta_{B_n,\mathbf{l}_n})N(\mathrm{ed}_n;\mu_{B_n,\mathbf{l}_n},\sigma^2_{B_n,\mathbf{l}_n}). \tag{8}$$

In this study, $g(\mathrm{pd}_n;\alpha_{B_n,\mathbf{l}_n},\beta_{B_n,\mathbf{l}_n})$ and $N(\mathrm{ed}_n;\mu_{B_n,\mathbf{l}_n},\sigma^2_{B_n,\mathbf{l}_n})$ are concurrently generated by the decision tree method,[58] for each break type.

Similarly, we simplify the general prosody-syntax model $P(\mathbf{B},\mathbf{p}|\mathbf{l},\mathbf{t})$ via assuming the independency of $(\mathbf{B},\mathbf{p})$ and **t**, and decomposing it into two models, i.e.,

$$P(\mathbf{B},\mathbf{p}|\mathbf{l},\mathbf{t}) \approx P(\mathbf{B},\mathbf{p}|\mathbf{l}) = P(\mathbf{p}|\mathbf{B},\mathbf{l})P(\mathbf{B}|\mathbf{l}) \approx P(\mathbf{p}|\mathbf{B})P(\mathbf{B}|\mathbf{l}), \tag{9}$$

where $P(\mathbf{p}|\mathbf{B})$ is a prosodic-state model describing the dynamics of **p** given with **B** and $P(B_n|\mathbf{l}_n)$ is a break-syntax model describing the relationship between **B** and the contextual linguistic feature sequence **l**. In this study, we realize $P(\mathbf{p}|\mathbf{B})$ by a Markov model:

$$P(\mathbf{p}|\mathbf{B}) \approx P(p_1)\left[\prod_{n=2}^{N} P(p_n|p_{n-1},B_{n-1})\right], \tag{10}$$

where $P(p_1)$ is the initial prosodic-state probability for syllable 1 and $P(p_n|p_{n-1},B_{n-1})$ is the prosodic-state transition probability from syllable $n-1$ to syllable n given B_{n-1}. We also simplify $P(\mathbf{B}|\mathbf{l})$ by separately modeling it for each syllable juncture:

$$P(\mathbf{B}|\mathbf{l}) = \prod_{n=1}^{N-1} P(B_n|\mathbf{l}_n). \tag{11}$$

Here $P(B_n|\mathbf{l}_n)$ is implemented by the decision tree method.[58]

B. Joint prosody labeling and modeling

A sequential optimization procedure based on the maximum likelihood (ML) criterion is proposed to jointly label the prosodic tags for all utterances of the training corpus and to estimate the parameters of the four prosodic models. It is

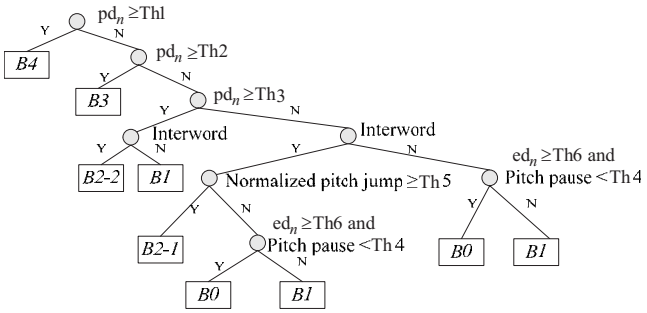


FIG. 2. The decision tree for initial break-type labeling.

divided into two main parts: *initialization* and *iteration*. The initialization part determines initial prosodic tags of all utterances and estimates initial parameters of the four prosodic models by a specially designed procedure. The iteration part first defines an objective likelihood function for each utterance by

$$Q = \left(\prod_{n=1}^{N} P(\mathbf{sp}_n | p_n, B_{n-1}^{n}, t_{n-1}^{n+1}) \right) \times \left(P(p_1) \prod_{n=2}^{N} P(p_n | p_{n-1}, B_{n-1}) \right) \times \left(\prod_{n=1}^{N-1} (P(\mathrm{pd}_n, \mathrm{ed}_n | B_n, \mathbf{l}_n) P(B_n | \mathbf{l}_n)) \right). \quad (12)$$

It then applies a multistep iterative procedure to update the labels of prosodic tags and the parameters of the four prosodic models sequentially and iteratively. In Secs. III B 1 and III B 2, we discuss the sequential optimization procedure in detail.

1. Initialization

The initialization part is further divided into two subparts: (a) a specially designed procedure to determine initial break labels of all syllable junctures and (b) a ML estimation process to estimate initial parameters of the four prosodic models and to determine the initial prosodic-state labels of all syllables using the information of initial break labels determined in the first subpart.

a. Initial labeling of break indices The initial break index of each syllable juncture is determined by a decision tree (see Fig. 2) designed based on a prior knowledge about break labeling/modeling gained in previous studies.[7,26,38,49–55,59,60] It is known that pause duration is the most important acoustic cue to specify breaks. Most word junctures with PM have long pauses so that they are most likely labeled as major break, or in our case *B*3 and *B*4. On the other hand, most intraword syllable junctures have very short pause duration so that they are generally labeled as nonbreak, or in our case *B*0 and *B*1. Moreover, *B*0 represents tightly coupled syllable juncture so that it is distinguished from *B*1 by having very short pitch pause duration and high energy-dip level. In-between these extreme situations, non-PM interword junctures with medium pause duration and with medium pitch jump are likely labeled as *B*2-2 and *B*2-1, respectively. By using the prior knowledge, we develop the algorithms to determine all thresholds of the decision tree (Th1–Th6) in a systematic way to avoid doing it manually or by trial and error. Detail of the algorithm is given in the Appendix.

b. Estimation of the initial parameters of the four prosodic models and prosodic-state indices The initializations of the break-acoustics model and the break-syntax model can be done independently with initial break indices of all syllable junctures being given. We realize them by the CART algorithm.[58] For the initialization of the break-acoustics model, the CART algorithm with the node splitting criterion of maximum likelihood gain is adopted to classify pause duration pd_n and energy-dip level ed_n for each break type B according to a question set Θ_1 derived from the contextual linguistic features $\mathbf{l}_n$. Each leave node represents the product of a gamma distribution $g(\mathrm{pd}_n; \alpha_{B,\mathbf{l}_n}, \beta_{B,\mathbf{l}_n})$ and a normal distribution $N(\mathrm{ed}_n; \mu_{B,\mathbf{l}_n}, \sigma^2_{B,\mathbf{l}_n})$. For the initialization of the break-syntax model $P(B_n|\mathbf{l}_n)$, a decision tree is built by using another question set Θ_2 derived also from $\mathbf{l}_n$ to classify break types.

The initializations of the syllable pitch contour model and prosodic-state indices are integrated together and performed by a progressive estimation procedure. Since the syllable pitch contour model is a multiparametric representation model to superimpose several APs of major affecting factors to form the surface syllable pitch contour, the estimation of an AP may be interfered by the existence of the APs of other types. It is therefore improper to estimate all initial parameters independently. We hence adopt a progressive estimation strategy to first determine the initial APs which can be estimated most reliably and then eliminate their affections from the surface pitch contours for the estimations of the remaining APs. In this study, the order of initial AP estimation is listed as follows: global mean $\boldsymbol{\mu}$, five tones $\boldsymbol{\beta}_t$, coarticulation $\{\boldsymbol{\beta}^f_{B,tp}, \boldsymbol{\beta}^b_{B,tp}, \boldsymbol{\beta}^f_{B_b,t_1}$, and $\boldsymbol{\beta}^b_{B_e,t_N}\}$, and prosodic states $\boldsymbol{\beta}_p$. Notice that the initial prosodic-state indices are assigned by vector quantization (VQ) of the pitch-level components of the residue pitch contours, and the APs are set to be the codewords obtained by VQ. Lastly, the initialization of the prosodic-state model $P(\mathbf{p}|\mathbf{B})$ is done using the labeled prosodic-state indices and break indices.

2. Iteration

The iteration is a multistep iterative procedure listed below.

Step 1. Update the APs of five tones $\boldsymbol{\beta}_t$ with all other APs being fixed.

Step 2. Update the APs of coarticulation $\{\boldsymbol{\beta}^f_{B,tp}, \boldsymbol{\beta}^b_{B,tp}, \boldsymbol{\beta}^f_{B_b,t_1}$, and $\boldsymbol{\beta}^b_{B_e,t_N}\}$ with all other APs being fixed, and then update $\mathbf{R}$.

Step 3. Relabel the prosodic-state sequence of each utterance by using the Viterbi algorithm so as to maximize Q defined in Eq. (12). Then, update the APs of prosodic-state $\boldsymbol{\beta}_p$, the prosodic-state model $P(\mathbf{p}|\mathbf{B})$, and $\mathbf{R}$.

Step 4. Relabel the break-type sequence of each utterance by using the Viterbi algorithm so as to maximize Q. Then, update the prosodic-state model $P(\mathbf{p}|\mathbf{B})$ and $\mathbf{R}$.

Step 5. Reconstruct the decision trees to update $P(\mathrm{pd}_n, \mathrm{ed}_n|B_n, \mathbf{l}_n)$ and $P(B_n|\mathbf{l}_n)$ by the CART algorithm using the question sets Θ_1 and Θ_2, respectively.

Step 6. Repeat Steps 1–5 until a convergence is reached.

IV. EXPERIMENTAL RESULTS

The proposed method was evaluated using an unlabeled Mandarin speech database. The database contained read speech of a female professional announcer. Its texts were all short paragraphs composed of several sentences selected from the Sinica Treebank corpus.[61] The database consisted of 380 utterances which contained in total 52 192 syllables. In this experiment, the number of prosodic states was properly set to be 16 because the root mean squared error (RMSE) of VQ saturated when the number of prosodic states was greater than 16. The sequential optimization procedure took 69 iterations to reach a convergence. Following is the presentation of the analyses, discussion, and findings of our experiment, which is arranged in the order that an examination and interpretation of the parameters of the four prosodic models was introduced in Secs. IV A–IV D, then, to evaluate the performance of the models proposed, explorations in the relationships between prosodic breaks and linguistic features of texts, the length of prosodic constituents, and the general pitch patterns of prosodic constituents obtained in our method were described in Secs. IV E–IV G, and, finally, to further verify the labeling outcomes generated by our models, a comparison conducted between human labeling and our labeling was given in Secs. IV H and IV I.

A. The syllable pitch contour model

We first examined the parameters of the syllable pitch contour model $P(\mathbf{sp}_n|p_n,B_{n-1}^n,t_{n-1}^{n+1})$. The covariance matrices of the original and normalized syllable log $F0$ contour feature vectors are shown below:

$$\mathbf{R}_{\mathbf{sp}} = \begin{bmatrix} 883.7 & 23.9 & -25.6 & -0.5 \\ 23.9 & 90.5 & 9.7 & -8.2 \\ -25.6 & 9.7 & 17.8 & -0.9 \\ -0.5 & -8.2 & -0.9 & 5.0 \end{bmatrix} \times 10^{-4} \Rightarrow \mathbf{R}_{\mathbf{sp}^r}$$

$$= \begin{bmatrix} 3.5 & 0.2 & -0.2 & 0.0 \\ 0.2 & 31.9 & 2.6 & -1.5 \\ -0.2 & 2.6 & 11.1 & 0.6 \\ 0.0 & -1.5 & 0.6 & 3.7 \end{bmatrix} \times 10^{-4}.$$

Obviously, all elements of $\mathbf{R}_{\mathbf{sp}^r}$ were much smaller than those of $\mathbf{R}_{\mathbf{sp}}$. This showed that the influences of the affecting factors considered were indeed essential to the variation in **sp**.

Figure 3 displays the APs of five tones. We find from the figure that the APs of the first four tones conformed well to the standard tone patterns found by Chao.[62] As for tone 5, its low dipping pattern resembles the pattern of tone 3 to some degree. This also matched the finding in the previous study about tone 5.[63]

Table I displays the APs (log $F0$ levels) and the distribution of the 16 prosodic states. It can be seen from Table I that these log $F0$ levels spanned widely to cover the whole dynamic range of log $F0$ variation with lower indices of prosodic state corresponding to lower log $F0$ levels, and the prosodic states distributed normally with relatively few located at the two extremes of high and low prosodic states.

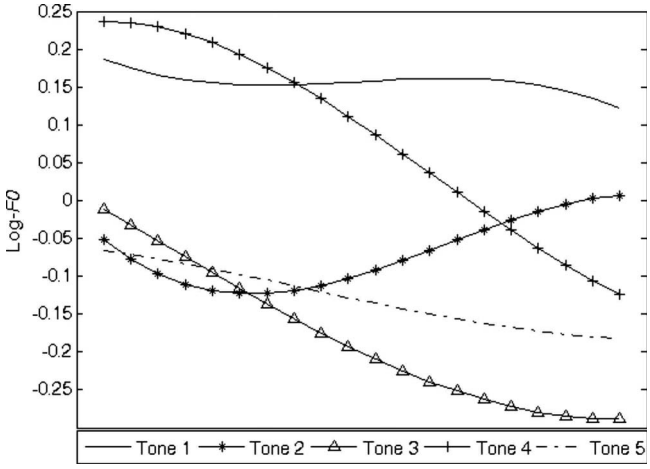


FIG. 3. The APs of five tones.

Figures 4(a) and 4(c) display the APs of forward and backward coarticulations, $\boldsymbol{\beta}_{B,tp}^f$ and $\boldsymbol{\beta}_{B,tp}^b$, for the three break types of $B0$, $B1$, and $B4$. These three break types were chosen on purpose to show extreme cases of intersyllable coarticulation: $B0$ for tightly coupling, $B1$ for normal coupling, and $B4$ for no coupling. Some interesting phenomena can be observed from the figure. First, it can be seen from Fig. 4(a) that most APs of forward coarticulation for $B0$ and $B1$, $\boldsymbol{\beta}_{B0,tp}^f$ and $\boldsymbol{\beta}_{B1,tp}^f$, were bended in their beginning parts. These bendings were to compensate the level mismatch between the beginning and ending parts of the log $F0$ contours of the tone pairs for highly coarticulated preceding and current syllables, so as to make their log $F0$ contours be concatenated more smoothly. For example, the upward bending at the beginning parts of $\{\boldsymbol{\beta}_{B,tp}^f|tp=(1,2),(1,3),(2,2),(2,3),(1,5)\}$ were due to H-L mismatches, while the downward bending at the beginning parts of $[\boldsymbol{\beta}_{B,tp}^f|tp=(3,1),(3,4),(5,1),(5,4),(4,1)(4,4)]$ corresponded to L-H mismatches. Similarly, it can be observed from Fig. 4(c) that the ending parts of the APs of backward coarticulation for $B0$ and $B1$, $\boldsymbol{\beta}_{B0,tp}^b$ and $\boldsymbol{\beta}_{B1,tp}^b$, were bended. But the degrees of their upward and downward bendings were generally smaller. This conformed to the observation reported in Ref. 64 that the carry-over effect on the syllable $F0$ contour influenced by the preceding syllable is much larger than the anticipation effect caused by the following syllable. Second, it can be found from Figs. 4(a) and 4(c) that most APs of forward and backward coarticulations for $B4$ with the same current tone looked similar and hence were nearly independent of their respective preceding and succeeding tones. This showed that the intersyllable coarticulation across a $B4$ break was relatively low as compared with those of $B0$ and $B1$. Moreover, many APs of forward and backward coarticulations for $B4$ were downward bended in their beginning and ending parts, respectively. They exhibited the onset and offset phenomena at the beginning and ending syllables of BG/PG. Furthermore, we find from Figs. 4(b) and 4(d) that most utterance initial and final patterns, $\boldsymbol{\beta}_{B_b,t}^f$ and $\boldsymbol{\beta}_{B_e,t}^b$, looked very similar to those of $\boldsymbol{\beta}_{B4,tp}^f$ and $\boldsymbol{\beta}_{B4,tp}^b$, respectively, to show the same onset and offset phenomena at the two types of utterance boundaries. We also find that $\boldsymbol{\beta}_{B_e,3}^b$ and $\boldsymbol{\beta}_{B_e,5}^b$ were two exceptional patterns which

TABLE I. The APs [log $F0$ levels, $\boldsymbol{\beta}_p(1)$] and the distribution [$P(p)$] of the 16 prosodic states.

State index p	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
$\boldsymbol{\beta}_p(1)$	−0.77	−0.50	−0.37	−0.28	−0.22	−0.16	−0.10	−0.05	0.01	0.06	0.12	0.17	0.24	0.31	0.38	0.49
$P(p)$	0.00	0.01	0.02	0.04	0.07	0.10	0.11	0.12	0.12	0.10	0.09	0.08	0.06	0.05	0.03	0.01

had lower levels. These probably resulted from the total relaxation of pronunciation at the utterance ending for these two tones. Third, it can be found from Fig. 4(c) that the APs of $\boldsymbol{\beta}^b_{B0,(3,3)}$ and $\boldsymbol{\beta}^b_{B1,(3,3)}$ were upward bended drastically in their ending parts. As combining with the AP of tone 3 shown in Fig. 3, these bendings would make the integrated log $F0$ patterns of the first syllable in a (3,3) tone pair change from middle-falling tone-3 shape to middle-rising tone-2 shape to fulfill the well-known 3-3 tone *sandhi* rule which says that the first tone 3 of a 3-3 tone pair will change to a tone 2. On the contrary, we find that the pattern $\boldsymbol{\beta}^b_{B4,(3,3)}$ did not bend upward. This showed that the 3-3 tone sandhi rule did not apply when the syllable juncture was a $B4$. Lastly, we made some comments to the APs of forward and backward coarticulations for $B2$-1, $B2$-2, and $B3$. Basically, the APs of $B2$-1 and $B3$ resembled to those of $B4$ but with smaller upward and downward bendings, and $B2$-2 had similar patterns to those of $B1$ but with smaller upward and downward bendings.

From above analyses, we find that the inferred syllable pitch contour model provides a meaningful interpretation to the variation in syllable pitch contour controlled by several major affecting factors. With this capability, the model can be used in Mandarin TTS to generate pitch contour if all tags of prosodic-state and break type can be properly predicted from the input text. It can also be used in Mandarin ASR to manipulate pitch information for tone discrimination.

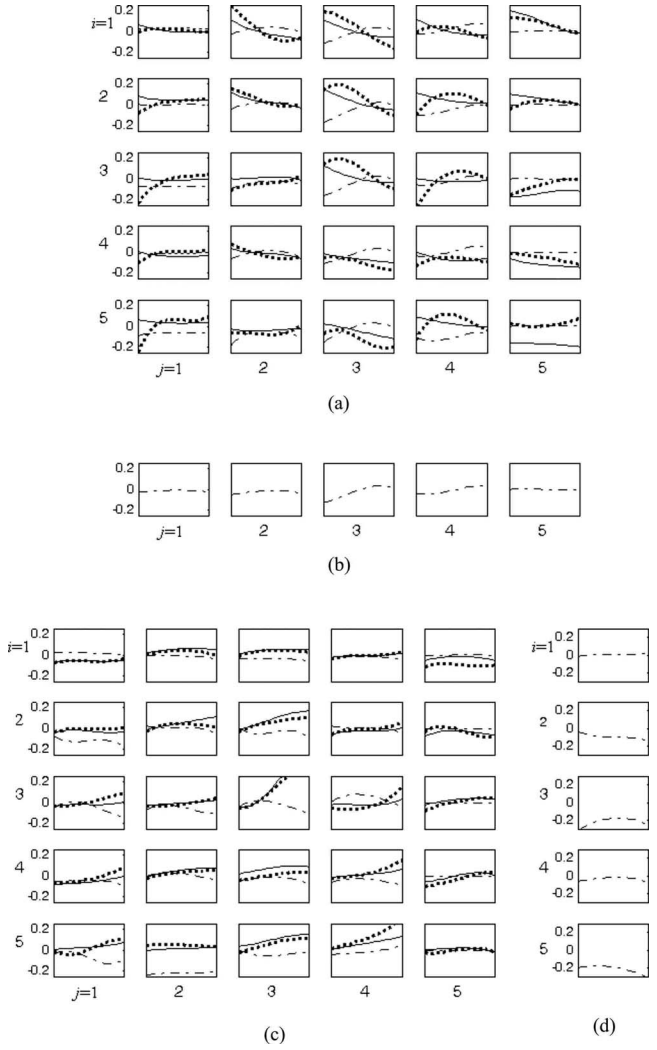


FIG. 4. The APs of (a) forward and (c) backward coarticulations, $\boldsymbol{\beta}^f_{B,ip}$ and $\boldsymbol{\beta}^b_{B,ip}$, for $B0$ (point line), $B1$ (solid line), and $B4$ (dashed line); and the APs of (b) utterance onset and (d) utterance offset, $\boldsymbol{\beta}^f_{B_b,i}$ and $\boldsymbol{\beta}^b_{B_e,i}$, for B_b and B_e. Here $tp=(i,j)$ and $t=j$ or i.

B. The break-acoustics model

The two break-acoustics models, $g(\mathrm{pd}_n;\alpha_{B_n,\mathbf{l}_n},\beta_{B_n,\mathbf{l}_n})$ and $N(\mathrm{ed}_n;\mu_{B_n,\mathbf{l}_n},\sigma^2_{B_n,\mathbf{l}_n})$, were built by the decision tree method using the question set Θ_1. One decision tree was constructed for each break type. Figure 5 displays the distributions of pause duration and energy-dip level for the root nodes of these six break types. It can be found from the

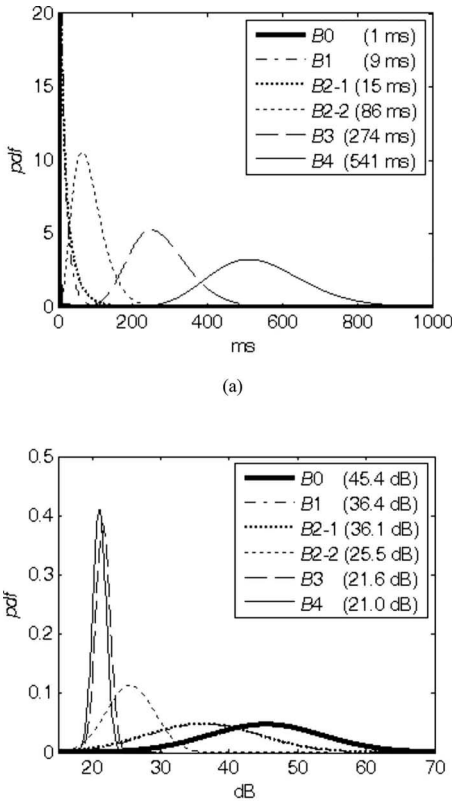


FIG. 5. The pdfs of (a) pause duration and (b) energy-dip level for the root nodes of these six break types. Numbers in ( ) denote the mean values.

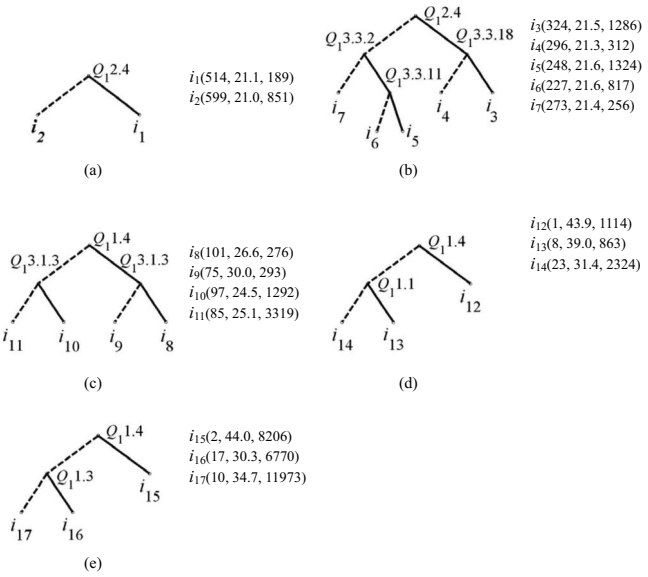


FIG. 6. The decision trees of the break-acoustics model for (a) *B*4, (b) *B*3, (c) *B*2-2, (d) *B*2-1, and (e) *B*1. The numbers in a bracket denote average pause duration in milliseconds (left), energy-dip level in decibels (middle), and sample count (right) of the associated node. Solid line indicates positive answer to the question and dashed line indicates negative answer.

figure that the break types of higher level were generally associated with longer pause duration and lower energy-dip level. *B*0 had very short pause duration and widespread energy-dip level with very high mean value. *B*1 and *B*2-1 had similar distributions of short pause durations and widespread high energy-dip level. *B*2-2 had medium long pause duration and medium high energy-dip level. Both *B*3 and *B*4 had widespread long pause duration and low energy-dip level. These conformed to the prior knowledge about break types.[4–7]

To further examine the model, we show its decision trees for the five break types of *B*4, *B*3, *B*2-2, *B*2-1 and *B*1 in Fig. 6. It is noted here that no tree split for *B*0 due to the relative uniformity on the acoustic prosodic features of its samples. Generally, the questions used to split trees of higher-level break types (*B*4 and *B*3) tended to be related to higher-level syntactic features, such as PM ($Q_1$2.4) and syntactic phrase size ($Q_1$3.1.3, $Q_1$3.3.2, $Q_1$3.3.11, and $Q_1$3.3.18). On the contrary, the questions of lower-level phonetic features ($Q_1$1.1, $Q_1$1.3, and $Q_1$1.4) tended to split trees of lower-level break types (*B*1 and *B*2-1).

From above discussions, we find that the inferred break-acoustics model describes the relationship of the break type of syllable juncture with the two intersyllable acoustic features and some contextual linguistic features very well. So it seems that the model can be used to predict major and minor breaks from acoustic and linguistic cues for some applications, such as segmenting speech into sentences and generation of punctuations from speech.

C. The prosodic-state model

We then examined the prosodic-state model. Figure 7 displays some most significant transitions of $P(p_n|p_{n-1},B_{n-1})$ for six break types. For *B*0 and *B*1, the general high-to-low, nearby-state transitions showed that the syllable log *F*0 level

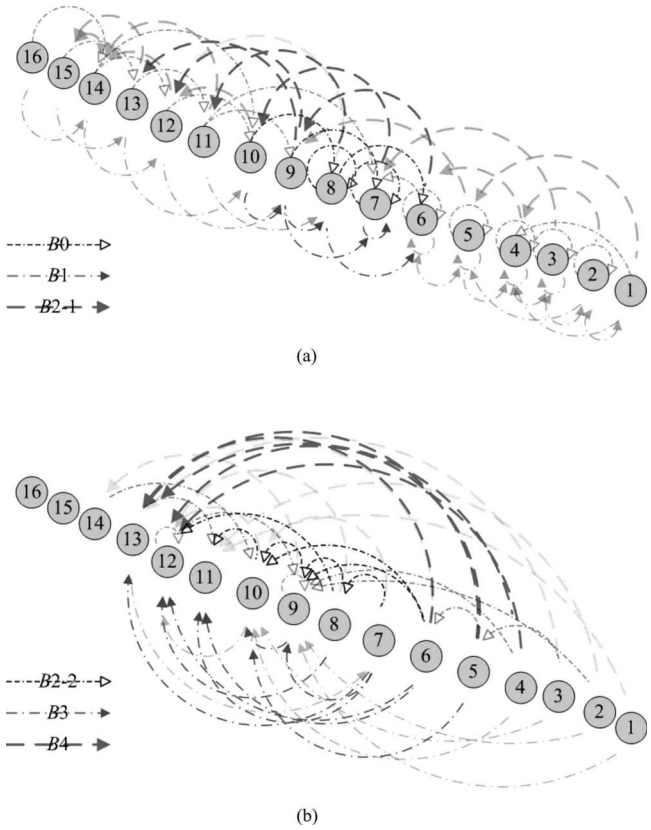


FIG. 7. The most significant prosodic-state transitions for (a) *B*0, *B*1, and *B*2-1, and (b) *B*2-2, *B*3, and *B*4. Here, the number in each node represents the index of the prosodic state. Note that bold and thin lines denote the primary and secondary state transitions, respectively.

declined slowly within PWs. We also find that some low-to-high, nearby-state transitions occurred within PWs of low pitch level. This demonstrated the sustaining phenomenon of the log *F*0 trajectory at the ending part of some PPhs. For *B*2-2, it had both high-to-low and low-to-high state transitions. For *B*2-1, *B*3, and *B*4, their low-to-high state transitions showed clearly the phenomena of syllable log *F*0 level resets across PWs, PPhs, and BG/PGs. Compared with these clear log *F*0 level resets, the resets of *B*2-2 were insignificant. Combining the results shown in Figs. 5 and 7, we find that *B*2-1 and *B*2-2 had different acoustic characteristics: *B*2-1 had significant log *F*0 reset with very short pause duration, while *B*2-2 had longer pause duration with low or no log *F*0 reset.

From above findings, since the prosodic states defined in our study mainly carry the full information of pitch-level variation in the upper three layers of prosodic structure (PW, PPh, or BG/PG), the prosodic-state model can roughly represent dynamic patterns of PW, PPh, and BG/PG and may be applied to pitch contour generation in Mandarin TTS.

D. The break-syntax model

The break-syntax model $P(B_n|\mathbf{l}_n)$ was built by the decision tree method using the question set Θ_2. Figure 8 displays the decision tree of the break-syntax model. The tree was divided into four subtrees, *T*3–*T*6, by the three questions of $Q_2$2.1.1 (PM?), $Q_2$2.1.3 (minor PM?), and $Q_2$1.3 (intra-

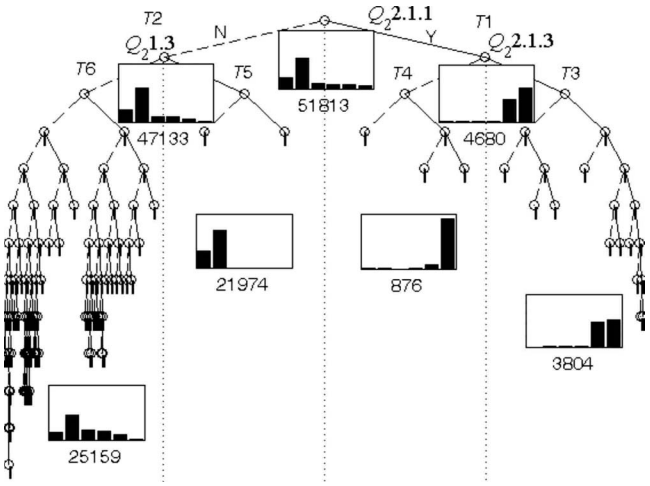


FIG. 8. The decision tree of the break-syntax model. The bar plot associated with a node denotes the distributions of these six break types (*B*0, *B*1, *B*2-1, *B*2-2, *B*3, and *B*4, from left to right) and the number is the total sample count of the node.

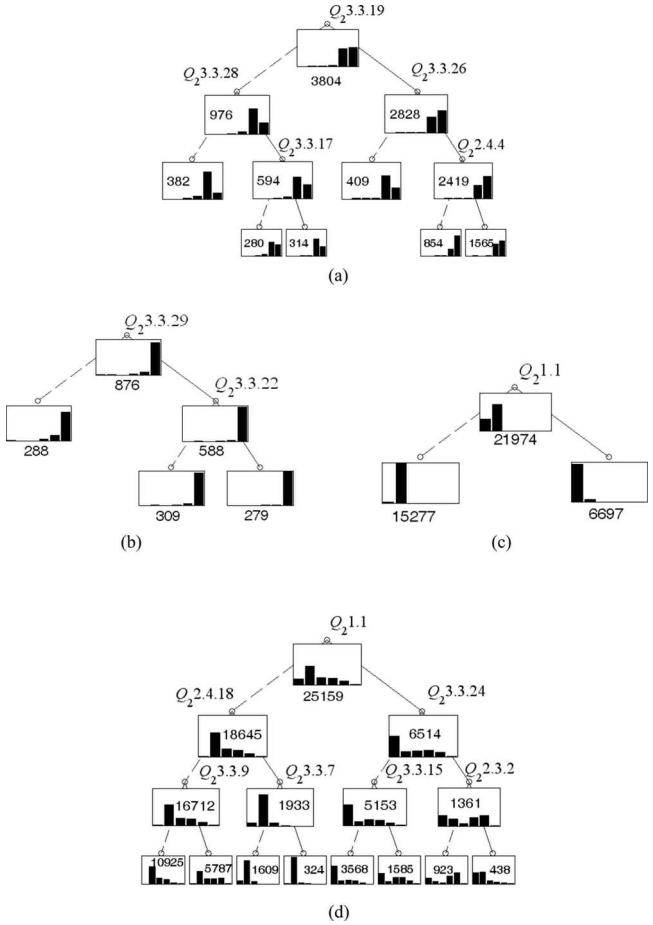


FIG. 9. The more detailed structures of subtrees of (a) *T*3, (b) *T*4, (c) *T*5, and (d) *T*6. Solid line indicates positive answer to the question and dashed line indicates negative answer.

word?). It can be seen from the figure that the root node of subtree *T*3, which corresponded to syllable juncture with minor PM, was mainly composed of *B*3 and *B*4. Similarly, the root nodes of subtrees *T*4 and *T*5, corresponding to major PM and intraword syllable juncture, were mainly composed of *B*4 and *B*0/*B*1, respectively. Due to the fact that the break-type constituents of both *T*4 and *T*5 were pure, they had very simple tree structures. On the contrary, subtree *T*6 was a miscellaneous collection of all other types of syllable juncture without PM. So, it had the most complex tree structure.

Figure 9 displays the more detailed structures of these four subtrees up to the fourth layer. From Figs. 9(a) and 9(b), we find that nodes in *T*3 and *T*4 were mainly split by questions related to high-level linguistic features such as $Q_2 3.3.19$ (Is the length of the following syntactic phrase/sentence greater than 6?) and $Q_2 3.3.29$ (Is the length of the preceding syntactic phrase/sentence greater than 7?). As shown in Fig. 9(c), *T*5 had two leaf nodes split by $Q_2 1.1$ (Does the following syllable have a null initial or initial in $\{m,n,l,r\}$?). The set associated with positive answer was mainly composed of *B*0, while another set was mainly composed of *B*1. As shown in Fig. 9(d), *T*6 was constructed by questions related to features of various levels, including $Q_2 1.1$, $Q_2 2.4.18$ (Is the preceding word "DE"?), $Q_2 2.3.2$ (Is the preceding word a function word?), $Q_2 3.3.24$ (Is the length of the preceding syntactic phrase greater than 2?), and so on. We also find from Fig. 9 that the purities of the break-type constituents were high for leaf nodes of *T*4 and *T*5, medium high for nodes of *T*3, and relatively low for most nodes of *T*6. This implies that it is difficult to correctly label (or predict) the break types of syllable junctures other than intraword and those with major PM by the break-syntax model using only linguistic features without the help of acoustic cues.

E. Analyses of the labeled break types

Since the purpose of announcers' broadcasting is to propagate information accurately to the audience relying exclusively on their audio perception, our well-trained informant skillfully manipulated as many segmental and prosodic cues as possible, such as clear and precise articulation, strategic variations in the fundamental frequency, volume, syllable length, and types of breaks. These prosodic information carried in the utterance speech, in turn, reflects the infor-

TABLE II. Statistics of break types labeled for 121 prefixes and 195 suffixes.

Labeled break type		*B*0	*B*1	*B*2-1	*B*2-2	*B*3	*B*4	Total count
Prefix	Preboundary	94	1289	460	545	193	5	2586
	Postboundary	584	1475	344	178	5	0	2586
Suffix	Preboundary	1046	2466	31	20	3	0	3566
	Postboundary	307	1479	272	482	568	458	3566

TABLE III. Statistics of break types labeled for the DE words.

Labeled break type	*B*0	*B*1	*B*2-1	*B*2-2	*B*3	*B*4	Total count
Preboundary	168	1600	146	1	0	0	1915
Postboundary	210	1035	331	294	41	4	1915

mant's mental grammar, her Mandarin linguistic competence that determines when to form a semantically appropriate word chunk, a PPh, or a larger unit, and hence where and how long a break in an utterance should be so that the informant's speech would sound natural, informative, and attention attracting to the audience.

As a research based on our informant's speech data rich in the Mandarin prosodic cues, our break-type-labeling model also can generate appropriate break types consistent with native speakers' psychological reality. To verify this point, we examined the relationship between some special groups of words/morphemes and their concurring break types that both our break-type-labeling model and the ordinary Mandarin native speakers would consistently produce. These special groups of words/morphemes include (1) affix morpheme; (2) DE; (3) Ng, Di, and T; (4) VE; (5) Caa and Cb; and (6) P.[65] The results are discussed in more detail as follows.

1. Set of affix morpheme

It is well known that prefixes and suffixes are bound morphemes that attach to their preceding or following heads to form units of complex words. Since the resultant form after combining the head and the affix is a unit, it is reasonable to predict that the breaks at the boundaries between the head and the affix tend to fall in *B*0 or *B*1 types. These phenomena were observed in our corpus. We found that some Mandarin Chinese monosyllabic prefixes, such as *bu-* "un-, dis-, in-," *ke-* "-able," *wu-* "un-, -less, without," etc.,[62,66] tend to join the following heads to form legitimate words as in *bu-li* "unfavorable," *bu-fang-bian* "inconvenient," *ke-wu* "detestable," *ke-sing* "feasible," *wu-sian* "limitless," and *wu-shuang* "unparalleled." Similarly, by attaching monosyllabic suffixes, such as *-bian* "side," *-zhe* "-er, -or," *-hua* "-ize," etc., to the preceding roots, we can derive complex words as in *lu-bian* "roadside, curb," *he-bian* "riverside," *zuo-zhe* "author, writer," *sing-zhe* "religious practicer," *gung-yie-hua* "industrialize," and *min-zhu-hua* "democratize."

Table II lists the statistics of the break types labeled for the syllable boundaries of 121 prefixes and 195 suffixes. It can be seen from the table that 79.6% of the postsyllable boundaries of these 121 prefixes and 98.5% of the presyllable boundaries of these 195 suffixes were labeled as *B*0 or *B*1. These prosodic findings reflect the fact that morphologically the combination of head and affix generates a lexical unit, and thus the break between them is determined to be the break type of intra-PW category by our method. The results were also consistent with some rules found in Refs. 59, 60, and 63.

2. Word set of DE

The words in the DE set particularly refer to *de*, *zhe*, and *di*, which serve multifunctions including a possessive marker, an adjective marker, and an adverbial marker.[65] They are characterized by the fact that a DE word can combine with a wide range of preceding syntactic constituents to form a possessive adjective as in a noun phrase (NP)-de structure: *xue-sheng-de quan-li* "students' right" to derive an adjective phrase as in a verb phrase (VP)-de structure: *se-siang-zhe qin* "nostalgia" or to function as an adverbial phrase as in a DM-de structure: *ke-ren yi-bo-bo-di yong-jin-dien-lai* "guest were flocking to the shop." Despite the variety of the preceding constituent, a DE word, similar to a suffix, builds closer connection with its preceding constituent to form a larger syntactic unit; consequently, it is predictable that the break at the DE words' preboundary position tends to fall into *B*0 and *B*1, which means a pause is hardly to be perceived at this juncture. It is also reasonable to infer that due to a looser connection between the DE words and the following constituent, less *B*0 and *B*1 would occur at the postboundary position.

The statistics in Table III indicates that the distribution of the break types labeled by our model just conformed with our anticipation; while 92.3% preboundary breaks of the DE words were *B*0 and *B*1, only 65% postsyllable boundaries of the DE words fell into the same types, which suggests that for the DE words, the majority of the neighboring breaks are unperceivable, and in most cases only at the postboundary

TABLE IV. Statistics of break types labeled for the word sets of Ng, Di, and T.

Labeled break type		*B*0	*B*1	*B*2-1	*B*2-2	*B*3	*B*4	Total count
Ng	Preboundary	97	420	19	12	0	2	550
	Postboundary	26	81	17	58	245	123	550
Di	Preboundary	107	83	12	1	0	2	205
	Postboundary	30	68	36	41	11	19	205
T	Preboundary	89	84	14	11	0	0	198
	Postboundary	0	5	1	2	22	168	198

TABLE V. Statistics of break types labeled for word set of VE.

Labeled break type	*B*0	*B*1	*B*2-1	*B*2-2	*B*3	*B*4	Total count
Postboundary	63	177	99	108	159	234	840

position can perceivable breaks be sensed. This result also matched the findings in Refs. 59, 60, and 63.

3. Word sets of Ng, Di, and T

Ng, Di, and T represent the word sets of Mandarin Chinese localizers, aspectual adverbs, and particles,[65] respectively. The distinctive shared feature of these sets of words is that almost all the words are no longer than two syllables in length and that when combining with other syntactic constituent to form a larger phrase, they are all positioned at the end of the derived phrase, such as *san-tian ho*$_{\mathrm{Ng}}$ "three days latter," *kai-hui dang-zhung*$_{\mathrm{Di}}$ "while the meeting is being held," and *bu-qu le-ma*$_{\mathrm{T}}$ "not going?." Due to the characteristic of being postpositioned in a phrase, these words are inclined to be incorporated with their preceding constituents, and predictably barely any pauses can be perceived at the preboundary position. The statistic results listed in Table IV indicate that our model's break-labeling performance just exactly met our expectation. As high as 94%, 93%, and 87% of the presyllable boundaries of the words in this category were labeled as *B*0 or *B*1.

On the other hand, it is also interesting to find that for the breaks at the postsyllable boundaries, 67% and 96% of them were labeled as *B*3/*B*4 especially for the Ng-set and T-set words, respectively. Further investigation reveals that most of the longer breaks were caused by a following PM, an index representing the occurrence of a detectable pause. Besides, because the T-set words are phrasal or sentential final particles and hence are highly likely to be followed by a PM, a much higher ratio of *B*3/*B*4 could be found.

4. Word set of VE

VE represents a class of transitive verbs that take a sentence as the object, such as *ren-wei* "to suppose/think/believe (that)," *gan-dao* "to feel (that)," *biao-she* "to show/indicate/mean/suggest (that)," etc.[65] It is evident that since the message carried in a sentential object, compared to a NP object for example, demands longer time to process mentally before being accurately expressed, a longer pause is reasonably anticipated to occur after a VE verb for information operation. Based on the statistic results listed in Table V, on the whole 72% postword boundaries of the VE verbs were labeled as breaks with distinctly audible pauses, namely, *B*2-1, *B*2-2, *B*3, or even *B*4, another quite favorable evidence that the break types labeled by our model were consistent with the pause duration people usually take in their utterances.

However, it cannot be neglected that no less than 28% postword boundaries of the VE verbs were labeled as *B*0 or *B*1, implying that seemingly our model still generated quite a few unexpected break types for the VE verbs. Further observation of the data, nevertheless, found two main reasons to account for this discrepancy of labeling. First, besides a sentential object, part of the VE verbs could also take a NP object, so the breaks occurring before a NP object were predictably shorter than before a sentential object. The other reason for the occurrence of *B*0/*B*1 after a VE verb is that to express attitudinal, temporal, spatial, or manner information about a VE verb, a small word from the DE, Di, Ng, or T sets (such as *de*, *zhe*, *le*, *guo*, etc.) was attached to the verb, and this attachment and the close connection between the small word and the VE verb caused no need to pause at the juncture. However, the originally expected long pause (*B*3/*B*4) after the VE verb did not actually disappear; it was retained and only lagged behind to occur after the VE verb.

5. Word sets of Caa and Cb

Caa and Cb are two subcategories of Mandarin conjunctions, representing conjunctive conjunctions and correlative conjunctions,[65] respectively. In the case of Caa, the arguments linked by the Caa conjunctions are words or phrases of identical syntactic categories and are usually associated in their meaning as in *feng*$_{\mathrm{N}}$ he$_{\mathrm{Caa}}$ *yu*$_{\mathrm{N}}$ "wind and rain," *re*$_{\mathrm{VH}}$ *hia-shi*$_{\mathrm{Caa}}$ *leng*$_{\mathrm{VH}}$ "hot or cold," *si*$_{\mathrm{Neu}}$ *zhi*$_{\mathrm{Caa}}$ *shi*$_{\mathrm{Neu}}$ *sui*$_{\mathrm{Nf}}$ "from four to ten years old," and the like. Upon observation, we found that people usually tend to take a longer pause at preword boundary than at the postword context, forming a sensible rhythmic variation and hence facilitating message delivery. The statistics of the labeling results in Table VI informs us that 90% of the Caa preboundary breaks were not shorter than *B*2-2, while, on the contrary, 98% of the postword breaks were not longer than *B*2-2, a labeling outcome verifying our observation of the Caa words' neighboring breaks; that is, longer pauses tended to occur at the boundary between the preceding argument and the conjunction. The results matched some findings in Ref. 38.

On the other hand, the Cb conjunctions function to join

TABLE VI. Statistics of break types labeled for word sets of Caa and Cb.

Labeled break type		*B*0	*B*1	*B*2-1	*B*2-2	*B*3	*B*4	Total count
Caa	Preboundary	5	32	1	127	214	26	405
	Postboundary	52	104	157	85	7	0	405
Cb	Preboundary	61	46	23	39	168	512	849
	Postboundary	135	284	166	95	150	19	849

TABLE VII. Statistics of break types labeled for word sets of P07 and P21.

Labeled break type		*B*0	*B*1	*B*2-1	*B*2-2	*B*3	*B*4	Total count
P07	Preboundary	0	39	0	9	32	9	89
	Postboundary	8	38	34	9	0	0	89
P21	Preboundary	1	168	12	24	88	53	346
	Postboundary	27	79	208	28	4	0	346

two clauses—a syntactic unit much larger than Caa's arguments—into a compound sentence, and therefore have higher potential to be preceded or followed by a PM in written texts to delimit the domain of a clause or a sentence; in read speech the occurrence of a PM elicits the announcer to take a longer pause to index a message transition or a piece of new message is coming. Our statistic results show that in the case of Cb conjunctions 80% of the preword boundaries and 20% of the postword boundaries were labeled as *B*3/*B*4, which means much more PMs occurred before Cb conjunctions than afterward.

6. Word set of P

P represents the class of Chinese prepositions, which precede a required argument and together play several semantic roles and indicate various relationships such as time, location, tool, purpose, etc. Although Chinese Knowledge and information Processing (CKIP) categorizes prepositions into 65 types,[65] only 13 types are active in the Sinica Treebank corpus. As for the adjacent pause of a preposition, it is reasonable to expect that due to the close connection of a preposition and its following argument, the pause at the postword boundary tends to be short. For convenience of illustration, only *ba/jiang* (labeled as P07) and *zai* (labeled as P21), two typical and most frequently used prepositions, are selected out as the representative examples for discussion.

The statistic results in Table VII show that on the whole for both *ba/jiang* and *zai* about 90% of the postword boundaries were labeled as breaks no longer than *B*2-1 (a break type caused by a pitch jump instead of lengthened pause duration), which indicates that the pauses at this juncture were either unperceivable or tending to be very short, again another confirmation of our model's sound labeling job. Besides, a closer look at the distribution of break-type percentages reveals that as high as 49% and 69% of the postword breaks were *B*2-1 for *ba/jiang* and *zai*, respectively. This statistics reflected our informant's idiosyncratic style of articulating prepositional phrase; namely, besides leaving no pauses, she often made a pitch jump between a preposition and the following argument to cause a sensible short pause.

On the other hand, as far as the labeling at the preword boundary is concerned, most labels were either *B*1 or *B*3/*B*4; that is, 46% and 41% of the labels were *B*3/*B*4 and 44% and 49% of them were *B*1 for *ba/jiang* and *zai*, respectively, which suggests that our informant either took quite a long pause or just no pause at the preword position. To explain this phenomenon, further examination on the data containing these two prepositions revealed that the informant's long breaks (*B*3/*B*4) before a preposition were contributed by a left PM, and in the remained cases she usually took no pause at this position.

F. Analyses of prosodic constituents

Based on the break-type labeling, we can divide the syllable sequence of each utterance into three types of prosodic constituents (i.e., PW, PPh, and BG/PG) to form a four-layer prosodic structure. Statistics in Table VIII shows that the average lengths for these three types of prosodic constituents are, respectively, 3.17 syllables or 1.85 lexical words (LWs) for PWs; 6.98 syllables, 4.02 LWs, or 1.69 PWs for PPhs; and 16.69 syllables, 9.62 LWs, 4.07 PWs, or 1.94 PPhs for BG/PGs.

According to the histograms displayed in Fig. 10, the length of each of these three prosodic constituents spans, respectively, from 1 to 12 syllables for PWs, from 1 to 33 syllables for PPhs, and from 1 to 99 syllables for BG/PGs. Besides, the histograms also reveal that quite a few PPhs and BG/PGs, whose average lengths are supposed to be about 6.98 and 16.69 syllables, respectively, are nevertheless no longer than three syllables in length. Further investigation into these oddly short PPhs and BG/PGs indicates that the main reason lies in several special structure patterns of these constituents that require a long pause to highlight their prominence for successful information processing. First of all, in the case of short BG/PGs, defined as a sequence of syllables bounded by a *B*4 on both sides, many of the particularly short BG/PGs, actually consisted of a monosyllabic subject and VE verb, which, as discussed in Sec. IV E 4, due to its sentential object was tending to be followed by a long break up to *B*4; accordingly, bounded by a *B*4 on both sides, the structure pattern of a subject plus a VE verb, both monosyllabic in length, could generate as many short BG/PGs, as possible.

As for the cases of short PPhs, defined as a sequence of syllables delimited by (1) a *B*3 at both sides or (2) a *B*3 and a *B*4 at each side, respectively, most of the *B*3s or *B*4s bounding the very short PPhs were actually caused by the

TABLE VIII. Statistics of three types of prosodic constituents. Value in parentheses denotes standard deviation.

Average length in	Prosodic constituent		
	PW	PPh	BG/PG
Syllable	3.17(1.74)	6.98(3.48)	16.69(9.49)
Lexical word	1.85(1.03)	4.01(2.17)	9.62(5.43)
PW	1.00	1.69(1.55)	4.07(2.90)
PPh	X	1.00	1.94(1.75)

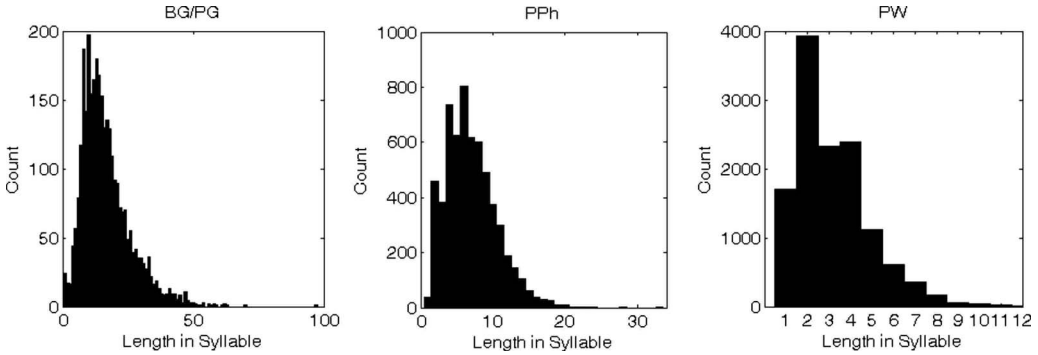


FIG. 10. Histograms of lengths for BG/PG, PPh, and PW.

existence of PMs that cued long pause duration. Table IX shows the statistic results of the short PPh instances with respect to the existence of PMs at their two endings. As shown in the table, 66% of one-syllable PPhs were bounded by PMs on both sides, and most of them were numbers that were used to enumerate events. On the other hand, in the case of two- or three-syllable PPhs, on the whole about 84% of them were delimited at least by a left-sided PM, which means that the majority of these PPhs occurred at the beginning of a sentence. In terms of the internal structure of the two-syllable PPhs, 91% of them were bisyllabic LWs functioned to express transitional relationships like contrast, comparison, reinforcement, or addition. As for the three-syllable PPhs, their structures were either a topicalized trisyllabic noun or any phrasal structure composed of two smaller syntactic elements as in a subject-VE structure (*wo ren-wui* "I suppose"), a preposition-noun structure (*you Chung-qing* "from Chung-qing"), a noun-localizer structure (*hun-zhan zhong* "in the scuffle"), etc., and the long pauses adjacent to these PPhs were, on the informant's part, strategies to cause prominent stress on these short phrases, and on the audience's part, offered the listeners longer time to process and catch the information with least distortion.

G. Pitch patterns of prosodic constituents

We then explored the log $F0$ patterns of the three prosodic constituents of PW, PPh, and BG/PGs,. First, we extracted the prosodic-state patterns from the observed pitch contour, $\mathbf{sp}_n$, by eliminating the influence of the current tone, the coarticulations from the two nearest neighboring tones, and the global mean, i.e.,

TABLE IX. Count of short PPh instances with respect to the existence of PM at their two endings.

	PPh length in syllable		
Count of PPh instances	1	2	3
No PMs on both sides	5	38	56
PM on right side only	1	8	28
PM on left side only	6	254	178
PMs on both sides	23	159	121
Total	35	459	383

$$\mathrm{pm}_n = \mathbf{sp}_n(1) - \boldsymbol{\beta}_{t_n}(1) - \boldsymbol{\beta}^f_{B_{n-1},tp_{n-1}}(1) - \boldsymbol{\beta}^b_{B_n,tp_n}(1) - \boldsymbol{\mu}(1) \text{ for } 1 \leqslant n \leqslant N, \quad (13)$$

where $\mathbf{x}(1)$ denotes the first dimension of vector $\mathbf{x}$. A sequence of pm_n delimited by $B2\text{-}1/B2\text{-}2/B3/B4$ at both sides is regarded as a prosodic-state pattern formed by integrating the log $F0$ mean patterns of the three prosodic constituents we considered. A model of prosodic-state pattern is therefore defined by

$$\mathrm{pm}_n = \mathrm{pm}^r_n + \beta_{\mathrm{PW}_n} + \beta_{\mathrm{PPh}_n} + \beta_{\mathrm{BG/PG}_n}, \quad (14)$$

where pm^r_n is the residual of log $F0$ mean at syllable n and β_{PW_n}, β_{PPh_n}, and $\beta_{\mathrm{BG/PG}_n}$ are the log $F0$ patterns of PW, PPh, and BG/PGs, with $\mathrm{PW}_n=(i,j)$, $\mathrm{PPh}_n=(i,j)$, and $\mathrm{BG/PG}_n=(i,j)$ denoting that syllable n is located at the jth place of an i-syllable PW, PPh, and BG/PGs, respectively. The model was trained by a sequential optimization procedure. After well training, the variances of $\mathbf{sp}_n(1)$, pm_n, and pm^r_n were 883.7×10^{-4}, 359.1×10^{-4}, and 191.2×10^{-4}, respectively. Hence, the total residual error (TRE), which is the percentage of sum-squared residue over the observed sum-squared

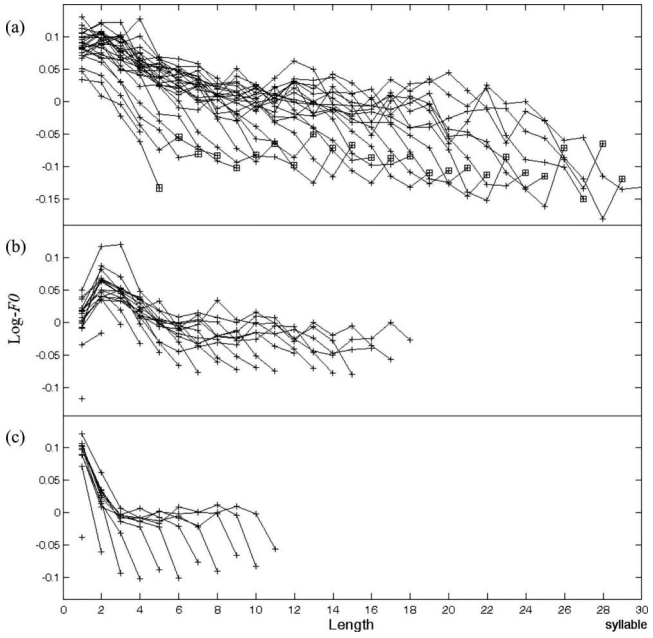


FIG. 11. The log $F0$ patterns of (a) BG/PG, (b) PPh, and (c) PW. The special symbol "□" in (a) indicates the ending syllable of a log $F0$ pattern.

TABLE X. Correlations between unsupervised and human-labeled breaks.

Human unsupervised	*b*1	*b*2	*b*3	*b*4	Total
*B*0	836	207	9	0	1052
*B*1	1970	726	70	0	2766
*B*2-1	81	313	53	1	448
*B*2-2	20	93	227	12	352
*B*3	0	0	137	260	397
*B*4	0	0	4	265	269
B_e	0	0	0	42	42
Total	2907	1339	500	580	5326

log *F*0 mean, is about 21.6% by the current representation.

Figure 11 displays the patterns of β_{PW_n}, β_{PPh_n}, and $\beta_{\mathrm{BG/PG}_n}$ with different lengths. It is noted that only the patterns calculated using more than 20 instances of prosodic-state patterns are displayed because we want to know their general log *F*0 patterns. It can be found from Fig. 11(a) that all $\beta_{\mathrm{BG/PG}}$ had declining patterns with dynamic range spanning approximately from −0.1 to 0.1. Moreover, most of them had short ending resets. From Fig. 11(b), we find that short β_{PPh} had rising-falling patterns, while long β_{PPh} had rising-falling-sustaining-falling patterns. Moreover, they had smaller dynamic range spanning approximately in [−0.07,007]. Lastly, we find from Fig. 11(c) that short β_{PW} showed high-falling patterns, while long β_{PW} showed falling-sustaining-falling patterns. Their dynamic range spanned approximately from −0.1 to 0.1.

From above analyses, we find that the prosodic-state tags possess rich information to represent the high-level prosodic constituents of the four-layer prosodic structure defined in this study. All these three types of log *F*0 patterns generally agree with the findings of previous studies on intonation patterns of Mandarin speech.[55,63,67,68] The superposition patterns $\beta_{\mathrm{PPh}}+\beta_{\mathrm{BG/PG}}$, and all these three patterns (β_{PW}, β_{PPh}, and $\beta_{\mathrm{BG/PG}}$), resembled the intonation patterns reported in the studies Tseng and co-workers[69–72] and the study of Chen *et al.*,[34] respectively. Furthermore, with this prosodically meaningful finding, these quantitative prosodic constituent patterns combining with the APs of tone and coarticulation (i.e., $\boldsymbol{\beta}_t$ and $\boldsymbol{\beta}^f_{B,tp}/\boldsymbol{\beta}^b_{B,tp}$) can be used in Mandarin TTS to generate pitch contour if all break type can be properly predicted from the input text. However, due to the fact that the errors of the current representation are still high, a further study to explore a more efficient representation is worthwhile doing in the future.

H. A comparison with human labeling

To further evaluate the performance of break labeling of the proposed method, a part of the Sinica Treebank corpus used in this study was labeled cooperatively by two experienced labelers working in the Phonetics Laboratory, Department of Foreign Languages and Literatures of National Chiao Tung University. The annotated dataset consisted of 42 utterances with 5326 syllables. The labeling system used was a ToBI-like one developed by the laboratory, which represents the Mandarin speech prosody by a four-layer structure containing syllable, PW, intermediate phrase, and intonation phrase. These four prosodic constituents are delimited by four break types of *b*1, *b*2, *b*3, and *b*4, respectively. Here *b*1 represents an implicit nonbreak index, *b*2 is a perceivable break index for PW boundary, *b*3 is a minor-break index, and *b*4 is a major-break index.

Table X displays the correlation matrix of the break indices labeled by the two methods. It can be found from Table X that 97.8% of human-labeled *b*4s, i.e., major breaks, were labeled as break indices of phrase or utterance boundaries (i.e., *B*3, *B*4, or B_e) in our method, and 96.5% of *b*1s, i.e., nonbreaks, were labeled as indices of SYL boundaries within PW (i.e., *B*0 or *B*1). This indicates that the two labeling methods were consistent for the two extreme cases of nonbreak and major break. It is also observed from the table that *b*3s mainly (73.6%) corresponded to break indices $\geq$*B*2-2, suggesting that the intermediate phrase boundaries in manual labeling, defined and perceived by the labelers as a minor break, were, to quite a certain extent, consistently judged as a clearly perceived short pause (*B*2-2) or medium pause (*B*3) in our labeling. However, in the cases of *b*2, 69.7% of them, defined as perceivable breaks, inconsistently corresponded to nonbreaks (*B*0 or *B*1) in our scheme. To account for such inconsistency, a statistics on the internal morphological and syntactic structures of the PWs delimited by *B*2 and *b*2 shows that (1) while as high as nearly 69.3% of PW-LW correspondence occurred in the human labeling, 40.0% of such correspondence was found in our method, and (2) while 41.2% of the PWs labeled by our method was cases of compound words or long phrases composed of at least four syllables, only 2.2% of the PWs in the similar types was judged by the labelers. This significant discrepancy in the demarcation of PWs between these two methods suggests that labelers, though trained to listen to the prosodic cues with visual aids of graphic user interface to label the breaks, tended to subjectively treat LWs as PWs or as pronunciation units rather than objectively and exclusively relied on the actual prosodic features in prosodic labeling. This inclination obviously resulted in shorter average lengths of prosodic constituents in human labeling. Figure 12 displays the histograms of length of the prosodic constituents formed by the two labeling methods. It can be found from the figure that the average lengths of PWs, PPhs, and BG/PGs labeled by our method were indeed longer than human-labeled PWs, intermediate phrases, and intonational phrases, respectively.

From the perspective of prosodic features, it can be

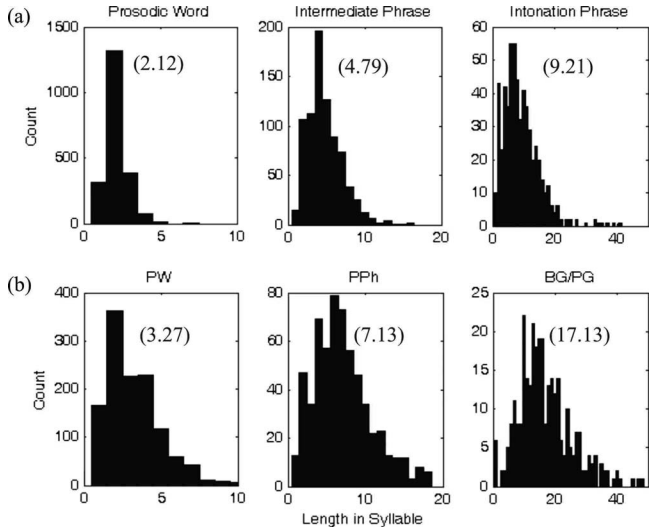


FIG. 12. The histograms of length of the prosodic constituents formed by (a) the human labelers and (b) the proposed methods. The numbers in ( ) represent the average length of prosodic constituents.

found from Figs. 13(a) and 13(b) that the similar histograms of pause duration and normalized pitch jump in the same rows represented labeling consistency in our method, while the distinguishable histograms in the same columns expressed labeling inconsistency in human labeling. Furthermore, Table XI displays symmetric Kullback–Leibler[73] (KL2) distances for the two break labeling methods to measure the difference between two acoustic feature distributions that belong to different break indices labeled by the same method. It can be found from Table XI that the KL2 distances for the proposed unsupervised method were generally greater than those of human labeling. Moreover, we find from Table XI(a) that the KL2 distances of pause duration were relatively large for all break index pairs of the proposed method except $(B1, B2\text{-}1)$; nevertheless the KL2 distances of normalized pitch jump for $(B1, B2\text{-}1)$ were large. On the contrary, we find from Table XI(b) that the KL2 distances of both acoustic features were low for $(b1, b2)$ of human labeling. This confirms that the six break types $B0-B4$ in our labeling have distinct characteristics of acoustic features but the break types in human labeling have less discriminated ones. Specifically, $B4$ has very large pause duration and significant pitch reset, $B3$ has large pause duration and pitch reset, $B2$-2 has medium pause duration, $B2$-1 and $B1$ have small pause duration but $B2$-1 has significant pitch reset and $B0$ has almost no pause duration. This property will be advantageous to our labeling method on those prosody modeling applications using acoustic features.

I. A labeling example

A typical example displaying the labeling results of the beginning part of a long utterance by the two methods is given in Fig. 14. We first examined the labeling results of our method. From Fig. 14(a), we find that the three PMs were labeled as two $B3$ and $B4$. One other $B3$ without PM appeared at the right boundary of a nine-syllable NP. Besides, there existed five $B2$-1 and four $B2$-2. They all appeared at interword junctures. We also find from Fig. 14(b) that all

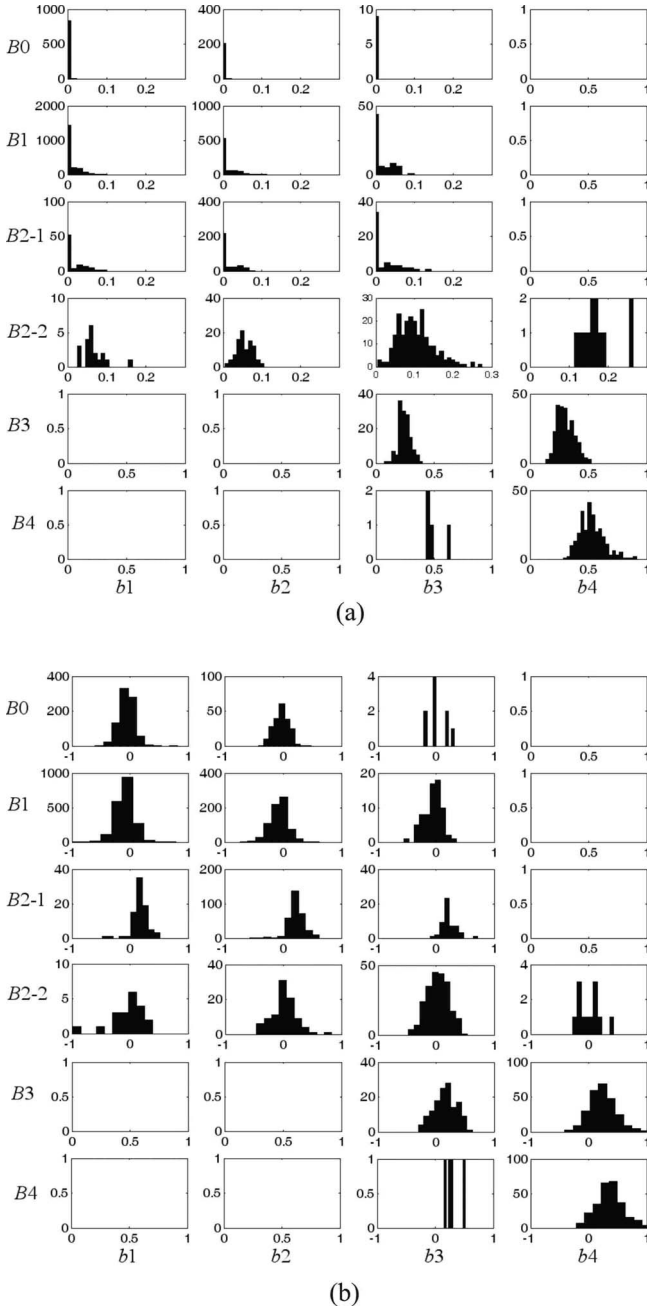


FIG. 13. The histograms of (a) pause duration (in seconds) and (b) normalized pitch jump (in log $F0$) for syllable-juncture instances belonging to subgroups with different break-index pairs labeled by the two methods.

three $B3$ and five $B2$-1 had clear normalized log $F0$ reset. Moreover, the curve of integrating APs of prosodic state and the global mean of pitch level showed smoother PW patterns derived via removing the tone and coarticulation effects from the observed zigzag curve of log $F0$ mean. We then compared the results of the two labeling methods. It can be found from Fig. 14(a) that aside from giving indices of breaks to all the above-mentioned breaks labeled by our method, human labelers gave four additional breaks to divide the nine-syllable-NP (*xing-zheng-yuan zhu-ji-chu de tong-ji*) PW into three PWs, and the two four-syllable compound-word PWs, "*jin-kou* (import) *jin-e* (the amount of money)" and "*qu-nian* (last year) *tong-qi* (the same period)," into four two-syllable words. To justify whether the deletions of these four human-

TABLE XI. KL2 distances measuring the difference between two acoustic feature distributions that belong to different break indices labeled by the same method: (a) the proposed method and (b) human labeling. Upper and lower triangular matrices represent KL2 distances for pause duration and normalized pitch jump, respectively.

		*B*0	*B*1	*B*2-1	*B*2-2	*B*3	*B*4
(a)	*B*0		2.63	3.39	23.59	23.42	22.77
	*B*1	0.19		0.16	14.21	23.28	22.66
	*B*2-1	4.59	4.87		11.92	21.17	20.62
	*B*2-2	0.52	0.72	2.79		13.84	18.85
	*B*3	1.66	2.12	1.25	1.43		12.71
	*B*4	3.69	4.18	0.36	2.50	0.88	
			*b*1	*b*2	*b*3	*b*4	
(b)		*b*1		0.12	6.83	23.16	
		*b*2	0.24		6.07	22.10	
		*b*3	0.60	0.36		10.56	
		*b*4	2.05	1.20	0.82		

labeled breaks were reasonable, we examined the pause durations of these four word junctures and the normalized pitch patterns of the three integrated PWs. The pause durations were 12, 40, 22, and 1 ms. Obviously, they were all not significant. Besides, as seen in Fig. 14(b) all the three normalized pitch patterns of nine-syllable-NP PW and two four-syllable compound-word PWs were smooth. So the deletions of these four breaks by our method seemed reasonable.

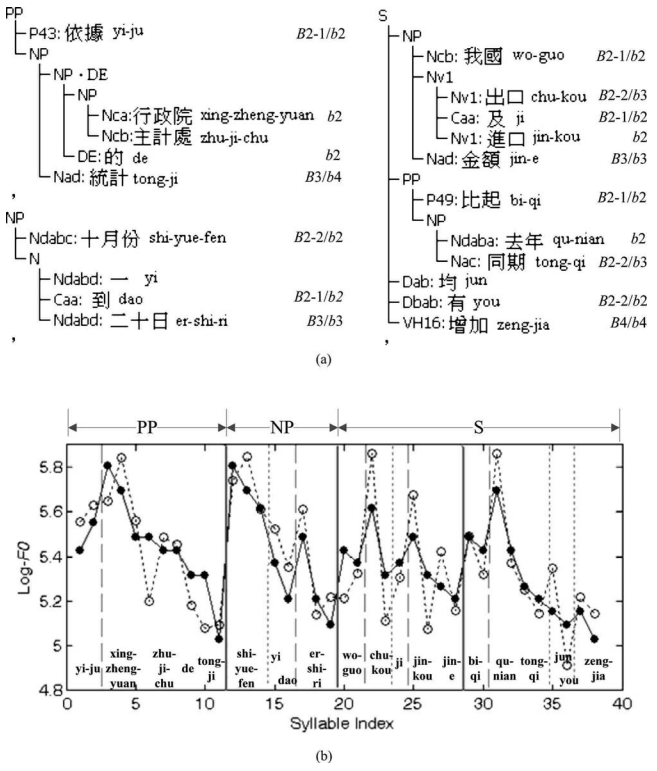


FIG. 14. An example of the automatic prosody labeling. (a) Syntactic trees with prosodic tags: uppercase *B* and lowercase *b* for break index labeled by our method and the human labeler, respectively, and (b) syllable log *F*0 means: observed (open circle) and prosodic state+global mean (close circle). Solid/dashed/dotted lines represent *B*3/*B*2-1/*B*2-2, respectively. The utterance is "*yi-ju* (according to) *xing-zheng-yuan* (the Executive Yuan) *zhu-ji-chu* (Directorate-General of Budget, Accounting and Statistics) *de* (DE) *tong-ji* (statistics), *shi-yue-fen* (October) *yi* (1st) *dao* (to) *er-shi-ri* (20th), *wo-guo* (our country) *chu-kou* (export) *ji* (and) *jin-kou* (import) *jin-e* (the amount of money) *bi-qi* (in comparison with) *qu-nian* (last year) *tong-qi* (the same period) *jun* (both) *you* (to have some) *zeng-jia* (increase).

V. CONCLUSIONS

In this paper, a new approach of joint prosody labeling and modeling for Mandarin speech has been proposed. It first employed four prosodic models to describe the relationship of two types of prosodic tags to be labeled with the input acoustic prosodic features and linguistic features, and then used a sequential optimization procedure to determine all prosodic tags and estimate the parameters of the four prosodic models jointly using the Sinica Treebank speech corpus. Experimental results showed that the estimated parameters of the four prosodic models were able to penetratingly explore and appropriately describe the hierarchy of Mandarin prosody. First, the syllable pitch contour model was able to interpret the variation in syllable pitch contour controlled by such affecting factors as lexical tones, adjacent breaks, and prosodic state. Next, the prosodic-state model was developed to clearly describe the declination effect of log *F*0 level within PW and the resets across PW, PPh, and BG/PG, and hence to extract the pitch patterns of each prosodic constituent. Then, the break-acoustics model could demonstrate the distinct acoustic characteristics for each of the six break types. The last model, the break-syntax model, was built to express the general relationship between the break type and the linguistic features of various levels. Besides, the performance of our models was further confirmed by the corresponding relationships found between the break indices labeled and their associated words which served as evidences to manifest the connections between prosodic and linguistic parameters, and it was also verified by our more consistent and discriminative prosodic feature distributions than those in human labeling by a quantitative comparison. In conclusion, the method we proposed to develop the joint prosody labeling and modeling for Mandarin speech was able to construct interpretive prosodic models and generate prosodic tags that were automatically and consistently labeled.

Some future works are worth doing. First, the syllable pitch contour model can be extended to jointly model syl-

lable pitch contour, syllable duration, and syllable energy level simultaneously. Second, the database with prosodic tags being properly labeled can be used to exploit the hierarchical structure of Mandarin prosody in more detail. Third, the break-syntax model can be extended to consider more linguistic features and applied to the problem of break-type prediction from linguistic features. Fourth, the break-acoustics model can be extended to include more acoustic and linguistic features and applied to the problems of speech segmentation and punctuation generation. Lastly, the four prosodic models can be used to provide useful prosodic information to assist in ASR.

ACKNOWLEDGMENTS

This work was supported by the NSC of Taiwan under Contract Nos. NSC95-2218-E-002-027 and NSC95-2752-E009-014-PAE. The authors would like to thank Academia Sinica, Taiwan for providing the Tree-Bank text corpus, and Dr. Ho-Hsien Pan of Phonetics Laboratory, Department of Foreign Languages and Literatures of National Chiao Tung University, Taiwan for her generous and helpful assistance in manually labeling our experimental database.

APPENDIX: THE ALGORITHM TO DETERMINE ALL THRESHOLDS OF THE DECISION TREE FOR INITIAL BREAK LABELING

1. Determinations of Th1, Th2, and Th3

Th1, Th2, and Th3 are pause-duration thresholds set to sequentially distinguish $B4$, $B3$, and $B2\text{-}2/B1$ with significant pause duration from other break types. First, the two gamma distributions for $B3$ and $B4$ are estimated using two clusters of pause duration samples of syllable juncture with PM clustered by VQ. The one with larger mean is regarded as the distribution for $B4$, and another is for $B3$. We then construct an empirical gamma distribution of pause duration $f_{B0/B1}(\mathrm{pd})$ for $B0/B1$ by using all samples of intraword juncture. An empirical distribution of pause duration $f_{B2\text{-}2}(\mathrm{pd})$ for $B2$-2 is then constructed by using all samples of interword juncture without PM but with apparent pause. Here, the condition of apparent pause is evaluated based on the criterion of $f_{B3}(\mathrm{pd}_n) > f_{B0/B1}(\mathrm{pd}_n)$ which can exclude non-PM interword samples with pause duration similar to those of $B0/B1$. Lastly, the thresholds Th3, Th2, and Th1 are set as the equal-probability intersections of $f_{B0/B1}(\mathrm{pd})$, $f_{B2\text{-}2}(\mathrm{pd})$, $f_{B3}(\mathrm{pd})$, and $f_{B4}(\mathrm{pd})$.

2. Determination of Th5

The pitch jump threshold Th5 is set to distinguish between $B2$-1 and $B0/B1$. We first define the normalized log $F0$ level jump by

$$\xi_n = (\mathbf{sp}_{n+1}(1) - \boldsymbol{\beta}_{t_{n+1}}(1)) - (\mathbf{sp}_n(1) - \boldsymbol{\beta}_{t_n}(1)), \qquad \text{(A1)}$$

where $\mathbf{x}(1)$ denotes the first dimension of vector $\mathbf{x}$. It is noted that the APs of five tones, $\boldsymbol{\beta}_t$, can be estimated in advance before break-type labeling by simply averaging all samples of each tone. Then two empirical Gaussian distributions of normalized log $F0$ level jump, $f_{\mathrm{intra}}(\xi)$ and $f_{\mathrm{PM}}(\xi)$, for intraword and PM junctures are constructed using all samples of intraword syllable junctures and all PM junctures, respectively. We then construct an empirical Gaussian distribution of normalized log $F0$ level jump $f_{B2\text{-}1}(\xi)$ for $B2$-1 by using all samples of interword junctures without PM but with apparent normalized log $F0$ level jump. The condition of apparent normalized log $F0$ level jump is evaluated based on the criterion of $f_{\mathrm{PM}}(\xi_n) > f_{\mathrm{intra}}(\xi_n)$ which can exclude non-PM interword junctures with normalized log $F0$ level jump similar to intraword juncture. Lastly, the threshold Th5 is set as the equal-probability intersection of $f_{\mathrm{intra}}(\xi)$ and $f_{B2\text{-}1}(\xi)$.

3. Determinations of Th4 and Th6

The $F0$ pause duration threshold Th4 and the energy-dip level threshold Th6 are set to distinguish between $B0$ and $B1$. Basically, $B1$ should have very short $F0$ pause duration and large energy-dip level because it represents tightly coupling syllable juncture. So, we simply set Th4 to be 1 frame (=10 ms). For Th6, the two Gaussian distributions for $B0$ and $B1$ are estimated using two clusters of energy-dip level samples of intraword juncture clustered by VQ. Then, the threshold Th6 is set as the equal-probability intersection of the two Gaussian distributions.

[1] E. Selkirk, "On prosodic structure and its relation to syntactic structure," *Nordic Prosody* (Tapir, Trondheim, Norway), Vol. **2**, pp. 111–140.
[2] E. Selkirk, *Phonology and Syntax: The Relation Between Sound and Structure* (MIT Press, Cambridge, MA, 1984).
[3] M. Beckman and J. Pierrehumbert, "Intonational structure in Japanese and English," *Phonology Yearbook 3* (Cambridge University Press, UK, 1986), pp. 255–309.
[4] A.-J. Li, Y.-Q. Zu, and Z.-Q. Li, "A national database design and prosodic labeling for speech synthesis," Proceedings of the Oriental COCOSDA Workshop 1999, pp. 13–16.
[5] A.-J. Li and M.-C. Lin, "Speech corpus of Chinese discourse and the phonetic research," Proceedings of the ICSLP 2000, Vol. **4**, pp. 13–18.
[6] J.-F. Cao, "Rhythm of spoken Chinese—Linguistic and paralinguistic evidences," Proceedings of the ICSLP 2000, Vol. **2**, pp. 357–360.
[7] C.-Y. Tseng, S.-H. Pin, Y.-L. Lee, H.-M. Wang, and Y.-C. Chen, "Fluent speech prosody: Framework and modeling," Speech Commun. special issue on quantitative prosody modeling for natural speech description and generation, **46**, 284–309 (2005).
[8] S.-H. Pin, Y.-L. Lee, Y.-C. Chen, H.-M. Wang, and C.-Y. Tseng, "A Mandarin TTS system with an integrated prosodic model," Proceedings of the ISCSLP 2004, pp. 169–172.
[9] N.-H. Pan, W.-T. Jen, S.-S. Yu, M.-S. Yu, S.-Y. Huang, and M.-J. Wu, "Prosody model in a Mandarin text-to-speech system based on a hierarchical approach," Proceedings of the ICME 2000, Vol. **1**, pp. 448–4511.
[10] S.-H. Chen, S.-H. Hwang, and Y.-R. Wang, "An RNN-based prosodic information synthesizer for Mandarin text-to-speech," IEEE Trans. Speech Audio Process. **6**, 226–239 (1998).
[11] Y. Liu, E. Shriberg, S. Stolcke, D. Hillard, M. Ostendorf, and M. Harper, "Enriching speech recognition with automatic detection of sentence boundaries and disfluencies," IEEE Trans. Audio, Speech, Lang. Process. **14**, 526–1540 (2006).
[12] Y. Gotoh and S. Renals, "Sentence boundary detection in broadcast speech transcripts," Proceedings of the ISCA Workshop: Automatic Speech Recognition: Challenges for the New Millennium ASR 2000, pp. 228–235.
[13] E. Shriberg, A. Stolcke, D. Hakkani-Tur, and G. Tur, "Prosody-based automatic segmentation of speech into sentences and topics," Speech Commun. **32**, 127–154 (2000).
[14] J.-H. Kim and P. C. Woodland, "A combined punctuation generation and speech recognition system and its performance enhancement using prosody," Speech Commun. **41**, 563–577 (2003).
[15] J.-H. Kim and P. C. Woodland, "The use of prosody in a combined system for punctuation generation and speech recognition," Proceedings of the

Eurospeech 2001, pp. 2757–2760.
[16]H. Christensen, Y. Gotoh, and S. Renals, "Punctuation annotation using statistical prosody models," Proceedings of the ISCA Workshop on Prosody in Speech Recognition and Understanding 2001, pp. 35–40.
[17]J.-F. Yeh and C.-H. Wu, "Edit disfluency detection and correction using a cleanup language model and an alignment model," IEEE Trans. Audio, Speech, Lang. Process. **14**, 1574–1583 (2006).
[18]M. Lease, M. Johnson, and E. Charniak, "Recognizing disfluencies in conversational speech," IEEE Trans. Audio, Speech, Lang. Process. **14**, 1566–1573 (2006).
[19]C.-K. Lin and L.-S. Lee, "Improved spontaneous Mandarin speech recognition by disfluency interruption point (IP) detection using prosodic features," Proceedings of the Eurospeech 2005, pp. 1621–1624.
[20]K. Chen and M. Hasegawa-Johnson, "How prosody improves word recognition," Proceedings of the ISCA International Conference on Speech Prosody 2004, pp. 583–586.
[21]K. Chen, M. Hasegawa-Johnson, A. Cohen, S. Borys, S.-S. Kim, J. Cole, and J.-Y. Choi, "Prosody dependent speech recognition on radio news corpus of American English," IEEE Trans. Audio, Speech, Lang. Process. **14**, 232–245 (2006).
[22]K. Chen and M. Hasegawa-Johnson, "Improving the robustness of prosody dependent language modeling based on prosody syntax dependence," Proceedings of the IEEE ASRU 2003, pp. 435–440.
[23]E. Shriberg and A. Stolcke, "Direct modeling of prosody: An overview of applications in automatic speech processing," Proceedings of the ISCA International Conference on Speech Prosody 2004, pp. 575–582.
[24]J.-H. Yang, Y.-F. Liao, Y.-R. Wang, and S.-H. Chan, "A new approach of using temporal information in Mandarin speech recognition," Proceedings of the ISCA International Conference on Speech Prosody 2006, Vol. **SPS4-3**.
[25]X. Lei and M. Ostendorf, "Word-level tone modeling for Mandarin speech recognition," Proceedings of the IEEE ICASSP 2007, Vol. **4**, pp. 665–668.
[26]C.-Y. Tseng, "Recognizing Mandarin Chinese fluent speech using prosody information—An initial investigation," Proceedings of the ISCA International Conference on Speech Prosody 2006.
[27]K. Silverman, M. Beckman, J. Pitrelli, M. Ostendorf, C. Wightman, P. Price, J. Pierrehumbert, and J. Hirschberg, "ToBI: A standard for labeling English prosody," Proceedings of the ICSLP 1992, Vol. **2**, pp. 867–870.
[28]A. Batliner, J. Buckow, H. Niemann, E. Noth, and V. Warnke, "The prosody module," in *Verbmobil: Foundations of Speech-to-Speech Translation*, edited by W. Wahlster (Springer, New York, 2000).
[29]M. Selting, *Prosody in Conversation* (Max Niemeyer, Tuebingen, Germany, 1995), in German.
[30]D. J. Hirst, "The symbolic coding of fundamental frequency curves: From acoustics to phonology," Proceedings of the International Symposium on Prosody 1994.
[31]P. A. Taylor, "The tilt intonation model," Proceedings of the ICSLP 1998, Vol. **4**, pp. 1383–1386.
[32]S.-H. Peng, M. K. M. Chan, C.-Y. Tseng, T. Huang, O.-J. Lee, and M. Beckman, "Towards a Pan-Mandarin system for prosodic transcription," in *Prosodic Typoloxgy: The Phonology of Intonation and Phrasing*, edited by S.-A. Jun (Oxford University Press, Oxford, 2005), pp. 230–270.
[33]A.-J. Li, "Chinese prosody and prosodic labeling of spontaneous speech," Proceedings of the ISCA International Conference on Speech Prosody 2002, pp. 39–46.
[34]G.-P. Chen, G. Bailly, Q.-F. Liu, and R.-H. Wang, "A superposed prosodic model for Chinese text-to-speech synthesis," Proceedings of the ISCSLP 2004, pp. 177–180.
[35]M.-S. Yu, N.-H. Pan, and M.-J. Wu, "A statistical model with hierarchical structure for predicting prosody in a Mandarin text-to-speech system," Proceedings of the ISCSLP 2002, pp. 21–24.
[36]G. Bailly and B. Holm, "SFC: A trainable prosodic model," Speech Commun. **46**, 348–364 (2005).
[37]M. Ostendorf and N. Veilleux, "A hierarchical stochastic model for automatic prediction of prosodic boundary location," Comput. Linguist. **20**, 27–52 (1994).
[38]X. Shen and B. Xu, "A CART-based hierarchical stochastic model for prosodic phrasing in Chinese," Proceedings of the ISCSLP 2000, pp. 105–109.
[39]H.-J. Peng, C.-C. Chen, C.-Y. Tseng, and K.-J. Chen, "Predicting prosodic words from lexical words—A first step towards predicting prosody from text," Proceedings of the ISCSLP 2004, pp. 173–176.
[40]J. Hirschberg and P. Prieto, "Training intonational phrasing rules automatically for English and Spanish text-to-speech," Speech Commun. **18**, 281–290 (1996).
[41]D.-W. Xu, H.-F. Wang, G.-H. Li, and T. Kagoshima, "Parsing hierarchical prosodic structure for Mandarin speech synthesis," Proceedings of the IEEE ICASSP 2006, Vol. **1**, pp. 14–19.
[42]X. Sun and T. H. Applebaum, "Intonational phrase break prediction using decision tree and n-gram model," Proceedings of the Eurospeech 2001, pp. 537–540.
[43]A. W. Black and P. Taylor, "Assigning phrase breaks from part-of-speech sequences," Proceedings of the Eurospeech 1997, pp. 995–998.
[44]Z. Sheng, J.-H. Tao, and D.-L. Jiang, "Chinese prosodic phrasing with extended features," Proceedings of the IEEE ICASSP 2003, Vol. **1**, pp. 492–495.
[45]J.-F. Li, G.-P. Hu, and R.-H. Wang, "Chinese prosody phrase break prediction based on maximum entropy model," Proceedings of the Interspeech 2004, pp. 729–732.
[46]Y.-Q. Shao, Y.-Z. Zhao, J.-Q. Han, and T. Liu, "Using different models to label the break indices for mandarin speech synthesis," Proceedings of the ICMLC 2005, Vol. **6**, pp. 3802–3807.
[47]J.-F. Li, G.-P. Hu, R.-H. Wang, and L.-R. Dai, "Sliding window smoothing for maximum entropy based intonational phrase prediction in Chinese," Proceedings of the IEEE ICASSP 2005, Vol. **1**, pp. 285–288.
[48]Z.-P. Zhao, T.-J. Zhao, and Y.-T. Zhu, "A maximum entropy Markov model for prediction of prosodic phrase boundaries in Chinese TTS," Proceedings of the IEEE GrC 2007, pp. 498–498.
[49]C. W. Wightman and M. Ostendorf, "Automatic labeling of prosodic patterns," IEEE Trans. Speech Audio Process. **2**, 469–481 (1994).
[50]K. Chen, M. Hasegawa-Johnson, and A. Cohen, "An automatic prosody labeling system using ANN-based syntactic-prosodic model and GMM-based acoustic-prosodic model," Proceedings of the IEEE ICASSP 2004, Vol. **1**, pp. 509–512.
[51]V. Rangarajan, S. Narayanan, and S. Bangalore, "Acoustic-syntactic maximum entropy model for automatic prosody labeling," Proceedings of the IEEE Spoken Language Technology Workshop 2006, pp. 74–77.
[52]X.-J. Ma, W. Zhang, Q. Shi, W.-B. Zhu, and L.-Q. Shen, "Automatic prosody labeling using both text and acoustic information," Proceedings of the IEEE ICASSP 2003, Vol. **1**, pp. 516–519.
[53]A. F. Muller, H. G. Zimmermann, and R. Neuneier, "Robust generation of symbolic prosody by a neural classifier based on autoassociators," Proceedings of the IEEE ICASSP 2000, Vol. **3**, pp. 1285–1288.
[54]J.-H. Tao, "Acoustic and linguistic information based Chinese prosodic boundary labeling," Proceedings of the TAL 2004, pp. 181–184.
[55]S.-H. Chen, W.-H. Lai, and Y.-R. Wang, "A statistics-based pitch contour model for Mandarin speech," J. Acoust. Soc. Am. **117**, pp. 908–925 (2005).
[56]S.-H. Chen, W.-H. Lai, and Y.-R. Wang, "A new duration modeling approach for Mandarin speech," IEEE Trans. Speech Audio Process. **11**, 308–320 (2003).
[57]S.-H. Chen and Y.-R. Wang, "Vector quantization of pitch information in Mandarin speech," IEEE Trans. Commun. **38**, 1317–1320 (1990).
[58]L. Breiman, J. Friedman, R. Olshen, and C. Stone, *Classification and Regression Trees* (Wadsworth, Belmont, CA, 1984).
[59]Y. Qian and W.-Y. Pan, "Prosodic word: The lowest constituent in the Mandarin prosody processing," Proceedings of the ISCA International Conference on Speech Prosody 2002, pp. 591–594.
[60]J.-H. Tao, H.-G. Dong, and S. Zhao, "Rule learning based Chinese prosodic phrase prediction," Proceedings of the IEEE NLP-KE 2003, pp. 425–432.
[61]C.-R. Huang, K.-J. Chen, F.-Y. Chen, Z.-M. Gao, and K.-Y. Chen, "Sinica Treebank: Design criteria, annotation guidelines, and pn-line interface," Proceedings of the Second Chinese Language Processing Workshop 2000, pp. 29–37.
[62]Y.-R. Chao, *A Grammar of Spoken Chinese* (Berkeley Press, Berkeley, CA, 1968).
[63]L.-S. Lee, C.-Y. Tseng, and M. Ouh-Young, "The synthesis rules in a Chinese text-to-speech system," IEEE Trans. Acoust., Speech, Signal Process. **37**, 1309–1320 (1989).
[64]Y. Xu, "Contextual tonal variations in Mandarin," J. Phonetics **25**, 61–83 (1997).
[65]K.-J. Chen and C.-R. Huang, "Part of speech (POS) analysis on Chinese language," CKIP Technical Report No. 93-05, Institute of Information Science, Academia Sinica, Taiwan, R.O.C., 1993 (in Chinese).
[66]Chinese knowledge Information Processing (CKIP), Academia Sinica, "An introduction to Academia Sinica balanced corpus for modern Mandarin Chinese," CKIP Technical Report No. 95-02, Institute of Information

Science, Academia Sinica, Taiwan, R.O.C. 1995 (in Chinese).
[67]C. Shih, "Declination in Mandarin," Proceedings of the ESCA Workshop on Intonation: Theory, Models and Applications 1997, pp. 293–296.
[68]Y. Yufang and W. Bei, "Acoustic correlates of hierarchical prosodic boundary in Mandarin," Proceedings of the ISCA International Conference on Speech Prosody 2002, pp. 707–710.
[69]C.-Y. Tseng and S.-H. Pin, "Mandarin Chinese prosodic phrase grouping and modeling: Method and implications," Proceedings of the TAL 2004, pp. 193–196.
[70]C.-Y. Tseng and S.-H. Pin, "Modeling prosody of Mandarin Chinese fluent speech via phrase grouping," Proceedings of the Speech and Language Systems for Human Communication (SPLASH-2004/Oriental-COCOSDA2004), 2004, pp. 53–57.
[71]C.-Y. Tseng and Z.-Y. Su, "Corpus approach to phonetic investigation—Methods, quantitative evidence and findings of Mandarin speech prosody," Proceedings of the Oriental COCOSDA Workshop 2006, pp. 123–138.
[72]C.-Y. Tseng, "Higher level organization and discourse prosody," Proceedings of the TAL 2006, pp. 23–34.
[73]S. Theodoridis and K. Koutroumbas, *Pattern Recognition* 2nd ed. (Elsevier, London, UK, 2003).

A study of lip movements during spontaneous dialog and its application to voice activity detection

David Sodoyer, Bertrand Rivet, Laurent Girin, Christophe Savariaux, and Jean-Luc Schwartz
GIPSA-lab, Department of Speech and Cognition, UMR 5126 CNRS, Grenoble-INP, Université Stendhal, Université Joseph Fourier, 46 Avenue Félix Viallet, 38031 Grenoble, France

Christian Jutten
GIPSA-lab, Department of Images and Signal, UMR 5126 CNRS, Grenoble-INP, Université Stendhal, Université Joseph Fourier, 46 Avenue Félix Viallet, 38031 Grenoble, France

(Received 19 June 2007; revised 3 October 2008; accepted 14 November 2008)

This paper presents a quantitative and comprehensive study of the lip movements of a given speaker in different speech/nonspeech contexts, with a particular focus on silences (i.e., when no sound is produced by the speaker). The aim is to characterize the relationship between "lip activity" and "speech activity" and then to use visual speech information as a voice activity detector (VAD). To this aim, an original audiovisual corpus was recorded with two speakers involved in a face-to-face spontaneous dialog, although being in separate rooms. Each speaker communicated with the other using a microphone, a camera, a screen, and headphones. This system was used to capture separate audio stimuli for each speaker and to synchronously monitor the speaker's lip movements. A comprehensive analysis was carried out on the lip shapes and lip movements in either silence or nonsilence (i.e., speech+nonspeech audible events). A single visual parameter, defined to characterize the lip movements, was shown to be efficient for the detection of silence sections. This results in a visual VAD that can be used in any kind of environment noise, including intricate and highly nonstationary noises, e.g., multiple and/or moving noise sources or competing speech signals. [DOI: 10.1121/1.3050257]

PACS number(s): 43.72.Ar, 43.72.Kb [KWG] Pages: 1184–1196

I. INTRODUCTION

A. Context: Audiovisual speech processing

Speech is a bimodal signal, both acoustic and visual. Many studies have shown that the visual modality improves the intelligibility of speech in noise when switching from the "audio only" condition to the "audio+speaker's face" condition (Sumby and Pollack, 1954; Erber, 1975; Benoît *et al.*, 1994; Robert-Ribes *et al.*, 1998). In parallel, McGurk and McDonald (1976) demonstrated that humans can even integrate conflicting audio and visual speech stimuli to perceive a "chimeric" speech stimulus. More recently, Grant and Seitz (2000) showed that viewing the speaker's face also improves the detection of speech in noise. Such results have been confirmed by Kim and Davis (2004) and Bernstein *et al.* (2004). More specifically, visual information helps pertinent acoustic features to be better extracted, i.e., "seeing to hear better," providing a different and complementary contribution to lipreading (Schwartz *et al.*, 2004). Additionally, visual speech information has been shown to irresistibly attract speaker's localization (Bertelson, 1999).

Concerning the nature of visual speech information, two major questions have been addressed. First, the oral region including the lips and jaw seems to be the major contributor to visual speech perception (see, e.g., Summerfield, 1979; Benoît *et al.*, 1996). Thomas and Jordan (2004) actually showed that the intelligibility of oral-movement display was more or less the same as that of whole-face-movement display. However, extra-oral movements also influence identification of visual and audiovisual speech, mostly due to the strong correlation between oral and extra-oral movements (Munhall and Vatikiotis-Bateson, 1998). Orofacial configurations can be basically characterized in terms of lip contours and specifically by the parameters of inner lip height, inner lip width, and lip protrusion (Summerfield, 1979; Abry and Boë, 1986; Benoît *et al.*, 1992, 1996). Second, the question of static versus dynamic processing of facial configurations has been largely discussed. Studies using pointlike displays, which remove fine spatial information, showed that movement seems to be crucial in the perceptual processing of visual speech in both noisy configurations (Rosenblum *et al.*, 1996) and conflicting McGurk stimuli (Rosenblum and Saldana, 1996). This led Munhall *et al.* (1996) to suggest that listeners might use the time-varying properties of visual speech for perceptual grouping and phonetic perception. Neurophysiological data seem to confirm the specific role of the dynamic processing of visual speech (Calvert and Campbell, 2003; Munhall *et al.*, 2002). This is compatible with Summerfield's (1987) suggestion that one possible metric for audiovisual integration is the pattern of changes over time in articulation, considering that listeners are sensitive to the dynamics of vocal tract change. Thereafter, a number of studies in the audiovisual speech literature have characterized the correlation between lower-face movement and the produced acoustic signal (Yehia *et al.*, 1998; Barker and Berthommier, 1999; Jiang *et al.*, 2002; Bailly and Badin, 2002; Goecke and Millar, 2003).

Following these considerations on the bimodal aspect of speech, an important number of technological studies have been undertaken in the past 20 years to integrate the visual modality into speech processing systems. The goal is to improve the performance and robustness (in noise) of different human-to-human telecommunication systems or human-computer interfaces. Petajan (1984) was the first to integrate visual speech information in an automatic speech recognition (ASR) system. Many studies followed, including recent advances going toward real-life implementations of bimodal ASR (Potamianos *et al.*, 2003b). Recently, audiovisual speech processing applications also concerned video indexing and retrieval (Huang *et al.*, 1999; Iyengar and Neti, 2001), audiovisual speech synthesis and talking heads (Yehia *et al.*, 2000; Bailly *et al.*, 2003; Cosi *et al.*, 2003; Gibert *et al.*, 2005), and audiovisual speech coding (Rao and Chen, 1996; Girin, 2004). In recent years, the visual modality has also been exploited for speech enhancement in (background) noise (Girin *et al.*, 2001; Deligne *et al.*, 2002; Potamianos *et al.*, 2003a) and more generally for speech source separation, i.e., for the extraction of a speech signal from complex mixtures using several microphones for both linear instantaneous mixtures (Sodoyer *et al.*, 2002, 2004) and convolutive mixtures (Wang *et al.*, 2005; Rivet *et al.*, 2007a).

B. Video characterization of silence versus nonsilence sections

Most of the time, studies addressing the characterization of lip patterns in speech production have been carried out in more or less controlled speech production contexts (typically "laboratory speech:" see, e.g., Abry and Boë, 1986; Benoît *et al.*, 1992; Goecke and Millar, 2003; Jiang *et al.*, 2002; Yehia *et al.*, 1998). Relatively poor attention has been paid to the description and characterization of these patterns during speech production in natural contexts, especially in spontaneous multispeaker conversation. Moreover, in such context, speech activity (i.e., actual speech production by a speaker of interest) alternates with many silence sections (i.e., sections where the speaker of interest does not produce sounds, whereas other speakers may actually do) and also with many nonspeech audible events such as murmurs, grunts, laughs, respiration intakes, expirations, lip noise, whispers, sighs, growls, and moans (Campbell, 2007). In spite of this, even poorer attention has been paid to lip patterns in silence and nonspeech contexts, although these patterns may exhibit a specific behavior to be considered in both audiovisual speech fundamental studies and technological applications.

This paper provides an attempt to fill this gap. The relationship between a speaker's lip movements and speech/nonspeech activity versus silence is investigated using signals from a spontaneous dialog. For this aim, the recording and the study of a "real-life" audiovisual corpus were achieved and are presented in this paper. This corpus consists of two speakers recorded in a spontaneous dialog situation (in French) during about 40 min. It is characterized by two properties. Firstly, it is based on a very clean audio (and, of course, video) recording process since each speaker is located in a separate room to completely avoid cross-speaker audio interferences in the recordings. Communication between the two speakers is effected using a specially designed equipment described in Sec. II A. Secondly, the audiovisual material is recorded in a lively dialog situation, in which various creative contexts lead the two speakers to have a spontaneous discussion (see also more details in Sec. II A). As a result, the recorded signals include speech and silence sections, as well as many different nonspeech audible events such as those mentioned above. It also contains many face expressions and movements with or without sound production [see the related work of Macho *et al.* (2005)]. Using this corpus and starting from a very simple hypothesis—the lips of a given speaker should move when he/she talks (or produces nonspeech sounds), whereas they should not move (or move less) when he/she does not utter sounds—the distributions of static and dynamic lip parameters are provided for the two conditions. Those distributions show how dynamic lip parameters can be associated with nonsilence sections (i.e., speech+nonspeech audible events) versus silence sections. Actually, the correspondence is not straightforward. Indeed, lip movements can occur during silence, and conversely speech or nonspeech oral production can occur with still lips. However, it is shown that a single dynamic lip parameter is more appropriate than static parameters for this characterization and that temporal integration of the dynamic parameter values can improve the "separability" of nonsilence sections versus silence sections from lip information.

C. Application to automatic voice activity detection

Finally, a technological application of the study is considered: the possibility of using visual information to automatically detect sound production and silence sections in a given audio channel. Such an algorithm is called a voice activity detector (VAD), and it is generally derived from audio information only. Among other applications, it can be used to drastically improve the performance of speech enhancement/separation techniques: *silence detection*, i.e., the detection of regions where the speaker of interest does not produce any sound, is used to identify properties of the noise or properties of the mixture configuration. These properties are then used to process the extraction of the speech signal of interest when it is detected as present in the mixture[1] (see, e.g., Ephraim and Malah, 1984; Abrard and Deville, 2003). Various types of audio VAD have been studied, and they can achieve good performance even with a low signal-to-noise ratio (SNR) (Le Bouquin-Jeannès and Faucon, 1995; Sohn *et al.*, 1999; Tanyer and Ozer, 2000; Ramírez *et al.*, 2005). However these techniques are based on the analysis of the acoustic signal, and consequently their performance depends strongly on the environment noise. Generally the noise has to be considered as stationary or weakly nonstationary and/or with a given power spectral density function or probability density function. Thus, when the noise is highly nonstationary with a low SNR (a concurrent speaker, for example), the audio VAD performance considerably decreases. In this case, visual information could be very useful since it is completely independent of the acoustic environment.[2] For instance, in a previous study,

De Cueto *et al.* (2000) used a basic visual voice activity detector (V-VAD) for detecting a speaker's speech activity in front of a computer. For this, either specific lip parameters or the average luminance of the mouth picture can be used (Iyengar and Neti, 2001). However, those studies are limited to the speaker's "intent to speak," useful for, e.g., turn-taking detection. The methods do not provide accurate segmentation of the content of a given speaker's sequences. More recently, Liu and Wang (2004) proposed a V-VAD based on Gaussian models. One Gaussian kernel was used to model the silence/nonspeech sections, and two kernels were used to model the speech sections.[3] However, little information is reported on the video processing, on the nature of the corpus that is used for setting and testing the V-VAD, and even on the visual information itself: it is not clear whether static or dynamic information is used. Also, the size of the experimental data is not compatible with real-life applications. The V-VAD proposed in the present paper specifically addresses these last remarks: it is based on real-life audiovisual data (and it is tested using these data) while remaining simple (given that lip shape parameters are available). Its efficiency is demonstrated by a series of detection scores [receiver operating characteristics (ROCs)]. As mentioned before, this V-VAD can be used in a speech enhancement system or a source separation system [see, for instance, Rivet *et al.* (2007b) for a first application of V-VAD to the speech source separation problem].

This paper is organized as follows. Section II presents the method, beginning with a description of the audiovisual corpus (Sec. II A) including the recording conditions and the definition of the video (lip) parameters used in this study. This is followed by the description of the audio (Sec. II B) and video (Sec. II C) processing applied to the data. The lip dynamic parameter used for silence versus nonsilence characterization and VAD is described in detail in Sec. II D. Section III presents the results of the study: in Sec. III A, the audio content of the corpus in terms of silence versus nonsilence sections is presented. Then, Sec. III B provides an analysis of the properties of the static and dynamic lip parameters in silence versus nonsilence sections. The performance of the proposed V-VAD in terms of ROC curves is given in Sec. III C. Section IV presents our conclusions.

II. METHOD

A. Description of the audiovisual corpus

To describe and characterize lip movements in relation with speech/sound production or nonproduction requires the acquisition of appropriate audiovisual data. An original audiovisual corpus was thus recorded and processed, consisting of a series of spontaneous dialogs between two male French speakers (JLS and LG). To obtain a set of conversation situations as natural as possible, several tasks were suggested to the speakers. These tasks were, e.g., different interactive games such as answering as fast as possible to a word association problem, finding the solution of riddles, or playing language games. In all these tasks, the interaction between speakers was totally spontaneous, thus including spontaneous turn taking, interruptions, hesitations, and possible cross-

FIG. 1. Illustrations of the audiovisual corpus recording session. The two speakers are in separate rooms. A specially designed equipment is used for the real-time transmission of audio and video signals between the speakers, as well as the recording of these signals.

overlapping between speakers. This led each of them to alternate between natural silence sections and speech sections of various sizes and contents. The corpus also contains many different kinds of audible and nonaudible nonspeech events, such as those mentioned in Sec. I.

The two speakers were placed and recorded in separate rooms. They both had a microphone and a microcamera fixed on a light helmet. The camera focused on the lip region to optimize the capture of labial information. Moreover, the speakers could hear and see each other using headphones and a monitor screen in front of them with real-time video feedback. This was necessary to ensure "naturalness and conviviality" during the conversation. Automatic time-code generators were used for postprocessing synchronization of all audio/video signals. Finally, these experimental settings enabled the conditions of a real face-to-face conversation to be simulated while the recorded audio signals (and, of course, the video signals) were perfectly separate. Illustrations of the recording session are given in Fig. 1.

The visual information extracted from this corpus consists of the time trajectories of two basic geometric parameters characterizing the lip contour (see Sec. I A), namely, inner width l_w and inner height l_h (Fig. 2). These parameters were extracted using the ICP "face processing system" (Lallouache, 1990), which is based on blue make-up, image thresholding with the Chroma-Key system, and contour tracking algorithms. The parameters were extracted every 20 ms (the video sampling frequency is 50 Hz) synchronously with the acoustic signal, which is sampled at 44.1 kHz. Thus, in the following, a signal *frame* is defined as a 20 ms section of acoustic signal together with a pair of lip parameters (l_w, l_h). A spontaneous audiovisual speech corpus for two speakers with a total duration of 40 min was finally

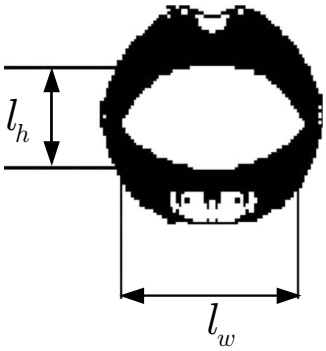


FIG. 2. The lip parameters used in this study: inner lip height (l_h) and inner lip width (l_w).

obtained, representing 120 000 vectors of audiovisual frames per speaker.

B. Audio analysis and silence/nonsilence labeling

The first phase of the corpus analysis consisted in the labeling of the 20 ms frames (corresponding to the video sampling) as "silence frames" or "nonsilence frames" based on the analysis of the audio signal and the dichotomy defined in Sec. I: Silence frames are defined as signal frames with no sound produced at all, and nonsilence frames contain speech and nonspeech acoustic events. It is important to note that these definitions are given here for each speaker independently (obviously, a silence frame for one speaker can be simultaneous with a nonsilence frame for the other speaker since the two tracks are recorded separately). Silence frames are mainly present between phrase boundaries that result from conversation turn taking and also in more or less long pauses within one speaker's "continuous" talk due to, e.g., hesitations.

The labeling into silence frames versus nonsilence frames was made semiautomatically with the algorithm proposed by Ramirez *et al.* (2004) and a manual verification. This algorithm measures the long-term spectral divergence between speech and environment noise and formulates the decision rule by comparing the long-term spectral envelope to the average noise spectrum, thus yielding a high discriminating decision rule and minimizing the average number of decision errors. The decision threshold is adapted to the measured noise. In our case, the environment noise was generally very low, and the results of this labeling were almost perfect. A manual verification of the entire corpus was made, and a very small number of errors were corrected. It can be noted that very short silences corresponding to the time periods preceding the release of unvoiced plosives are not considered as silence frames, even though they may happen to be slightly greater than 20 ms. This is because of the nature of the audio detection algorithm that considers longer signal sections. Conveniently, this is coherent with the definition and processing of the temporal integration step that we propose in Sec. II D.

C. Video preprocessing

As mentioned before, the extracted visual information is the time trajectory of the geometric parameters l_w and l_h characterizing the lip contour. The measures provided by the face processing system, although very accurate, are slightly noisy. Since a dynamic video parameter is calculated from the derivatives of the temporal trajectories, computed by a difference operator, which is very sensitive to noise, the lip parameter trajectories have to be filtered (smoothed). This is not a trivial task for such signals since labial parameter trajectories are highly nonstationary signals: slow variations in time can be followed by drastic changes, for instance, when lips are closing. Therefore, it is difficult to remove noise in regions with slow variations while respecting the abrupt variations provided by natural lip movements. In our study, a technique based on spline functions was used. A basic version of this technique has been successfully used in a previous study using audiovisual corpora (Girin, 2004), and this process is refined here as follows.

The basic principle of the spline smoothing consists in locally fitting (noisy) data $x(i)$ with a cubic spline $s(t_i)$ defined as piecewise polynomial functions, where each piece is described using a cubic polynomial. The fitting is based on the minimization of the following criterion:

$$f = p\sum_{j=1}^{J} w(j)|x(j) - s(t_j)|^2 + (1-p)\int \left(\frac{\partial s}{\partial t}\right)^2 dt. \qquad (1)$$

The first term is a weighted least-squares error between data and the spline model [the weights are given by $w(i)$], and the second term stands for the smoothness of the resulting curve. Balancing these two constraints is made possible by setting the parameter p at an appropriate value between 0 and 1. For instance, $p=0$ produces a least-squares straight line fit to the data, $p=1$ produces a cubic spline interpolate, and intermediate values provide a trade-off between close fit and smoothness.

In the proposed video processing system, the nonstationary property of the lip movements is taken into account by adaptively tuning the p parameter according to the signal dynamics. Relatively large p values must be used in time sections with high natural variations of the lip parameters to closely track these variations. On the contrary, relatively small p values must be used in quasistationary regions to adequately remove the noise. Thus the lip parameter signals l_w and l_h are segmented in time sections depending on the value of their local (sliding) variance $C(t)=1/N\Sigma_{n=-N/2}^{N/2}\nu(t+n)^2$, with $N=6$ [$\nu(t)$ represents a visual parameter (l_w or l_h), and t denotes the time index of 20 ms frames].

Each section is then fitted with a cubic spline whose parameter p is determined as a function of this variance. More specifically, this automatic smoothing process for each visual parameter $\nu(t)$ is the following:

- Compute for each frame the local variance $C(t)$.
- Search sections of consecutive frames with a variance $C(t)$ lower than a fixed threshold $C_{\min}$ defining a quasistationary signal section. Then all other frames are considered as nonstationary. This provides alternations of quasistationary sections and nonstationary sections with variable lengths.
- For each section i compute the mean of $C(t)$ over the section,

$$\bar{C}_i = \frac{1}{T_i} \sum_{t=t_i}^{t_i+T_i-1} C(t) \tag{2}$$

(T_i denotes the size of the section i, and t_i denotes the index of the first frame of the section) and compute p_i so that

$$p_i = \begin{cases} p_{\min} & \text{if } \log_{10} \bar{C}_i < \lambda_{\min} \\ \dfrac{p_{\max} - p_{\min}}{\lambda_{\max} - \lambda_{\min}} \log_{10} \bar{C}_i - \dfrac{p_{\min}\lambda_{\max} - p_{\max}\lambda_{\min}}{\lambda_{\max} - \lambda_{\min}} & \text{if } \lambda_{\min} \leqslant \log_{10} \bar{C}_i \leqslant \lambda_{\max}, \\ p_{\max} & \text{if } \log_{10} \bar{C}_i > \lambda_{\max} \end{cases} \tag{3}$$

where the thresholds are fixed as

$$\lambda_{\min} = \log_{10}\left(\frac{\text{std}(C)}{50}\right), \quad p_{\min} = 0.0001,$$

$$\lambda_{\max} = \log_{10}(5\ \text{std}(C)), \quad p_{\max} = 0.8.$$

Finally, the weights $w(i)$ of Eq. (1) are assumed to be equal to 1 for all data. This process is applied on each parameter $l_w(t)$ and $l_h(t)$ to obtain the smoothed visual parameters $\tilde{l}_w(t)$ and $\tilde{l}_h(t)$.[4] An illustration of the results obtained with this process is given in Sec. III B.

D. A dynamic lip parameter for silence versus nonsilence characterization and automatic silence detection

In Sec. I A, we have briefly discussed the importance of the lip *movements* (as opposed to static lip shapes) for characterizing audiovisual speech. In a preliminary work, lip movements have been shown to be good candidates to characterize the opposition between silence and nonsilence activities (Sodoyer *et al.*, 2006), the lip shape variations being generally smaller in silence sections. Therefore, following this previous work, we chose to describe the lip shape movements with one dynamic parameter, summing the absolute values of the two lip parameter derivatives (Sodoyer *et al.*, 2006),

$$\pi(t) = \left|\frac{\partial \tilde{l}_w(t)}{\partial t}\right| + \left|\frac{\partial \tilde{l}_h(t)}{\partial t}\right|. \tag{4}$$

Large $\pi(t)$ values indicate significant lip movements and should index nonsilence frames, while low values corresponding to small lip movements (or no movement at all) should index silence sections. Note that this dynamic parameter exploits the complementarity between the two lip parameters for many speech sequences (see Fig. 3). Indeed, the variations of $\tilde{l}_w(t)$ may characterize rounding movements during which lip height may not change much, and vice versa, the variations of $\tilde{l}_h(t)$ may characterize opening/closing movements during which lip width may not change much. For example, in Fig. 3, the variations of the width parameter are larger than the variations of the height parameter between 278.5 and 278.8 s, and the contrary occurs between 278 and 278.2 s.

However, the situation is not so simple. On the one hand, instantaneous large $\pi(t)$ values can correspond to local short lip movements in silence sections (e.g., smiles, grimacing, funny faces, or changes in the lip "rest position"). This is likely to produce silence detection errors (silence classified as nonsilence). On the other hand, local lip stability within speech gestures can lead to low local $\pi(t)$ values providing false alarms (speech classified as silence). To overcome these problems, $\pi(t)$ values are then summed over time. Therefore, the parameter $\rho(t)$ is defined from the filtering of $\pi(t)$ as

$$\rho(t) = h(t) * \pi(t), \tag{5}$$

with $h(t)$ being the truncated version of a first-order low-pass filter defined by

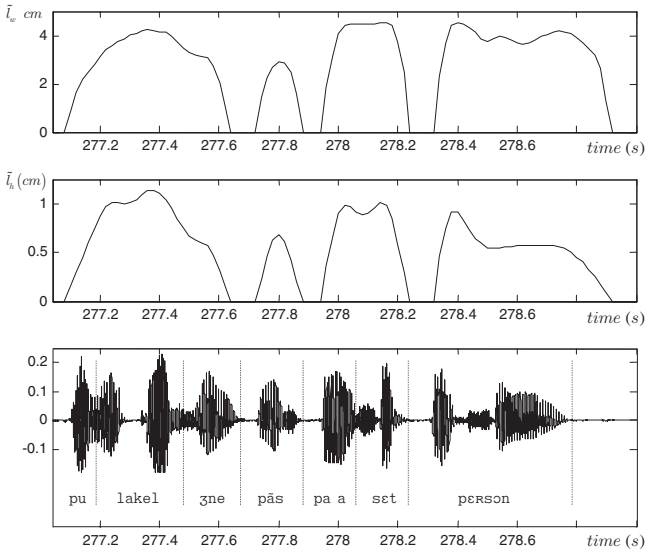


FIG. 3. Example of lip parameter trajectories: (top) inner width parameter, (middle) inner height parameter, and (bottom) corresponding acoustic signal.

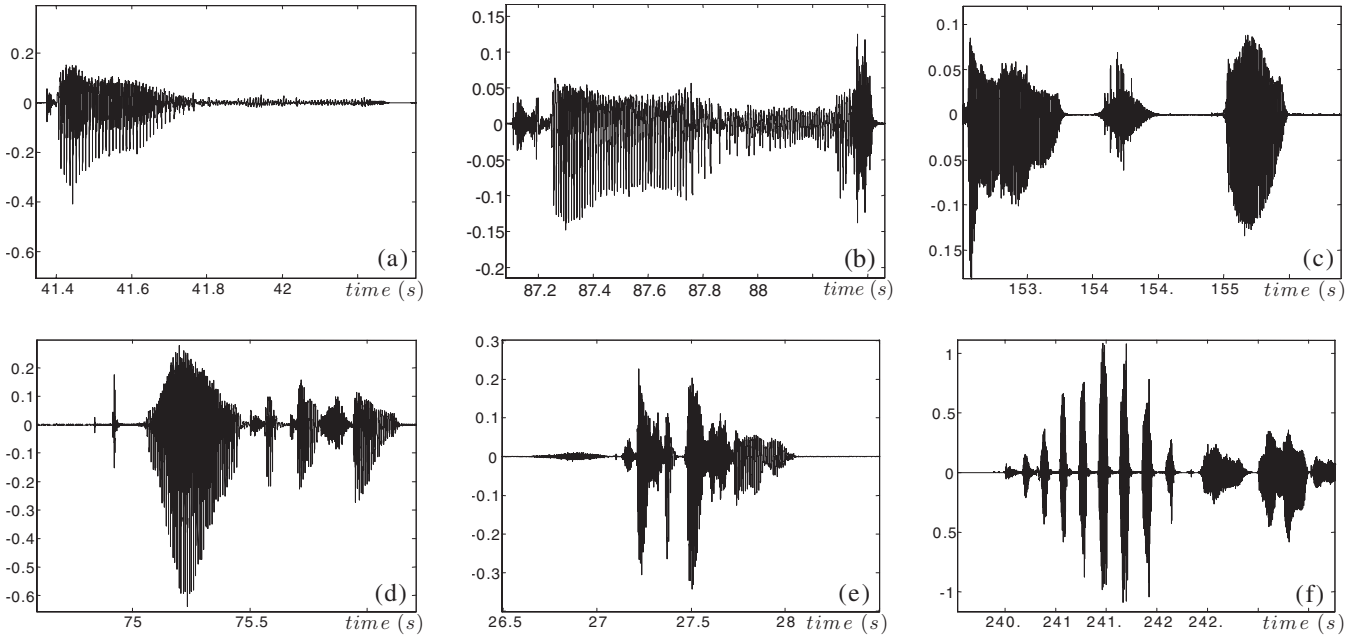


FIG. 4. Examples of sounds present in the spontaneous speech corpus. [(a) and (b)] typical hesitation sound in French ["euh," a long [∅]; included in the sequence in (b)]; (c) sound of "Mmmm…;" (d) snap of the lips before speech; (e) respiration intake; (f) laugh.

$$h(t) = \frac{1}{\tau}\sum_{t=0}^{T-1} \exp\left(\frac{-t}{\tau}\right), \tag{6}$$

where τ is the time constant of the filter and T is the number of integrated frames. These two parameters must be adequately chosen so that the filter significantly decreases the influence of isolated and accidental high $\pi(t)$ values in silence sections. On the other hand, the filter should not blur small but significant movements in nonsilence sections. In our study, for the sake of simplicity, the filter length is fixed to $T=100$ samples (or 2 s), and several representative values for τ are tested in Sec. III B (the τ value has the role of a memory factor over the past $\pi(t)$ values: the smaller the τ, the shorter the memory).

Finally, the video-based automatic acoustic silence detection is achieved for each frame by comparing $\rho(t)$ to a threshold ρ_{th} that remains to be determined. Therefore, the problem can be formalized by the following hypotheses:

- H_s: The audio frame belongs to a silence section.
- H_{ns}: The audio frame belongs to a nonsilence section.

Then, the audio frame index will respect the following rule:

$$\rho(t) \underset{H_{ns}}{\overset{H_n}{\lessgtr}} \rho_{\text{th}}. \tag{7}$$

That is, if $\rho(t) < \rho_{\text{th}}$ the frame t is considered as silence, else it is considered as nonsilence. This test is what is here referred to as V-VAD.

III. QUANTITATIVE ASSESSMENT

A. Audio analysis results

The audio processing described in Sec. II B has been applied to the corpus for each speaker (JLS and LG). As mentioned before, each frame (about 120 000 20 ms frames per speaker) was automatically labeled as silence or nonsilence before a systematic manual verification. To illustrate the diversity of the corpus, Fig. 4 shows several audio sequences for both speakers. These examples illustrate the need for a distinction between silence and nonsilence rather than speech versus nonspeech. Some audio sections with a significant amount of energy (nonsilence), e.g., Fig. 4(a) between 41.8 and 42.3 s, Fig. 4(d) between 74.7 and 74.9 s, or Fig. 4(e), between 26.5 and 27.1 s, are not speech but rather grunts or murmurs. Table I presents some quantitative results, derived from the analysis, which provide a characterization of the corpus. The number of frames labeled as silence versus nonsilence is quite close for speakers JLS and LG (51% and 58% of the total corpus, respectively). If a "silence section" is defined as a section composed of contiguous silence frames, and if a "nonsilence section" is defined as a section composed of contiguous nonsilence frames, 691 silence sections and 695 nonsilence sections are obtained for speaker JLS, and 603 silence sections and 607 nonsilence sections are obtained for speaker LG, with respective average time lengths of 1.73 and 1.93 s for the first speaker and 2.55 and 1.85 s for the second one. The corresponding standard deviations are quite high (the section length ranges from one to more than 2000 frames, that is, 40 s), illustrating the diversity of dialog situations. Figure 5 shows the duration histograms of silence and nonsilence sections. In both cases, more than 90% of the sections have a duration lower than 4 s.

B. Video characterization of silence versus nonsilence

For each speaker, the labial parameters $l_w(t)$ and $l_h(t)$ were smoothed with the preprocessing described in Sec. II C. Figure 6 shows the results of this process. It can be seen that

TABLE I. Characteristics of the audiovisual corpus processed in this study. The frame size is 20 ms. The data in this table are derived from the semiautomatic audio process of Ramirez *et al.* (2004), with manual verification.

		JLS		LG	
Number of silence sections		695		603	
	Mean duration (s)	1.73		2.55	
	Standard deviation of duration (s)	2.13		3.49	
	Minimum duration (s)	0.02		0.04	
	Maximum duration (s)	22.98		41.98	
Number of nonsilence sections		691		607	
	Mean duration (s)	1.93		1.85	
	Standard deviation of duration (s)	2.08		1.85	
	Minimum duration (s)	0.02		0.02	
	Maximum duration (s)	16.7		12.8	
N	Total number of frames	119 996		119 996	
N_s	Number of silence frames	61 373	(51% of N)	69 162	(58% of N)
N_{ns}	Number of nonsilence frames	58 623	(49% of N)	50 834	(42% of N)
N_z	Number of frames with $\tilde{l}_w(t)$ and $\tilde{l}_h(t)$ null	22 658	(19% of N)	26 249	(22% of N)
N_{zns}	Number of nonsilence frames with $\tilde{l}_w(t)$ and $\tilde{l}_h(t)$ null	5915	(10% of N_{ns})	4908	(10% of N_{ns})
N_{zs}	Number of silence frames with $\tilde{l}_w(t)$ and $\tilde{l}_h(t)$ null	16 743	(27% of N_s)	21 341	(31% of N_s)

the adaptive spline filter efficiently removes the measurement noise: slowly varying sections seem correctly smoothed, whereas fast parameter variations in highly nonstationary sections are preserved. Figure 7 shows the distribution of the resulting lip parameters for both speakers, separately for the audio silence frames and the nonsilence frames. First, differences between the distributions for the two speakers can be noticed. These differences are simply due to interindividual differences in lip shapes and gestures. Despite these differences, the two distributions have similar shapes in the nonsilence context [Figs. 7(a) and 7(c)]. For each speaker, the resulting organization of the labial space is classical for speech configurations (Benoît *et al.*, 1992; Robert-Ribes *et al.*, 1998), assuming that the additional nonspeech gestures do not smear the global trends. For example, we can distinguish closed lip shapes [$\tilde{l}_w(t)=0$ and $\tilde{l}_h(t)=0$] corresponding to bilabials in any vocalic context, rounded lip shapes (e.g., [y], [u], at around $\tilde{l}_w(t)=2$ cm and $\tilde{l}_h(t)=0.25$ cm, and consonants in rounded contexts), spread lip shapes (e.g., [i], at around $\tilde{l}_h(t)=3.5$ cm and $\tilde{l}_h(t)=0.6$ cm, and consonants in spread contexts), and open lip shapes (e.g., [a], at around $\tilde{l}_w(t)=3.5$ cm and $\tilde{l}_h(t)=1$ cm, see also Fig. 3, and consonants in open contexts). Notice that closed lip shapes represent 10% of nonsilence frames for both speakers (see Table I). This is a typical example of the difficulty to associate a given lip shape to a given audio class: in this specific case, a speaker actually spoke or emitted sounds with his mouth shut (during short periods). Now, let us con-

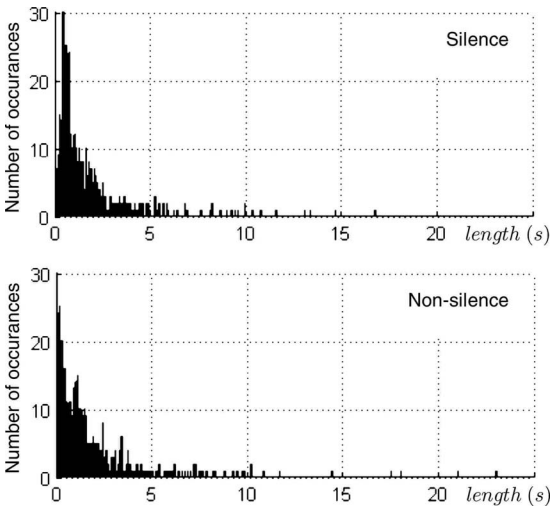


FIG. 5. Histograms of the time length (in seconds) of (top) silence sections and (bottom) nonsilence sections for speaker LG.

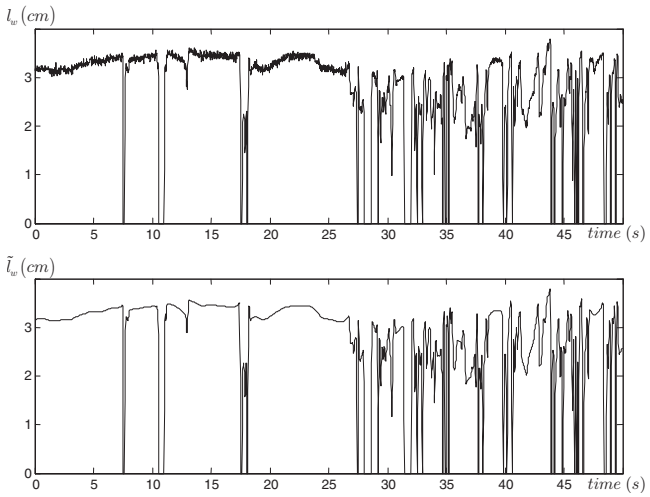


FIG. 6. A lip width parameter trajectory filtered with the adaptive spline technique. Top: raw parameter; bottom: smoothed parameter. The slowly varying sections are efficiently smoothed, while the abrupt changes are preserved.

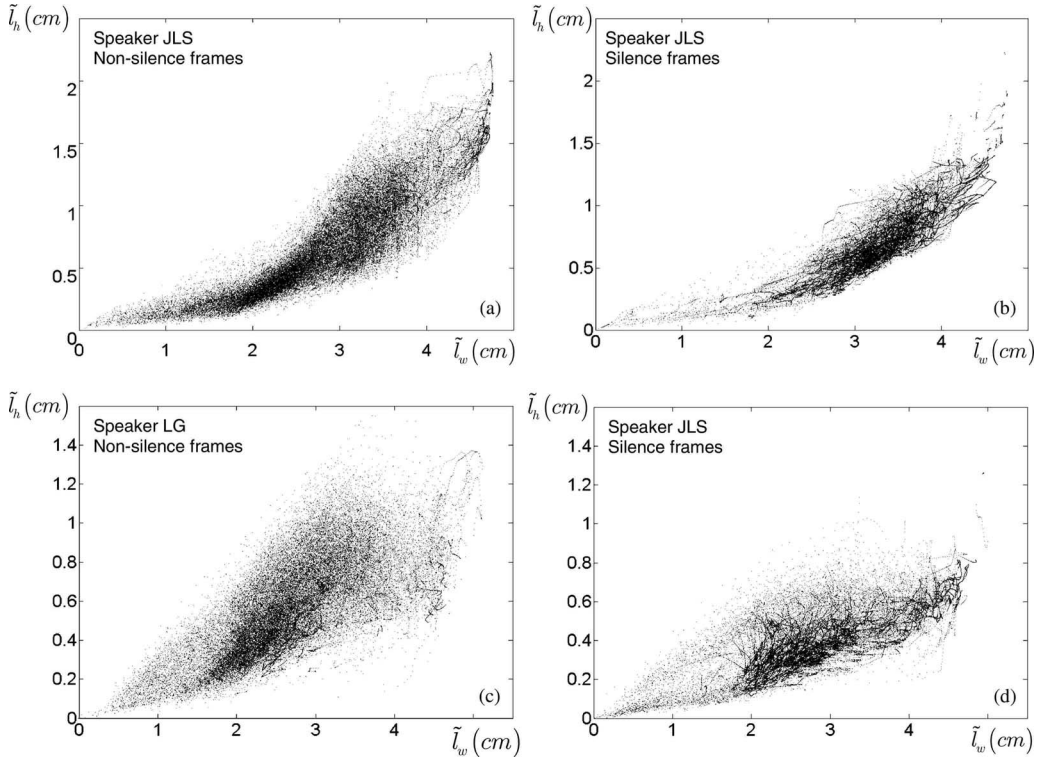


FIG. 7. Distribution of the visual parameters for the two speakers JLS [top: (a) and (b)] and LG [bottom: (c) and (d)] and for the nonsilence frames [left: (a) and (c)] and silence frames [right: (b) and (d)].

sider the visual parameter distribution associated with audio silence, in Figs. 7(b) and 7(d). These figures show that an important subset of visual parameters corresponding to silence frames is located in a subregion within the general set of speech shapes displayed in Figs. 7(a) and 7(c). Besides, another important subset of lip configurations is grouped around the origin, which corresponds to closed lips. Table I, however, shows that closed lip shapes represent only 27%–30% of the lip shapes associated with silence frames. This is much more than the 10% proportion in nonsilence frames but quite far from the totality of silence frames. Altogether, it appears that closed lip shapes are present in both distributions and thus cannot be systematically associated with a silence frame. More generally, since most values of the distribution of static visual parameters $[\tilde{l}_w(t),\tilde{l}_h(t)]$ associated with either silence frames or nonsilence frames are located in the same region, this information is not sufficient to characterize audio silence versus nonsilence. This confirms the need for a dynamic characterization of lip gestures.

A first illustration of this is given in Fig. 8, which provides the same plots as in Fig. 7, but for the derivatives of the parameters (on a log scale for a better concentration of the values). We can see that although still overlapping, the silence and nonsilence distributions are globally much better separated than previously, with the distributions for nonsilence frames being concentrated in higher parameter values than for silence frames. Also, the differences in the distributions between the two speakers seem to be much smaller in this case than in the static case for both silence and nonsilence frames.

Figure 9 displays the distribution (here as an histogram) of the dynamic parameter $\rho(t)$ for the entire corpus respectively for speaker JLS (left column) and speaker LG (right column) and for four values of the time constant τ corresponding to the summation of one frame (that is no actual temporal summation), five frames (100 ms integration), ten frames (200 ms), and 100 frames (2 s). The underlying goal is to tune the temporal integration window so that the distributions of $\rho(t)$ corresponding to the silence sections (the histogram plotted in black in Fig. 9) and to the nonsilence sections (the histogram plotted in white) are as separate as possible. Each of these two distributions is grossly distributed among two classes: the first one is a peak on the left part of the figure corresponding to no lip movement (including, of course, stable closed lips), and the second one is a kernel on the right part of the figure corresponding to the presence of lip movements. The two kernels associated with silence frames (plotted in black) and nonsilence frames (plotted in white) are centered on different locations, the nonsilence kernel being to the right of the silence kernel. This confirms that nonsilence sections are generally associated with larger/faster movements of the lips than silence sections. However, the two kernels are strongly overlapping for the one-frame integration, as shown in Figs. 9(a) and 9(e) since short lip movements can occur during audio silences. Furthermore, the distribution peak associated with stable closed lips on the left part of these figures contains a large contribution of nonsilence frames since short stable lip shapes can occur during speech/sound activity. An optimal temporal integration window is required, which should provide the best separation of

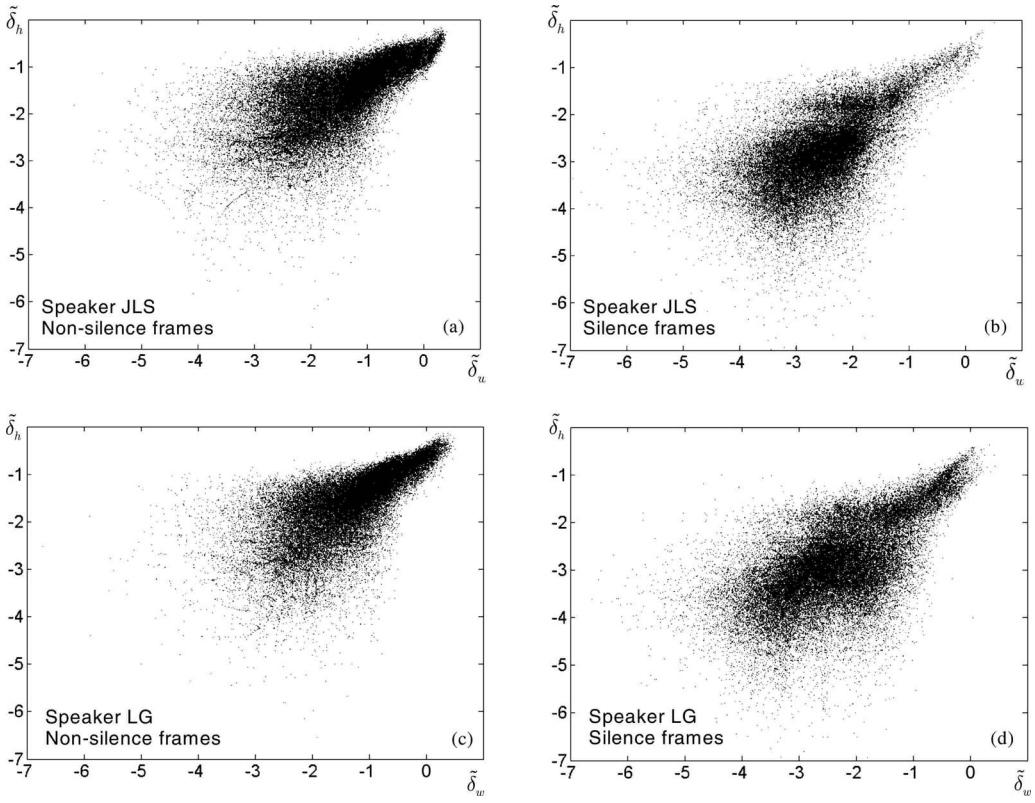


FIG. 8. Distribution of the (absolute values of the) derivatives of the lip parameters (on a log scale: $\tilde{\delta}_h = \log_{10}|\partial \tilde{l}_h / \partial t|$ and $\tilde{\delta}_w = \log_{10}|\partial \tilde{l}_w / \partial t|$) for the two speakers JLS [top: (a) and (b)] and LG [bottom: (c) and (d)] and for the nonsilence frames [left: (a) and (c)] and silence frames [right: (b) and (d)].

these kernels, while reducing the proportion of no-movement values associated with nonsilence frames. Too large a time constant [as in Figs. 9(d) and 9(h)], while successfully addressing this last point, mixes the silence and nonsilence kernels too much, losing the discrimination between silence and nonsilence audio frames for moving lips. However, the histograms plotted in Figs. 9(c) and 9(g) show that a suitable time summation around five to ten frames (100–200 ms) can largely improve the discrimination between silence and nonsilence sections (actually the optimal value is likely to be closer to 5 than to 10): in this case, the white portion of the peak at the origin is quite small, and the black and white kernels are better separated than in the other configurations. Notice finally that the dynamic parameter $\rho(t)$ provides less difference between speakers than the static labial parameters, as was already observed in Fig. 8. This could be important for a future multispeaker application.

C. Automatic video-based silence detection

The proposed V-VAD of Sec. II D was tested on the 120 000 frames of the corpus and for the different settings of the time integrations: 1 frame (instantaneous case) and 5, 10, 20, and 100 frames. In each case, the results of automatic silence frame detection using the V-VAD were compared with the reference labels provided by the acoustic semiautomatic identification process presented in Sec. II B. This test has been done for each speaker.

Figure 10 shows an example of silence detection. This figure represents the time trajectory of the lip parameters $\tilde{l}_w(t)$ and $\tilde{l}_h(t)$ [Figs. 10(a) and 10(b)], of their respective derivatives [Figs. 10(c) and 10(d)], and of the dynamic parameters $\pi(t)$ and $\rho(t)$ with their corresponding detection thresholds [Figs. 10(e) and 10(f)] for about 7 s of signal produced by speaker JLS. Figure 10(g) represents the corresponding speech waveform with the detected and reference silence regions. This figure illustrates the different possible relations between visual and acoustic data: movement of the lips in nonsilence (e.g., from 29.7 to 30.6 s) and in silence (e.g., just before 31.5 s or between 32 and 32.3 s), nonmovement of the lips in silence with opened lips (e.g., from 31.2 to 31.4 s) and closed lips (from 31.5 to 31.9 s), and nonmovement in nonsilence (from 30.9 to 31.1 s). The V-VAD, adequately tuned (τ=20), performs quite well. The silence section of this sequence has been detected. Obviously, the V-VAD fails to avoid a false detection between 31 and 31.2 s, but this is a tough configuration: part of this mistakenly detected section is a long nonsilence section with still lip shape, corresponding to a drawling sentence ending. Moreover, the V-VAD has shrunk the actual silence section. But, on the other hand, it discards several possible false detections in the speech section between 32.5 and 36 s in spite of both closed lip sections and small movements in some regions.

More general results are presented in Fig. 11 as ROC. These curves represent the percentage of correct silence detection (defined as the ratio between the number of detected silence frames and the actual number of silence frames) as a function of the percentage of false silence detection (defined as the ratio between the number of nonsilence frames detected as silence frames and the actual number of nonsilence

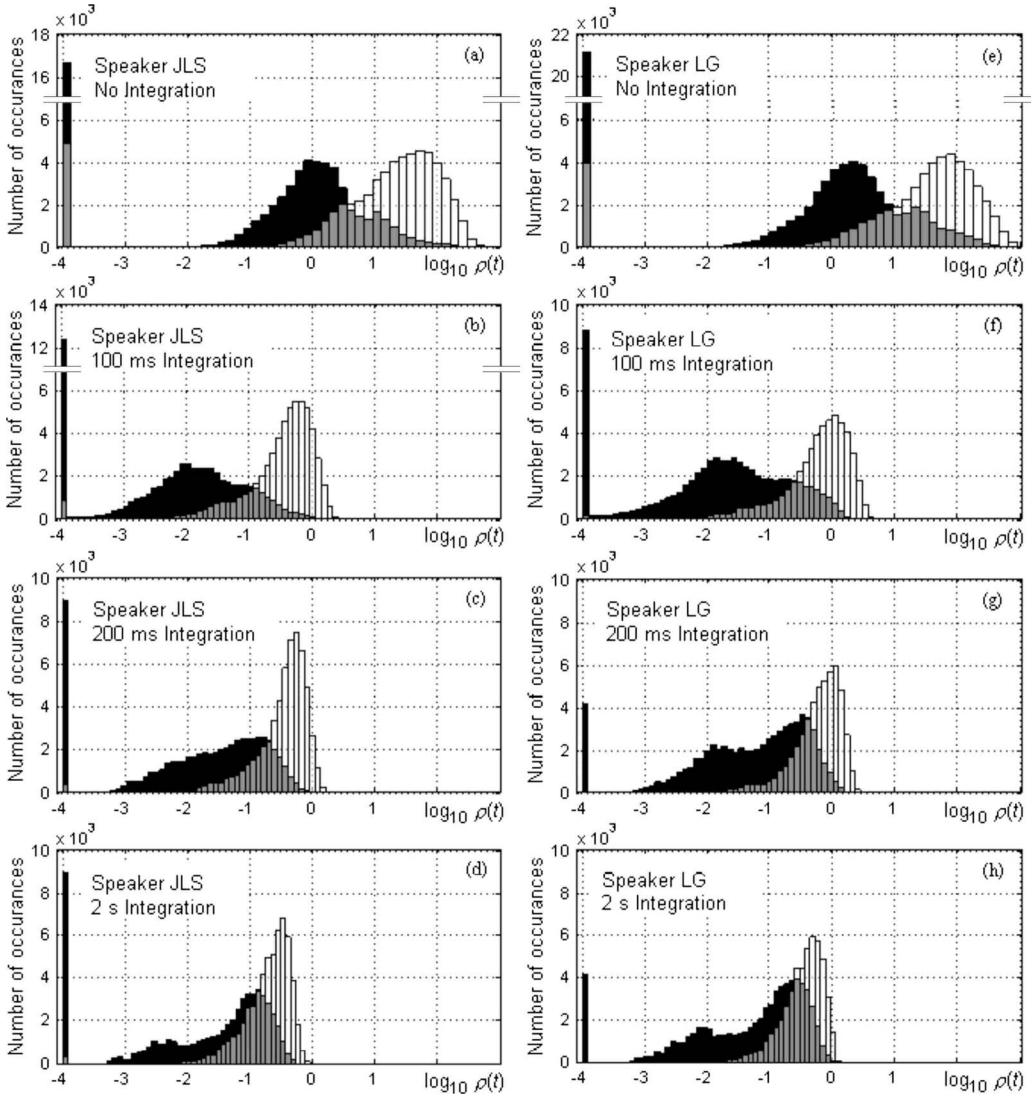


FIG. 9. Distribution of $\log_{10}(\rho(t))$ for the two speakers JLS (left column) and LG (right column) and for different configurations of the time integration. Note that the value $\rho(t)=0$ (no movement) has been arbitrarily fixed to 10^{-4} for visualization of the origin.

frames). To obtain those curves, the threshold ρ_{th} was varied between the minimum and the maximum of $\rho(t)$ (however, when using the V-VAD, one would set ρ_{th} to a fixed value, ensuring a good trade-off between hit rate and false alarm, possibly using the ROC curves as charts). It can be seen from those curves that the benefit of low-pass filtering the parameter $\rho(t)$ is significant. By decreasing the influence of short stable periods in actual speech or sound production, it enables the false silence detection ratio to be decreased significantly. Symmetrically, by decreasing the influence of short/small lip movements in silence, it improves the silence detection ratio. The time integration must be set carefully. When no time integration is performed, the false silence detection scores are moderate (e.g., the point 20%–80% for speaker JLS and 22%–80% for speaker LG). On the contrary, too large a time integration ($\tau=100$ frames corresponding to 2 s) dramatically decreases the silence detection ratio. Finally, the ROC performances are significantly improved with suitable time integration. For instance, using $\tau=$five frames (corresponding to 100 ms) efficiently decreases the false silence detection ratio without decreasing the silence detection ratio: ROC scores of 12%–80% and 15%–80% are obtained for speaker JLS and speaker LG, respectively.

As a complementary result, Fig. 12 shows the ROC curves obtained when $l_w(t)$ and $l_h(t)$ are used in Eq. (4), i.e., unfiltered visual parameters, instead of $\tilde{l}_w(t)$ and $\tilde{l}_h(t)$, to compute $\rho(t)$ with Eq. (5). In this case, lower performances are obtained, which confirms the importance of the preprocessing. Moreover, the role of integration is more important in this case because it also reduces the influence of the measurement noise coming from the lip parameter extraction system. This explains that the difference between the results of Figs. 11 and 12 is particularly important if no integration is performed (e.g., 37%–80% in the no-integration case compared to 17%–80% with adequate integration). The results with temporal integration are quite close with or without preprocessing for speaker JLS, although they are better with the preprocessing than without the preprocessing for speaker LG. This seems to be due to greater measurement noise for this last speaker.

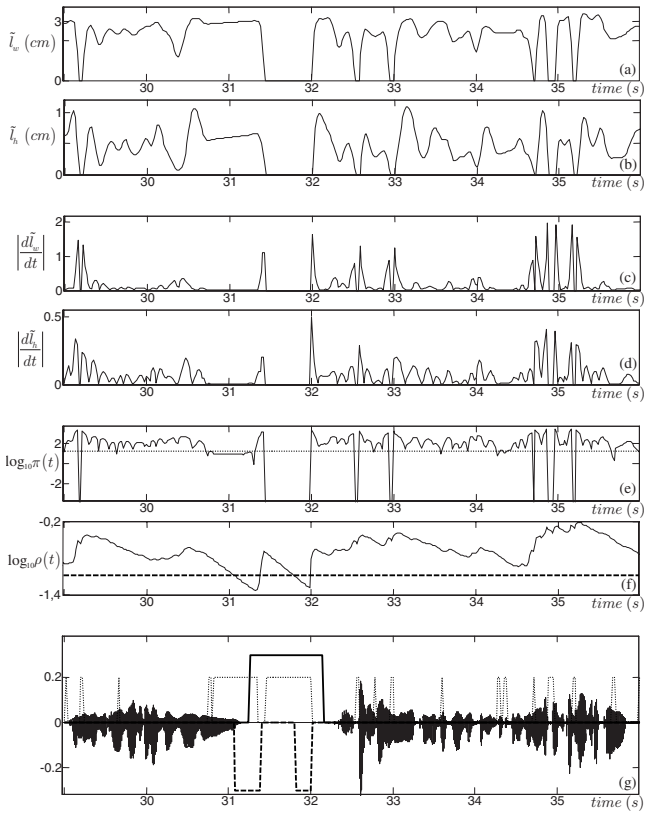


FIG. 10. Silence detection on a sequence of the recorded corpus. [(a) and (b)] Static lip parameters $\tilde{l}_w(t)$ and $\tilde{l}_h(t)$. [(c) and (d)] Their derivatives (absolute values). [(e) and (f)] Instantaneous detection parameter $\pi(t)$ and integrated detection parameter $\rho(t)$ (for τ=20 frames=400 ms) on a log-scale; the dotted and dashed lines are, respectively, the threshold for $\pi(t)$ and for $\rho(t)$. (g) Acoustic signal with silence reference (solid line), frames detected as silence using $\pi(t)$ (dotted line), and frames detected as silence using $\rho(t)$ (dashed line).

IV. CONCLUSION

This paper had two objectives. The first one was to describe the recording and processing of an audiovisual corpus in natural interaction situations. The second objective was to use this corpus to characterize the visual information provided by a speaker's lips during the different dialog phases, with a particular focus on silence sections. An automatic simple and efficient visual voice activity detector was derived from this analysis.

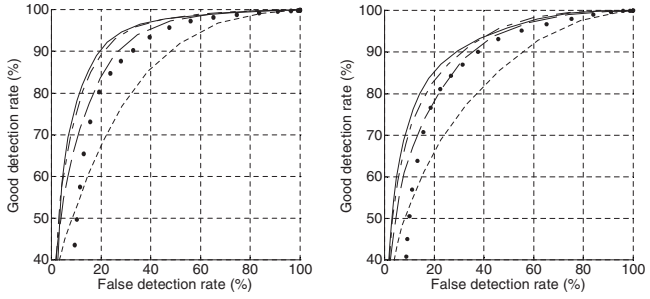


FIG. 11. ROC silence detection curves for the two speakers JLS (left) and LG (right). For each speaker, five integration durations of the visual parameter $\rho(t)$ are used: No integration (dotted line), 100 ms (τ=5, solid line), 200 ms (τ=10, dash-dot line), 400 ms (τ=20, dashed line), and 2 s (τ=100, small dashed line).

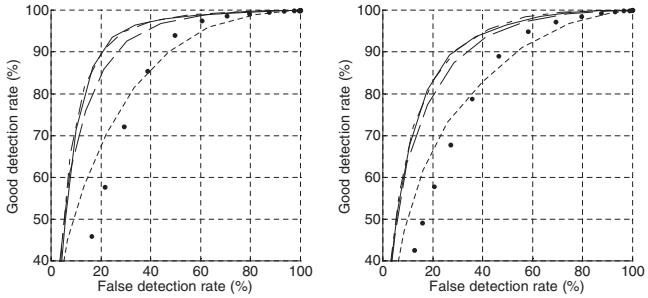


FIG. 12. ROC silence detection curves for the two speakers JLS (left) and LG (right). Here, the visual parameter $\rho(t)$ has been computed [using Eq. (5)] with unfiltered lip parameters l_h and l_w in Eq. (4). For each speaker, five integration durations of the visual parameter $\rho(t)$ are used: No integration (dotted line), 100 ms (τ=5, solid line), 200 ms (τ=10, dash-dot line), 400 ms (τ=20, dashed line), and 2 s (τ=100, small dashed line).

Regarding the first objective, let us recall that the corpus contains about 40 min of signal, providing a rich set of audiovisual data for two speakers in a realistic situation of spontaneous dialog (in French). This corpus is dedicated to fundamental studies in speech and language sciences, as well as to the assessment of audiovisual speech processing systems. The design of such a corpus is not a straightforward task. It requires specific recording equipment and protocol. In addition, as was pointed out in this paper, the preprocessing of the video data is not trivial (although it can be easily implemented after adequate settings). This corpus can be downloaded free of charge from http://www.icp.inpg.fr, assuming it is used for scientific/nonprofit purposes.

Regarding the second objective, the results show that the instantaneous lip shapes in silence and nonsilence frames are largely overlapping. Consequently, such straightforward information cannot be efficiently used for silence versus nonsilence automatic classification of speech sequences. In contrast, lip movements can provide adequate information: A single dynamical parameter processed with suitable temporal integration and threshold has been shown to be appropriate for efficient silence (versus nonsilence) detection. The detection scores have shown that the resulting V-VAD (actually a visual *silence* detector) can be exploitable in real speech processing applications such as enhancement, source separation, and recognition in noise, with, e.g., a 12% false alarm rate versus an 80% hit rate. It is of primary importance to remember that these performance scores are completely independent of the acoustic environment, a property that is not ensured by classical acoustic VAD. Note finally that in the perspective of a "real world" implementation, the blue make-up used for labial information extraction is not a limitation of the proposed method. In a recent study (Aubrey *et al.*, 2007), it has been shown that the dynamic information provided by Eq. (6) is equivalent (in terms of detection scores) to the information provided by a retina model applied on raw black and white images of the lip region, with natural lips (i.e., without make-up).

Further investigations will be conducted to increase the V-VAD performance. They could incorporate an adaptive decision threshold, taking into account the image quality and/or the interspeaker variability. Another perspective is to use both video and audio information together to increase the

detection performance, either taking a decision from a fusion of the decisions provided independently by audio and video information or using both sources of information to feed a single decision process. This would lead to the design of an audiovisual VAD, which seems to us an important outcome for future developments in audiovisual speech processing systems. The V-VAD that has been presented in this study provides a good basis for such development.

[1]Note that this explains why all throughout the paper we consider the distinction between silence sections and nonsilence sections (including speech and nonspeech audible events) rather than the distinction between speech and nonspeech (including silence and nonspeech audible events). Accordingly, the term *voice activity* is to be understood as covering speech and nonspeech audible events (while voice *inactivity* would correspond to silence). The term VAD is a usual denomination in the speech processing literature.

[2]Yet a dependence can be found by considering "the Lombard effect" (Lombard, 1911; Lane and Tranel, 1971): The speaker may increase his/her articulatory efforts (and thus modify the speech characteristics) to improve communication efficiency in noise. This does not reduce the interest of the visual speech information (on the contrary, the movements of the visible articulators may be exaggerated by the Lombard effect).

[3]These authors prefer to classify between speech and nonspeech sections rather than between silence and nonsilence sections as we do even if it seems less appropriate for use in enhancement/separation applications.

[4]Actually, it is not applied in regions where the parameters are equal to zero, or more specifically, the zero value in those regions is not modified since (i) the zero signal is not noisy, and (ii) this avoids unwanted oscillations or overshoots of the spline-filtered parameters after fast lip closing or before fast lip opening regions. In practice, implementing this precaution is a trivial task.

Abrard, F., and Deville, Y. (**2003**). "Blind separation of dependent sources using the "time-frequency ratio of mixture" approach," in Proceedings of the International Symposium on Signal Processing and Its Applications (ISSPA), Paris, France, pp. 81–84.

Abry, C., and Boë, L. J. (**1986**). "Laws for lips," Speech Commun. **5**, 97–104.

Aubrey, A., Rivet, B., Hicks, Y., Girin, L., Chambers, J., and Jutten, C. (**2007**). "Comparison of appearance models and retinal filtering for visual voice activity detection," in Proceedings of the European Signal Processing Conference (EUSIPCO), Poznan, Poland.

Bailly, G., and Badin, P. (**2002**). "Seeing tongue movements from outside," in Proceedings of the International Conference on Spoken Language Processing (ICSLP), Denver, CO, pp. 1913–1916.

Bailly, G., Berard, M., Elisei, F., and Odisio, M. (**2003**). "Audiovisual speech synthesis," Speech Technol., **6**, 331–346.

Barker, J. P., and Berthommier, F. (**1999**). "Estimation of speech acoustics from visual speech features: A comparison of linear and non-linear models," in Proceedings of the Conference on Audio-Visual Speech Processing (AVSP), Santa Cruz, CA, pp. 112–117.

Benoît, C., Guiard-Marigny, T., Le Goff, B., and Adjoudani, A. (**1996**). "Which components of the face humans and machines best speechread?" in *Speechreading by Man and Machine: Models, Systems and Applications*, NATO Advanced Studies Institute, Series F: Computer and System Sciences, edited by D. G. Stork and M. E. Hennecke (Springer, New York), pp. 315–328.

Benoît, C., Lallouache, T., Mohamadi, T., and Abry, C. (**1992**). "A set of French visemes for visual speech synthesis," in *Talking Machines: Thèories, Models, and Designs*, edited by G. Bailly, C. Benoit, and T. R. Sawallis (North-Holland, Amsterdam), pp. 485–504.

Benoît, C., Mohamadi, T., and Kandel, S. (**1994**). "Effects of phonetic context on audio-visual intelligibility of French," J. Speech Hear. Res. **37**, 1195–1293.

Bernstein, L. E., Takayanagi, S., and Auer, E. T., Jr. (**2004**). "Auditory speech detection in noise enhanced by lipreading," Speech Commun. **44**, 5–18.

Bertelson, P. (**1999**). "Ventriloquism: A case of crossmodal perceptual grouping," in *Cognitive Contributions to the Perception of Spatial and Temporal Events*, edited by G. Aschersleben, T. Bachmann, and J. Müsseler (Elsevier, Amsterdam), pp. 347–362.

Calvert, G. A., and Campbell, R. (**2003**). "Reading speech from still and moving faces: The neural substrates of visible speech," J. Cogn Neurosci. **15**, 57–70.

Campbell, N. (**2007**). "Approaches to conversational speech rhythm: Speech activity in two-person telephone dialogues," in Proceedings of the International Congress of Phonetic Sciences (ICPhS), Sarrebrücken, Germany, pp. 343–348.

Cosi, P., Fusaro, A., and Tisato, G. (**2003**). "LUCIA: A new Italian talking-head based on a modified Cohen-Massaro's labial coarticulation model," in Proceedings of the European Conference on Speech Communication and Technology (EuroSpeech), Geneva, Switzerland, pp. 2269–2272.

De Cueto, P., Neti, C., and Senior, A. W. (**2000**). "Audio-visual intent-to-speak detection in human-computer interaction," in Proceedings of the International Conference on Acoustics, Speech, and Signal Processing (ICASSP), Istanbul, Turkey, pp. 2373–2376.

Deligne, S., Potamianos, G., and Neti, C. (**2002**). "Audio-visual speech enhancement with AVCDCN (audiovisual codebook dependent cepstral normalization)," in Proceedings of the International Conference on Spoken Language Processing (ICSLP), Denver, CO, pp. 1449–1452.

Ephraim, Y., and Malah, D. (**1984**). "Speech enhancement using a minimum mean-square error short-time spectral amplitude estimator," IEEE Trans. Acoust., Speech, Signal Process. **32**, 1109–1121.

Erber, N. P. (**1975**). "Auditory-visual perception of speech," J. Speech Hear Disord. **40**, 481–492.

Gibert, G., Bailly, G., Beautemps, D., Elisei, F., and Brun, R. (**2005**). "Analysis and synthesis of three-dimensional movements of the head, face, and of a speaker using cued speech," J. Acoust. Soc. Am. **118**, 1144–1153.

Girin, L. (**2004**). "Joint matrix quantization of face parameters and LPC coefficients for low bit rate audiovisual speech coding," IEEE Trans. Speech Audio Process. **12**, 265–276.

Girin, L., Schwartz, J.-L., and Feng, G. (**2001**). "Audio-visual enhancement of speech noise," J. Acoust. Soc. Am. **109**, 3007–3020.

Goecke, R., and Millar, J. B. (**2003**). "Statistical analysis of relationship between audio and video speech parameters Australian English," in Proceedings of the Conference on Audio-Visual Speech Processing (AVSP), Saint-Jorioz, France, pp. 133–138.

Grant, K. W., and Seitz, P. (**2000**). "The use of visible speech cues for improving auditory detection of spoken sentences," J. Acoust. Soc. Am. **108**, 1197–1208.

Huang, J., Liu, Z., Wang, Y., Chen, Y., and Wong, E. (**1999**). "Integration of multimodal feature for video scene classification based on HMM," in Proceedings of the Workshop Meeting on Multimedia Signal Processing (MMSP), Copenhagen, Denmark, pp. 53–58.

Iyengar, G., and Neti, C. (**2001**). "A vision-based microphone switch for speech intent detection," in Proceedings of the Workshop at the International Conference on Computer Vision (ICCV) on Recognition, Analysis and Tracking of Face and Gestures in Real Time Systems (RATFG-RTS), Vancouver, Canada, pp. 101–105.

Jiang, J., Alwan, A., Keating, P. A., Auer, E. T., and Bernstein, L. E. (**2002**). "On the relationship between face movements, tongue movements and speech acoustics," EURASIP J. Appl. Signal Process. **11**, 1174–1188.

Kim, J., and Davis, C. (**2004**). "Investigating the audio-visual speech detection advantage," Speech Commun. **44**, 19–30.

Lallouache, T. (**1990**). "Un poste visage-parole. Acquisition et traitement des contours labiaux (*A device for the capture and processing of lip contours*)," in Proceedings of the XVIII Journées d'Étude sur la Parole (JEP), Montréal, Canada, pp. 282–286 (in French).

Lane, H., and Tranel, B. (**1971**). "The Lombard sign and the role of hearing in speech," J. Speech Hear. Res. **14**, 677–709.

Le Bouquin-Jeannès, R., and Faucon, G. (**1995**). "Study of a voice activity detector and its influence on a noise reduction system," Speech Commun. **16**, 245–254.

Liu, P., and Wang, Z. (**2004**). "Voice activity detection using visual information," in Proceedings of the International Conference on Acoustics, Speech, and Signal Processing (ICASSP), Montreal, Canada, pp. 609–612.

Lombard, E. (**1911**). "Le signe de l'élévation de la voix (*The sign of voice rise*)," Annales des maladies de l'oreille et du larynx **37**, pp. 101–119, (in French).

Macho, D., Padrell, J., Abad, A., Nadeu, C., Hernando, J., McDonough, J., Wölfel, M., Klee, U., Omologo, M., Brutti, A., Svaizer, P., Potamianos, G., and Chu, S. M. (**2005**). "Automatic speech activity detection, source localization and speech recognition on the CHIL seminar corpus," in In-

ternational Conference on Multimedia and Expo (ICME), Amsterdam, The Netherlands, pp. 876–879.
McGurk, H., and McDonald, J. (**1976**). "Hearing lips and seeing voices," Nature (London) **264**, 746–748.
Munhall, K. G., Gribble, P., Sacco, L., and Ward, M. (**1996**). "Temporal constraints on the McGurk effect," Percept. Psychophys. **58**, 351–362.
Munhall, K. G., Servos, P., Santi, A., and Goodale, M. (**2002**). "Dynamic visual speech perception in a patient with visual form agnosia," NeuroReport **13**, 1793–1796.
Munhall, K. G., and Vatikiotis-Bateson, E. (**1998**). "The moving face during speech communication," in *Hearing by Eye II: Advances in the Psychology of Speechreading and Auditory-Visual Speech*, edited by R. Campbell, B. Dodd, and D. Burnham (Psychology, London), pp. 123–139.
Petajan, E. D. (**1984**). "Automatic lipreading to enhance speech recognition," in Proceedings of the Global Telecommunications Conference (GLOBCOM), Atlanta, GA, pp. 265–272.
Potamianos, G., Neti, C., and Deligne, S. (**2003a**). "Joint audio-visual speech processing for recognition and enhancement," in Proceedings of the Conference on Audio-Visual Speech Processing (AVSP), Saint-Jorioz, France, pp. 95–104.
Potamianos, G., Neti, C., and Gravier, G. (**2003b**). "Recent advances in the automatic recognition of visual speech," Proc. IEEE **91**, 1306–1326.
Ramirez, J., Segura, J. C., Bemtez, C., de la Torre, A., and Rubio, A. (**2004**). "Efficient voice activity detection algorithms using long-term speech information," Speech Commun. **42**, 271–287.
Ramírez, J., Segura, J. C., Benítez, C., García, L., and Rubio, A. (**2005**). "Statistical voice activity detection using a multiple observation likelihood ratio test," IEEE Signal Process. Lett. **12**, 689–692.
Rao, R., and Chen, T. (**1996**). "Cross-modal predictive coding for talking head sequences," in Proceedings of the International Conference on Acoustics, Speech, and Signal Processing (ICASSP), Atlanta, GA, pp. 2058–2061.
Rivet, B., Girin, L., and Jutten, C. (**2007a**). "Mixing audiovisual speech processing and blind source separation for the extraction of speech signals from convolutive mixtures," IEEE Trans. Audio, Speech, Lang. Process. **15**, 96–108.
Rivet, B., Girin, L., and Jutten, C. (**2007b**). "Visual voice activity detection as a help for speech source separation from convolutive mixtures," Speech Commun. **49**, 667–677.
Robert-Ribes, J., Schwartz, J. L., Lallouache, T., and Escudier, P. (**1998**). "Complementary and synergy in bimodal speech: Auditory, visual, and audio-visual identification of French oral vowels in noise," J. Acoust. Soc. Am. **6**, 3677–3689.
Rosenblum, L. D., Johnson, J. A., and Saldana, H. M. (**1996**). "Visual kinematic information for embellishing speech in noise," J. Speech Hear. Res. **39**, 1159–1170.
Rosenblum, L. D., and Saldana, H. M. (**1996**). "An audiovisual test of kinematic primitives for visual speech perception," J. Exp. Psychol. Hum. Percept. Perform. **22**, 318–331.
Schwartz, J. L., Berthommier, F., and Savariaux, C. (**2004**). "Seeing to hear better: Evidence for early audio-visual interactions in speech identification," Cognition **93**, 69–78.
Sodoyer, D., Girin, L., Jutten, C., and Schwartz, J. L. (**2002**). "Separation of audio-visual speech sources: A new approach exploiting the audiovisual coherence of speech stimuli," EURASIP J. Appl. Signal Process. **11**, 1165–1173.
Sodoyer, D., Girin, L., Jutten, C., and Schwartz, J. L. (**2004**). "Further experiments on audio-visual speech source separation," Speech Commun. **44**, 113–125.
Sodoyer, D., Rivet, B., Girin, L., Jutten, C., and Schwartz, J. L. (**2006**). "An analysis of visual speech information applied to voice activity detection," in Proceedings of the International Conference on Acoustics, Speech, and Signal Processing (ICASSP), Toulouse, France, pp. 601–604.
Sohn, J., Kim, N. S., and Sung, W. (**1999**). "A statistical model based voice activity detection," IEEE Signal Process. Lett. **6**, 1–3.
Sumby, W. H., and Pollack, I. (**1954**). "Visual contribution to speech intelligibility in noise," J. Acoust. Soc. Am. **26**, 212–215.
Summerfield, Q. (**1979**). "Use of visual information for phonetic perception," Phonetica **36**, 314–331.
Summerfield, Q. (**1987**). "Some preliminaries to a comprehensive account of audio-visual speech perception," in *Hearing by Eye: The Psychology of Lip-Reading*, edited by B. Dodd and R. Campbell (Erlbaum, London), pp. 3–51.
Tanyer, S. G., and Ozer, H. (**2000**). "Voice activity detection in nonstationary noise," IEEE Trans. Speech Audio Process. **8**, 478–482.
Thomas, S. M., and Jordan, T. R. (**2004**). "Contributions of oral and extraoral facial movement to visual and audiovisual speech perception," J. Exp. Psychol. **30**, 873–888.
Wang, W., Cosker, D., Hicks, Y., Sanei, S., and Chambers, J. A. (**2005**). "Video assisted speech source separation," in Proceedings of the International Conference on Acoustics, Speech, and Signal Processing (ICASSP), Philadelphia, pp. 425–428.
Yehia, H., Kuratate, T., and Vatikiotis-Bateson, E. (**2000**). "Facial animation and head motion driven by speech acoustics," in Proceedings of the Seminar on Speech Production: Models and Data and CREST Workshop on Models of Speech Production: Motor Planning and Articulatory Modelling, Kloster Seeon, Germany, pp. 265–268.
Yehia, H., Rubin, P., and Vatikiotis-Bateson, E. (**1998**). "Quantitative association of vocal-tract and facial behavior," Speech Commun. **26**, 23–43.

The dependencies of phase velocity and dispersion on volume fraction in cancellous-bone-mimicking phantoms

Keith A. Wear[a)]
U.S. Food and Drug Administration, Center for Devices and Radiological Health, HFZ-142, 12720 Twinbrook Parkway, Rockville, Maryland 20852

(Received 29 August 2008; revised 20 November 2008; accepted 22 November 2008)

Frequency-dependent phase velocity was measured in eight cancellous-bone-mimicking phantoms consisting of suspensions of randomly oriented nylon filaments (simulating trabeculae) in a soft-tissue-mimicking medium (simulating marrow). Trabecular thicknesses ranged from 152 to 356 μm. Volume fractions of nylon filament material ranged from 0% to 10%. Phase velocity varied approximately linearly with frequency over the range from 300 to 700 kHz. The increase in phase velocity (compared with phase velocity in a phantom containing no filaments) at 500 kHz was approximately proportional to volume fraction occupied by nylon filaments. The derivative of phase velocity with respect to frequency was negative and exhibited nonlinear, monotonically decreasing dependence on volume fraction. The dependencies of phase velocity and its derivative on volume fraction in these phantoms were similar to those reported in previous studies on (1) human cancellous bone and (2) phantoms consisting of parallel nylon wires immersed in water. [DOI: 10.1121/1.3050310]

PACS number(s): 43.80.Jz, 43.80.Ev [TDM] Pages: 1197–1201

I. INTRODUCTION

Bone sonometry is now an accepted method for diagnosis of osteoporosis (Langton and Njeh, 2004; Laugier, 2008). Speed of sound (SOS) in cancellous bone is highly correlated with bone mineral density (Rossman *et al.*, 1989; Tavakoli and Evans, 1991; Zagzebski *et al.*, 1991; Njeh *et al.*, 1996; Laugier *et al.*, 1997; Nicholson *et al.*, 1998; Hans *et al.*, 1999; Trebacz and Natali, 1999), which is an indicator of systemic osteoporotic fracture risk (Cummings *et al.*, 1993). Calcaneal SOS (in combination with broadband ultrasound attenuation) has been shown to be predictive of hip fractures in women in prospective studies (Hans *et al.*, 1996; Miller *et al.*, 2002; Hans *et al.*, 2004; Huopio *et al.*, 2004; Schott *et al.*, 2005; Krieg *et al.*, 2006).

Despite the clinical utility of the SOS, the mechanisms responsible for variations of the SOS in cancellous bone are not well understood yet. Unlike soft tissues, which typically exhibit positive dispersion (phase velocity increasing with ultrasonic frequency) (O'Donnell *et al.*, 1981), cancellous bone exhibits negative dispersion (Nicholson *et al.*, 1996; Strelitzki and Evans, 1996; Droin *et al.*, 1998; Wear, 2000a, 2000b). This negative dispersion may be explained using a stratified model (Brekhovskikh, 1980; Hughes *et al.*, 1999; Wear, 2001; Lin, 2001), modified Biot–Attenborough theory (Lee *et al.*, 2003), a restricted-bandwidth form of the Kramers–Kronig dispersion relations (Waters and Hoffmeister, 2005), or from the interference of two or more positively dispersive pulses (Marutyan *et al.*, 2006; Marutyan *et al.*, 2007; Bauer *et al.*, 2008; Anderson *et al.*, 2008).

Measurements on cancellous-bone-mimicking phantoms can provide insight into the determinants of phase velocity and dispersion. The present study complements a previous study, which showed that for phantoms consisting of parallel nylon wires immersed in water, phase velocity and dispersion are primarily determined by volume fraction (VF) occupied by nylon wires (Wear, 2005). In the present study, a new phantom design is utilized. Water is replaced with soft-tissue-mimicking material, which is a more realistic surrogate for marrow than water. Also, in the present study, parallel nylon wires are replaced with randomly oriented nylon filaments. In cancellous bone, the orientation of trabeculae is somewhere in between these two extremes.

II. METHODS

A. Phantoms

Eight phantoms containing nylon filaments (simulating trabeculae) in proprietary soft-tissue-mimicking material (simulating marrow) (CIRS Inc., Norfolk, VA) were interrogated. Figure 1 shows a picture of a phantom. A reference phantom containing only soft-tissue-mimicking material was also interrogated. Table I shows the phantom properties. Three of the phantoms contained nylon filaments with diameter equal to 152 μm, which is reasonably close to the mean trabecular thickness in human calcaneus, 127 μm (Ulrich *et al.*, 1999). The dimensions for all phantoms were $80 \times 60 \times 25$ mm^3. The scanning window dimensions for all phantoms were 60×50 mm^2.

Nylon is a useful surrogate for cancellous bone material. The longitudinal sound speed in nylon (2600 m/s) is near the low end of the range reported for mineralized bone material (2800–4000 m/s, near 500 kHz) (Duck, 1990). The dependencies of phase velocity and dispersion on VF in phantoms consisting of parallel nylon wires in water are similar to those in human cancellous bone (Wear, 2005). Nylon wires exhibit frequency-dependent scattering similar to that exhibited by cancellous bone (Wear, 1999, 2004).

[a)]Electronic mail: keith.wear@fda.hhs.gov

FIG. 1. (Color online) A phantom containing nylon filaments.

A previously reported phantom design, consisting of cubic granules of gelatin immersed in epoxy, has been shown to be useful for the prediction of the dependences of phase velocity, dispersion, and attenuation on porosity of cancellous bone (Clarke *et al.*, 1994; Strelitzki *et al.*, 1997).

B. Ultrasonic methods

A Panametrics (Waltham, MA) 5800 pulser/receiver was used. Samples were interrogated in through-transmission in a water tank using a pair of coaxially aligned, Panametrics 500 kHz, broadband, 0.75 in. (1.9 cm) diameter, 1.5 in. (3.8 cm) focal length transducers. The propagation path between transducers was twice the focal length. Received radio frequency signals were digitized (8 bits, 10 MHz) using a LeCroy (Chestnut Ridge, NY) 9310C Dual 400 MHz oscilloscope and stored on a computer [via general purpose interface bus (GPIB)] for off-line analysis. The transducers were maintained in constant positions. Each phantom was suspended in the water tank by a two-dimensional stage that enabled movement of the phantom in the two dimensions perpendicular to the beam propagation direction. Attenuation measurements were made at 15 positions throughout each phantom scanning window corresponding to a 3×5 grid in which neighboring measurements were separated by 1 cm.

Frequency-dependent phase velocity, $c_p(f)$, was computed using

TABLE I. Properties of phantoms.

Filament diameter (μm)	Filament length (mm)	Filament number density (No./cc)	Volume fraction (%)
⋯	⋯	0	0
152	10	100	1.8
203	10	100	3.2
229	10	100	4.1
330	10	100	8.5
356	10	100	9.9
152	12	100	2.2
229	12	100	3.3
152	12	200	4.4

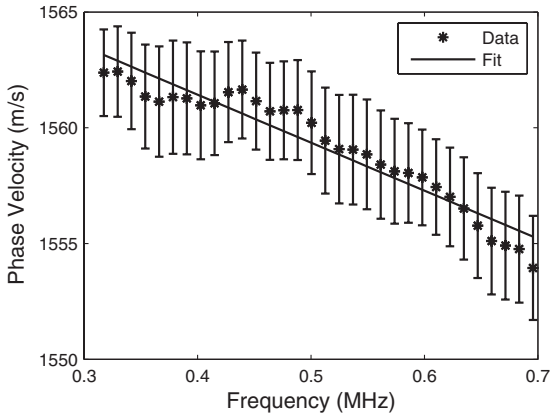


FIG. 2. Measurements of phase velocity vs frequency for one phantom. A linear fit is also shown. The error bars denote standard errors.

$$c_p(f) = \frac{c_w}{1 + \dfrac{c_w \Delta\phi(f)}{2\pi f d}}, \tag{1}$$

where f is frequency, $\Delta\phi(f)$ is the difference in unwrapped phases (see next paragraph) of the received signals with and without the phantom in the water path, d is the phantom thickness (2.5 cm), and c_w is the temperature-dependent SOS in distilled water given by (Kaye and Laby, 1973)

$$c_w = 1402.9 + 4.835T - 0.047\,016T^2 + 0.000\,127\,25T^3 \text{ m/s}, \tag{2}$$

where T is the temperature in °C. Temperature, measured with a digital thermometer, was about 20 °C for these measurements, which meant that c_w was about 1482 m/s.

The unwrapped phase difference, $\Delta\phi(f)$, was computed as follows. Fast Fourier transforms (FFTs) of the digitized received signals were taken. The phase of the signal at each frequency was taken to be the inverse tangent of the ratio of the imaginary to real parts of the FFT at that frequency. Since the inverse tangent function yields principal values between $-\pi$ and π, the phase had to be unwrapped by adding an integer multiple of 2π to all frequencies above each frequency where a discontinuity appeared.

Dispersion was characterized by the slope, dc_p/df, of a linear least-squares regression fit of $c_p(f)$ versus f over the range from 300 to 700 kHz, which roughly corresponded to the system −6 dB bandwidth.

III. RESULTS

Figure 2 shows measurements of phase velocity (c_p) versus frequency (f) for one phantom. Phase velocity declined approximately linearly with frequency for all phantoms.

Figure 3 shows measurements of c_p(500 kHz) versus VF on all the phantoms. A linear fit, $c_p(500\text{ kHz}) = 1530 + 3.3\text{ VF m/s}$ (where VF is expressed as a percentage), is in good agreement with the data.

Figure 4 shows measurements of dc_p/df versus VF for all phantoms. A power law fit, $dc_p/df = 5.6 - 0.18\text{ VF}^{2.1}$ is also shown. Values for dc_p/df ranged from +5 to −22 m/s MHz, which is consistent with values reported in

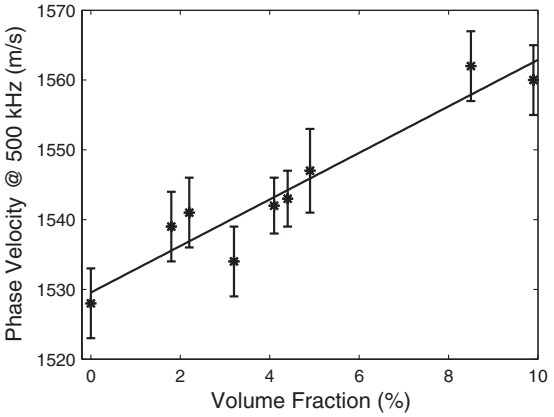


FIG. 3. Average phase velocity at 500 kHz vs VF occupied by nylon filaments. The error bars denote standard deviations.

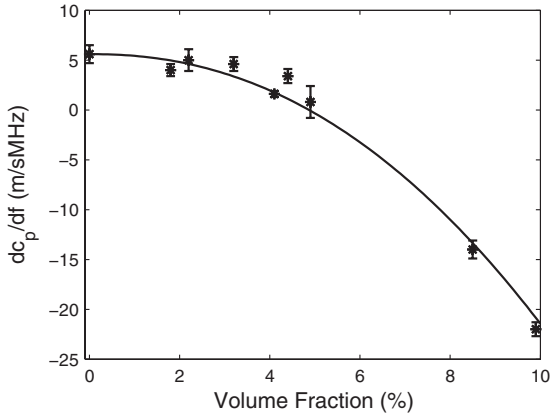


FIG. 4. dc_p/df vs VF occupied by nylon filaments. The error bars denote standard deviations.

human calcaneus *in vitro*. A greater range for dc_p/df has been measured *in vivo*. See Table II.

IV. DISCUSSION

Phase velocity (c_p) in phantoms, consisting of randomly distributed nylon filaments (simulating trabeculae) immersed in soft-tissue-mimicking material (simulating marrow), is an approximately linear, monotonically increasing function of VF. The derivative of phase velocity with respect to frequency, dc_p/df, in these phantoms is a nonlinear, monotonically decreasing function of VF. Both phase velocity and dc_p/df appear to be primarily determined by VF. Since group velocity is determined by c_p and dc_p/df (Duck, 1990; Wear, 2005), then group velocity must also be primarily determined by VF.

A previous study showed similar trends for phase velocity and dc_p/df in phantoms consisting of parallel nylon wires immersed in water (Wear, 2005). In Fig. 5, the change in phase velocity (compared with the zero VF level) is plotted versus VF for both the previous and current phantom designs. In Fig. 6, the change in dc_p/df (compared with the zero VF level) is plotted versus VF for both phantom designs. Despite significant differences in microarchitecture and fluid media for the two types of phantoms, the effect of VF on phase velocity and dc_p/df is remarkably similar. Moreover, it was argued previously that the VF dependencies of phase velocity and dc_p/df in phantoms consisting of parallel nylon wires immersed in water were similar to those reported in cancellous bone *in vitro* (Wear, 2005; Wear *et al.*, 2005). These empirical results, taken collectively, suggest that changes in phase velocity and dc_p/df observed in human cancellous bone are primarily determined by VF.

Nicholson *et al.* (2001) measured phase velocity in 69 human calcaneal cancellous bone cubes. They found that, after the data were adjusted for density (which is a measure of bone quantity rather than microarchitecture), there was no significant dependence of phase velocity on microarchitectural parameters. The dominant role of bone quantity (as opposed to microarchitecture) was also seen in a three-dimensional simulation study by Haïat *et al.* (2007), in which variations in VF of micro-CT reconstructions of human cancellous femur explained 94% of variations in SOS.

In measurements on human cancellous lumbar spine, Hans *et al.* (1999) found the SOS to be approximately 2%–3% higher in the axial direction compared with the sagittal and coronal directions. Since bone VF is identical in all three orientations, these results suggest that the arrangement of trabeculae, not just the quantity of trabecular material (i.e., VF), does play a role in determining the SOS. In these experiments, however, the comparison was between extreme differences in angle between the ultrasound propagation direction and the predominant trabecular direction: approximately parallel (axial) versus approximately perpendicular (sagittal and coronal). Less dramatic variations in trabecular arrangement, such as those in the phantom experiments and

TABLE II. Estimates of the first derivative of phase velocity with respect to frequency, dc_p/df, in human calcaneus from Nicholson *et al.* (1996, Table 1), Strelitzki and Evans (1996, Table 2), Droin *et al.* (1998, Table 1), Wear (2000a, Table 1), and Wear (2007, Table 2). N is the number of calcaneus samples upon which measurements were based.

Author(s)		N	Frequency range (kHz)	Age range (years)	dc_p/df (mean ± standard deviation) (m/s MHz)
Nicholson *et al.* (1996)	*In vitro*	70	200–800	22–76	−40
Strelitzki and Evans (1996)	*In vitro*	10	600–800	Unknown	−32 ± 27
Droin *et al.* (1998)	*In vitro*	15	200–600	69–89	−15 ± 13
Wear (2000a)	*In vitro*	24	200–600	Unknown	−18 ± 15
Wear (2007)	*In vivo*	73	300–600	21–78	−59 ± 52

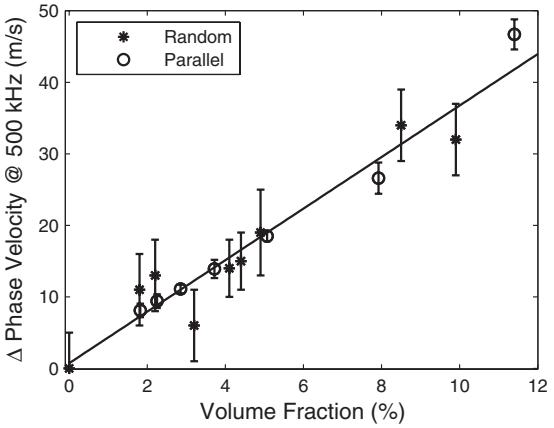


FIG. 5. Change in phase velocity (compared with the zero VF level) vs VF for the present study, which utilized randomly distributed nylon filaments in soft-tissue-mimicking material, and a previous study (Wear, 2005), which utilized phantoms consisting of parallel nylon wires immersed in water. The error bars denote standard deviations.

simulation study described above, may not necessarily produce significant variations in the SOS or phase velocity.

Anisotropy is more pronounced in bovine cancellous bone. Hosokawa and Otani (1998) found fast wave velocity to be approximately 25% greater in the direction parallel to the predominant trabecular orientation compared with other directions in bovine cancellous tibia. Hughes *et al.* (1999) found fast wave velocity to be approximately 100% greater in the parallel direction in bovine cancellous tibia and femur. Hoffmeister *et al.* (2000) found SOS to be approximately 25% higher in the parallel direction in bovine cancellous tibia. Given the dependence of cancellous bone microstructure on loading conditions, and the differences in loading conditions between humans and cows, the enhanced anisotropy in bovine cancellous bone is perhaps not surprising.

V. CONCLUSION

The experiments on phantoms reported here reinforce previous results on phantoms and human cancellous bone *in vitro* in which VF of trabeculae is the dominant determinant of phase velocity. In bovine cancellous bone, however, both VF and trabecular orientation are significant determinants of phase velocity.

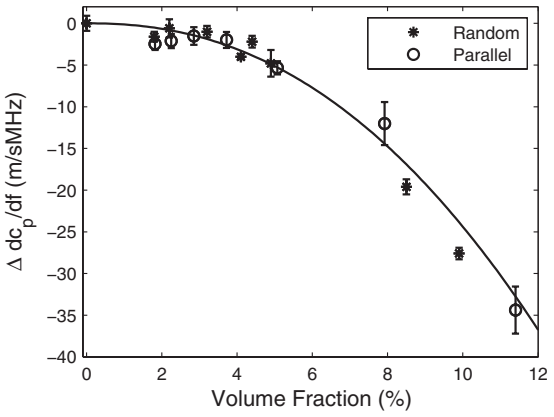


FIG. 6. Change in dc_p/df (compared with the zero VF level) vs VF for the present study, which utilized randomly distributed nylon filaments in soft-tissue-mimicking material, and a previous study (Wear, 2005), which utilized phantoms consisting of parallel nylon wires immersed in water. The error bars denote standard deviations.

ACKNOWLEDGMENTS

The author is grateful to Laura Perfetti and Heather Miller, C.I.R.S., Norfolk, VA, for assistance in phantom design and construction.

Anderson, C. C., Marutyan, K. R., Holland, M. R., Wear, K. A., and Miller, J. G. (**2008**). "Interference between wave modes may contribute to the apparent negative dispersion observed in cancellous bone," J. Acoust. Soc. Am. **124**, 1781–1789.

Bauer, A. Q., Marutyan, K. R., Holland, M. R., and Miller, J. G. (**2008**). "Negative dispersion in bone: The role of interference in measurements of the apparent phase velocity of two temporally overlapping signals," J. Acoust. Soc. Am. **123**, 2407–2414.

Brekhovskikh, L. M. (**1980**). *Waves in Layered Media* (Academic, New York).

Clarke, A. J., Evans, J. A., Truscott, J. G., Milner, R., and Smith, M. A. (**1994**). "A phantom for quantitative ultrasound of trabecular bone," Phys. Med. Biol. **39**, 1677–1687.

Cummings, S. R., Black, D. M., Nevitt, M. C., Browner, W., Cauley, J., Ensrud, K. E., Genant, H. K., Palermo, L., Scott, J., and Vogt, T. M. (**1993**). "Bone density at various sites for prediction of hip fractures," Lancet **341**, 72–75.

Droin, P., Berger, G., and Laugier, P. (**1998**). "Velocity dispersion of acoustic waves in cancellous bone," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **45**, 581–592.

Duck, F. A. (**1990**). *Physical Properties of Tissue* (Cambridge University Press, Cambridge).

Haïat, G., Padilla, F., Peyrin, F., and Laugier, P. (**2007**). "Variation of ultrasonic parameters with microstructure and material properties of trabecular bone: A 3D model simulation," J. Bone Miner. Res. **22**, 665–674.

Hans, D., Dargent-Molina, P., Schott, A. M., Sebert, J. L., Cormier, C., Kotzki, P. O., Delmas, P. D., Pouilles, J. M., Breart, G., and Meunier, P. J. (**1996**). "Ultrasonographic heel measurements to predict hip fracture in elderly women: The EPIDOS prospective study," Lancet **348**, 511–514.

Hans, D., Schott, A. M., Duboeuf, F., Durosier, C., and Meunier, P. J. (**2004**). "Does follow-up duration influence the ultrasound and DXA prediction of hip fracture? The EPIDOS prospective study," Bone (N.Y.) **35**, 357–363.

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.

Hosokawa, A., and Otani, T. (**1998**). "Acoustic anisotropy in bovine cancellous bone," J. Acoust. Soc. Am. **103**, 2718–2722.

Hughes, E. R., Leighton, T. G., Petley, G. W., and White, P. R. (**1999**). "Ultrasonic propagation in cancellous bone: A new stratified model," Ultrasound Med. Biol. **25**, 811–821.

Huopio, J., Kroger, H., Honkanen, R., Jurvelin, J., Saarikoski, S., and Alhava, E. (**2004**). "Calcaneal ultrasound predicts early postmenopausal fractures as well as axial BMD. A prospective study of 422 women," Osteoporosis Int. **15**, 190–195.

Kaye, G. W. C., and Laby, T. H. (**1973**). *Table of Physical and Chemical Constants* (Longman, London).

Krieg, M., Cornuz, J., Ruffieux, C., Melle, G. V., Buche, D., Dambacher, M. A., Hans, D., Hartl, F., Hauselmann, H. J., Kraenzlin, M., Lippuner, K., Neff, M., Pancaldi, P., Rizzoli, R., Tanzi, F., Theiler, R., Tyndall, A., Wimpfheimer, C., and Burckhardt, P. (**2006**). "Prediction of hip fracture risk by quantitative ultrasound in more than 7000 Swiss women ⩾70 years of age: Comparison of three technologically different bone ultrasound devices in the SEMOF study," J. Bone Miner. Res. **21**, 1457–1463.

Langton, C. M. and Njeh, C. F., eds. (**2004**). *The Physical Measurement of Bone* (Institute of Physics, Bristol).

Laugier, P. (**2008**). "Instrumentation for in vivo ultrasonic characterization of bone strength," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **55**,

1179–1196.
Laugier, P., Droin, P., Laval-Jeantet, A. M., and Berger, G. (**1997**). "In vitro assessment of the relationship between acoustic properties and bone mass density of the calcaneus by comparison of ultrasound parametric imaging and quantitative computed tomography," Bone (N.Y.) **20**, 157–165.
Lee, K. I., Roh, H., and Yoon, S. W. (**2003**). "Acoustic wave propagation in bovine cancellous bone: Application of the modified Biot-Attenborough model," J. Acoust. Soc. Am. **114**, 2284–2293.
Lin, W., Qin, Y. X., and Rubin, C. (**2001**). "Ultrasonic wave propagation in trabecular bone predicted by the stratified model," Ann. Biomed. Eng. **29**, 781–790.
Marutyan, K. R., Bretthorst, G. L., and Miller, J. G. (**2007**). "Bayesian estimation of the underlying bone properties from mixed fast and slow mode ultrasonic signals," J. Acoust. Soc. Am. **121**, EL8–EL14.
Marutyan, K. R., Holland, M. R., and Miller, J. G. (**2006**). "Anomalous negative dispersion in bone can result from the interference of fast and slow waves," J. Acoust. Soc. Am. **120**, EL55–EL61.
Miller, P. D., Siris, E. S., Barrett-Connor, E., Faulkner, K. G., Wehren, L. E., Abbott, T. A., Chen, Y., Berger, M. L., Santora, A. C., and Sherwood, L. M. (**2002**). "Prediction of fracture risk in postmenopausal white women with peripheral bone densitometry: Evidence from the national osteoporosis risk assessment," J. Bone Miner. Res. **17**, 2222–2230.
Nicholson, P. H. F., Lowet, G., Langton, C. M., Dequeker, J., and Van der Perre, G. (**1996**). "Comparison of time-domain and frequency-domain approaches to ultrasonic velocity measurements in trabecular bone," Phys. Med. Biol. **41**, 2421–2435.
Nicholson, P. H. F., Muller, R., Cheng, X. G., Ruegsegger, P., Van der Perre, G., Dequeker, J., and Boonen, S. (**2001**). "Quantitative ultrasound and trabecular architecture in the human calcaneus," J. Bone Miner. Res. **16**, 1886–1892.
Nicholson, P. H. F., Muller, R., Lowet, G., Cheng, X. G., Hildebrand, T., Ruegsegger, P., Van Der Perre, G., Dequeker, J., and Boonen, S. (**1998**). "Do quantitative ultrasound measurements reflect structure independently of density in human vertebral cancellous bone?," Bone (N.Y.) **23**, 425–431.
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.
O'Donnell, M., Jaynes, E. T., and Miller, J. G. (**1981**). "Kramers-Kronig relationship between ultrasonic attenuation and phase velocity," J. Acoust. Soc. Am. **69**, 696–701.
Rossman, P., Zagzebski, J., Mesina, C., Sorenson, J., and Mazess, R. (**1989**). "Comparison of speed of sound and ultrasound attenuation in the os calcis to bone density of the radius, femur and lumbar spine," Clin. Phys. Physiol. Meas. **10**, 353–360.'
Schott, M., Hans, D., Duboeuf, F., Dargent-Molina, P., Jajri, T., Breart, G., and Meunier, P. J. (**2005**). "Quantitative ultrasound parameters as well as bone mineral density are better predictors of trochanteric than cervical hip fractures in elderly women. Results from the EPIDOS study," Bone (Osaka) **37**, 858–863.
Strelitzki, R., and Evans, J. A. (**1996**). "On the measurement of the velocity of ultrasound in the os calcis using short pulses," Eur. J. Ultrasound **4**, 205–213.
Strelitzki, R., Evans, J. A., and Clarke, A. J. (**1997**). "The influence of porosity and pore size on the ultrasonic properties of bone investigated using a phantom material," Osteoporosis Int. **7**, 370–375.
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.
Trebacz, H., and Natali, A. (**1999**). "Ultrasound velocity and attenuation in cancellous bone samples from lumbar vertebra and calcaneus," Osteoporosis Int. **9**, 99–105.
Ulrich, D., van Rietbergen, B., Laib, A., and Rüegsegger, P. (**1999**). "The ability of three-dimensional structural indices to reflect mechanical aspects of trabecular bone," Bone (Osaka) **25**, 55–60.
Waters, K. R., and Hoffmeister, B. K. (**2005**). "Kramers-Kronig analysis of attenuation and dispersion in trabecular bone," J. Acoust. Soc. Am. **118**, 3912–3920.
Wear, K. A. (**1999**). "Frequency dependence of ultrasonic backscatter from human trabecular bone: Theory and experiment," J. Acoust. Soc. Am. **106**, 3659–3664.
Wear, K. A. (**2000a**). "Measurements of phase velocity and group velocity in human calcaneus," Ultrasound Med. Biol. **26**, 641–646.
Wear, K. A. (**2000b**). "The effects of frequency-dependent attenuation and dispersion on sound speed measurements: Applications in human trabecular bone," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **47**, 265–273.
Wear, K. A. (**2001**). "A stratified model to predict dispersion in trabecular bone," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **48**, 1079–1083.
Wear, K. A. (**2004**). "Measurement of 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.
Wear, K. A. (**2005**). "The dependences of phase velocity and dispersion on trabecular thickness and spacing in trabecular bone-mimicking phantoms," J. Acoust. Soc. Am. **118**, 1186–1192.
Wear, K. A. (**2007**). "Group velocity, phase velocity, and dispersion in human calcaneus *in vivo*," J. Acoust. Soc. Am. **121**, 2431–2437.
Wear, K. A., Laib, A., Stuber, A. P., and Reynolds, J. C. (**2005**). "Comparison of measurements of phase velocity in human calcaneus to Biot theory," J. Acoust. Soc. Am. **117**, 3319–3324.
Zagzebski, J. A., Rossman, P. J., Mesina, C., Mazess, R. B., and Madsen, E. L. (**1991**). "Ultrasound transmission measurements through the os calcis," Calcif. Tissue Int. **49**, 107–111.

A characterization of Guyana dolphin (*Sotalia guianensis*) whistles from Costa Rica: The importance of broadband recording systems

Laura J. May-Collado[a)]
Department of Environmental Science and Policy, George Mason University, MSN 5F2, 4400 University Drive, Fairfax, Virginia 22030 and Department of Biology, University of Puerto Rico, San Juan, Puerto Rico 00931

Douglas Wartzok[b)]
Department of Biological Sciences, Florida International University, 11200 SW 8th Street, Miami, Florida 33199

(Received 18 July 2008; revised 20 November 2008; accepted 5 December 2008)

Knowledge of the whistle structure in Guyana dolphins comes mostly from Brazilian populations where recordings have been made using limited bandwidth systems (18 and 24 kHz). In Brazil, Guyana dolphin whistle frequency span is 1.34–23.89 kHz, but authors have suggested that limits of their recording system may underestimate frequency span. Whistles of Guyana dolphins from Costa Rica were studied using a broadband recording system. How bandwidth limitations affect the understanding of whistle structure and species classification between sympatric dolphin species was evaluated. In addition, whistles were compared to Brazilian populations. Guyana dolphin whistle frequency span was 1.38 up to 48.40 kHz, greater than previously reported. Bandwidth limitations explained 89% of the whistle variation between studies, and increase in bandwidth improved the whistle classification of Guyana dolphins. Whistle duration and minimum frequency were the most important variables in dolphin species classification. Finally, after accounting for differences in recording systems, Costa Rican Guyana dolphins whistled with significantly higher frequency than Brazilian populations, providing evidence for a postulated increase in frequency from south to north. The study concludes that equipment with an upper frequency limit of at least 50 kHz (150 kHz for harmonics) is required to capture the entire whistle repertoire of the Guyana dolphin.
 [DOI: 10.1121/1.3058631]

PACS number(s): 43.80.Ka [WWA] Pages: 1202–1213

I. INTRODUCTION

The Guyana dolphin (*Sotalia guianensis*) previously considered a marine form of the freshwater Tucuxi dolphin (*Sotalia fluviatilis*) (da Silva and Best, 1996) is today recognized as a separate species based on morphological (Monteiro-Filho *et al.*, 2002) and molecular evidence (Cunha *et al.*, 2005; Caballero *et al.*, 2007). The species occurs in bays, estuaries, river mouths, and shallow coastal waters along the western Atlantic Ocean from Southern Brazil to Northern Nicaragua, and possibly Honduras (da Silva and Best, 1996; Carr and Bonde, 2000; Edwards and Schnell, 2001; Flores, 2002).

Despite the relatively broad distribution of the species most of what is known about its acoustic behavior and biology is from populations along the Brazilian coast from which echolocation clicks, pulsed sounds (e.g., calls and gargles), and whistles have been described (e.g., Wiersma, 1982; Terry, 1983; Monteiro-Filho and Monteiro, 2001; Azevedo and Simão, 2002; Erber and Simão, 2004; Azevedo and Van Sluys, 2005; Rossi-Santos and Podos, 2006). Whistles are the most studied sound type, and several whistle acoustic variables have been recently described from Brazilian populations (Monteiro-Filho and Monteiro, 2001; Azevedo and Simão, 2002; Erber and Simão, 2004; Azevedo and Van Sluys, 2005; Pivari and Rosso, 2005; Rossi-Santos and Podos, 2006). Monteiro-Filho and Monteiro (2001) first described Guyana dolphin whistles as low in frequency (up to 6 kHz) but a more extensive study revealed a much wider whistle frequency range (1.34–23.89 kHz) (Azevedo and Van Sluys, 2005). However, as noted by Azevedo and Van Sluys (2005), some of the recorded whistles looked "cut off" by the upper frequency limit of their recording systems, suggesting Guyana dolphins can emit high frequency whistles exceeding the 24 kHz recording limit. Several toothed whale species have been shown to emit whistles with high fundamental maximum frequencies, up to 24 kHz in spinner dolphins and Atlantic spotted dolphins (e.g., Lammers *et al.*, 1997, 2003; Oswald *et al.*, 2004), 29 and 41 kHz in bottlenose dolphins (Boisseau, 2005; May-Collado and Wartzok, 2008), 35 kHz in white-beaked dolphins (Rasmussen and Miller, 2002; Rasmussen *et al.*, 2006), 24 kHz striped and common dolphins (Oswald *et al.*, 2004), and 48.10 kHz in botos (May-Collado and Wartzok, 2007).

The importance of selecting recording systems with bandwidth appropriate for the study species is fundamental

a)Author to whom correspondence should be addressed. Electronic mail: lmaycollado@gmail.com. URL: delphinids.com
b)Electronic mail: wartzok@fiu.edu

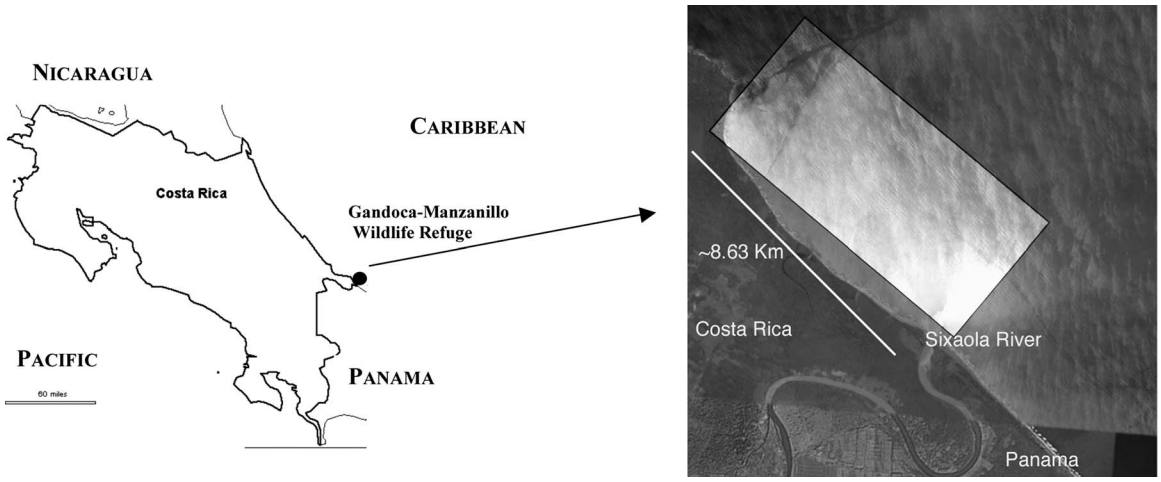


FIG. 1. Map showing the location of the Gandoca-Manzanillo Wildlife Refuge (9°59.972′ N, 82°60.530′ W) in Costa Rica and the surveyed area.

in understanding dolphin whistle structure and its geographical variation (Bazúa-Durán and Au, 2002; Au *et al.*, 1999), as well as for species classification (Oswald *et al.*, 2004). Acoustic methods have become an important tool for species identification in the field, but the success of such methods relies on the use of recording systems and analysis bandwidths proper for the species under study (Oswald *et al.*, 2004). Oswald *et al.* (2004) showed how an increase in recording system bandwidth improved the correct whistle classification for four sympatric dolphin species in the Eastern Tropical Pacific.

Although the understanding of Guyana dolphin whistle acoustic structure is growing, knowledge remains disproportionately concentrated on Brazilian populations and to a limited portion (below 24 kHz) of the frequency span of the species. Populations in other areas need to be studied to determine if there are latitudinal gradients in whistle parameters and recordings need to be made with equipment capable of recording over a greater bandwidth. The goal of this study is to (1) describe whistles of a small resident population of Guyana dolphins from Costa Rica (at its northern limit) using a broadband recording system, (2) evaluate the effect of the frequently used bandwidth recording systems (18 and 24 kHz) on whistle structure and whistle classification with respect to the sympatric bottlenose dolphin, and finally (3) compare whistle structure between Costa Rican and Brazilian populations to provide insights on whistle geographic variation after accounting for differences in recording systems' bandwidth.

II. METHODS

A. Study site

The only resident population of Guyana dolphins in Costa Rica inhabits the protected waters of the Gandoca-Manzanillo Wildlife Refuge on the southern Caribbean coast of Costa Rica (Fig. 1) (May-Collado, 2008). An ongoing photoidentification study suggests the population is relatively small and shows high site fidelity (Gamboa-Poveda and May-Collado, 2006). In addition to the Guyana dolphins, bottlenose dolphins (*Tursiops truncatus*) are also common in the Refuge, where the two commonly form mixed-species groups (Forestell *et al.*, 1999; Acevedo-Gutiérrez *et al.*, 2005; Gamboa-Poveda and May-Collado, 2006). Overall ambient noise levels (third octave) in the Refuge at the following frequencies 2, 6, 10, 14, and 18 kHz are 99.58, 98.18, 98.61, 104.02, and 92.10 dB, respectively (see May-Collado and Wartzok, 2008).

B. Whistle recordings and analysis

Surveys and recordings were carried out from a 10 m fiberglass boat with two engines (215 hp/4-stroke) and were restricted to an area of approximately 9.83 km^2 within the Refuge (Fig. 1). Because of the commonality of mixed-species groups in the area and the omnidirectional nature of our recording system it was important to ensure that only single-species groups of Guyana dolphins were present during the recording sessions. Therefore, only groups recorded under excellent weather conditions that allowed unambiguous confirmation and that no other dolphin species was present were used. Guyana and bottlenose dolphins contrast greatly in their fin morphology and surface behavior, allowing for confident distinction between single-species and mixed-species groups at relatively long distances. Twelve single-species groups of Guyana dolphins were recorded and 422 high quality whistles were selected for analysis. Whistles were recorded during a variety of behaviors, particularly foraging, traveling, and socializing, as well as in the presence and absence of other boats in addition to the research boat (see Table I).

Guyana dolphin signals were recorded using a broadband system consisting of a RESON hydrophone (−203 dB re 1 V/μPa, 1 Hz to 140 kHz) connected to AVISOFT recorder and Ultra Sound Gate 116 (sampling rate 400–500 kHz, 16 bits) that sent the signals to a laptop. All recording sessions were made with the research boat engine off. Recordings were made continuously in files of 2–3 min at sampling rates ranging from 384 to 500 kHz. Recordings were obtained over four periods of 1 week each (July 2004, September 2005, November 2005, and September 2006).

TABLE I. Total recorded and analyzed time for each study site. Note that the total number of whistles emitted is given only for the three most common behavioral categories during recording sessions.

Year	No. of individuals/ No. of whistles	No. of whistle per behavior			Total recorded time (min)/ analyzed (min)	Total recorded time (min) in the presence of just the research boat/ plus other boats
		Social	Foraging	Travel		
Total	155[a]/422	148	166	91	1465.89/529.10	308.5/220.6
2004	76/181	38	103	39	525.14/374	240.65/133.35
2005	15/110	···	44	52	622.35/74.76	31.87/42.89
2006	64/131	110	19	···	318.40/80.34	35.98/44.36

[a]The total number of individuals present in all recording sessions does not represent different animals. About 60% of the animals were the same based on photo-ID data.

Table I provides information on time recorded and analyzed in relation to documented behavioral activities and boat presence (in addition to the research boat).

Guyana dolphin whistles were analyzed manually using the program RAVEN 1.1 (Cornell Laboratory of Ornithology, New York) with a fast Fourier transform size of 1024 points, an overlap of 50%, and using a 512–522 sample Hann window. High quality whistles are those with a clear and dark contour from start to end (see Fig. 2). The maximum number of whistles to be analyzed per group was based on four times the number of individuals present in the group (for a similar method, see Azevedo and Van Sluys, 2005). Since the recordings were continuous but segmented into acoustic files of 2–3 min, dolphin whistles were selected from every other file. For each selected file all high quality whistles were selected avoiding oversampling those whistles with the same contour. Based on simultaneous photoidentification taken during each recording session, it is known that at least 60% of the photoidentified dolphins were consistently present across recording sessions. Thus, whistle selection was done with the purpose of minimizing oversampling of individuals more than groups.

Seven standard whistle variables were measured on the fundamental frequency of each: starting frequency (SF), ending frequency (EF), minimum frequency (MinF), maximum frequency (MaxF), delta frequency (DF=MaxF−MinF), duration (s), and number of inflection points (see, e.g., Wang *et al.*, 1995; Oswald *et al.*, 2003, 2004; Erber and Simão

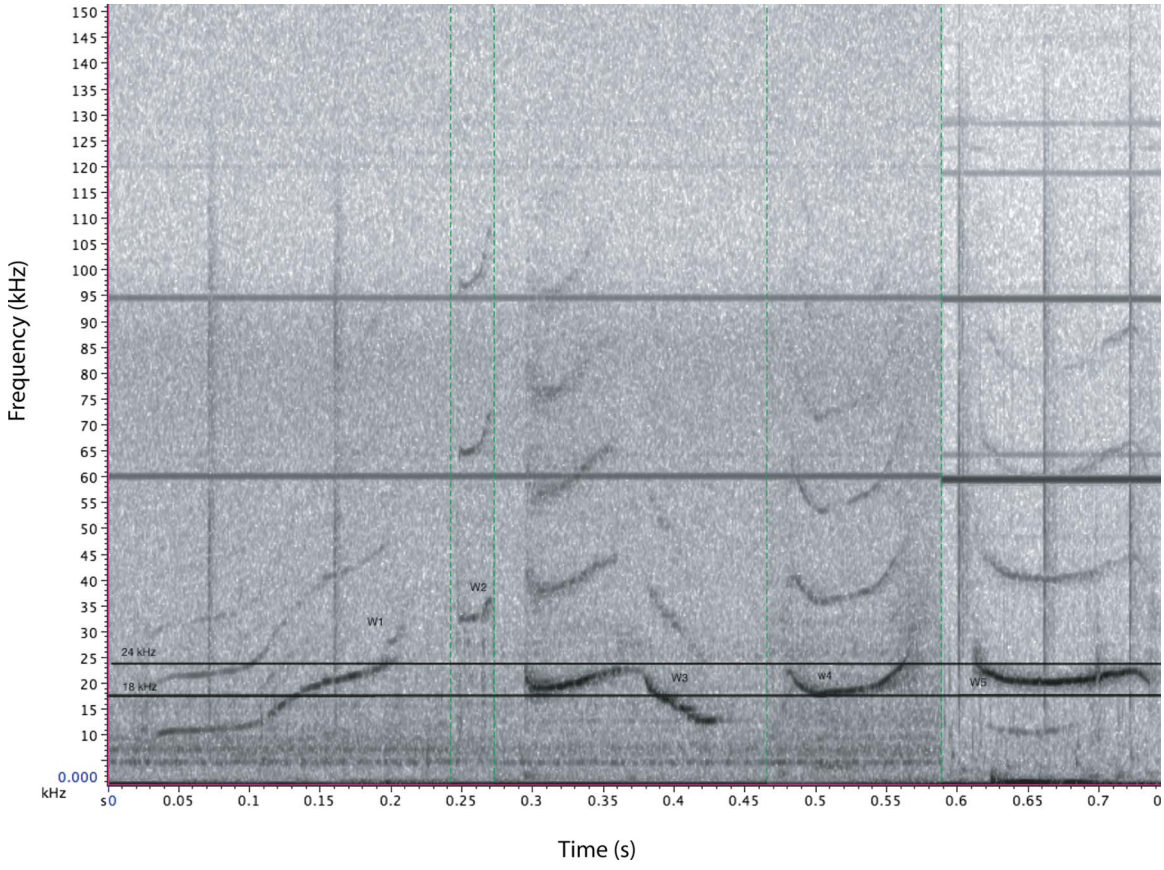


FIG. 2. (Color online) Examples of different whistles (fundamental and harmonics) emitted by Guyana dolphins from Gandoca-Manzanillo Wildlife Refuge, Costa Rica. The horizontal lines represent the mimicked bandwidth limits at 18 and 24 kHz.

TABLE II. Summary of descriptive statistics of whistle parameters for Guyana dolphins from Costa Rica and Brazil. The light-gray rows provide summary statistics for the pairwise comparisons between studies (n=whistle sample size, SD=standard deviation, CV=coefficient of variation, and *=significant results a the level of $p \leq 0.02$). Note that some of these studies in Brazil referred to the Guyana dolphin (*Sotalia guianensis*) as the marine ecotype of the tucuxi also referred as marine tucuxi or estuarine dolphin (*Sotalia fluviatilis*) but two separate species are recognized today. ψ=The whistle subsample 2 containing whistles with maximum frequency below 18 kHz were further subsampled to obtained only whistles emitted during foraging activities, n=48 whistles. See bottlenose dolphin whistle characteristics in May-Collado and Wartzok, 2008.

Study	Recording bandwidth (kHz)	Stats	MinF	MaxF	DeltaF	StartF	EndF	PeakF	$\frac{1}{4}$ Freq	$\frac{1}{2}$ Freq	$\frac{3}{4}$ Freq	Duration	IP	Harmonics
Monteiro-Filho and Monteiro (2001) Cananeia, Southern, Brazil (During various behaviors)	~8	Mean Range CV% n=214	0.3	6	⋯	⋯	⋯	⋯	⋯	⋯	⋯	0.07–0.21	⋯	⋯
Azevedo and Simão (2002) Guanabara Bay, Brazil (During various behaviors)	18	Mean±SD Range CV% n=5086	⋯	⋯	⋯	7.9±2.9	12.7±4.5	⋯	⋯	⋯	⋯	0.102.5±0.081	⋯	⋯
vs Whistle subsample 1		t-test df p-value	⋯	⋯	⋯	10.03 5192 <0.0001*	2.78 5192 0.0054*	⋯	⋯	⋯	⋯	NS	⋯	⋯
Erber and Simão (2004) Sepetiba Bay, Brazil (During various behaviors)	24	Mean±SD Range CV% n=3350	10.521 ±4.518 1.031–10.98 42.9%	13.312 ±4.85 1.171–17.49 27.7%	12.803±7.05 1–7.21 177.5%	10.704 ±4.97 1.03–11.06 46.5%	13.312 ±5.863 3.2–16.83 34.8%	⋯	11.11 ±4.72 2.74–12.07 39.1%	13.66 ±6.186 2.33–15.11 45.3%	15.368 ±6.441 2.053–21.7 41.9%	0.789 ±3.12 0.09–2.28 395.3%	1.3 ±1 0–9 110.5%	1.4±1 1–10 68.3%
vs Whistle subsample 2		t-test df p-value	4.51 3770 <0.0001*	23.41 3770 <0.0001*	15.17 3770 <0.0001*	8.96 3770 <0.0001*	14.02 3770 <0.0001*	⋯	17.82 3770 <0.0001*	5.02 3770 <0.0001*	6.82 3770 0.048	3.92 3770 <0.0001*	16.60 3770 <0.0001*	8.63 3770 <0.0001*
Azevedo and Van Sluys (2005) Southern and Northern Brazil (During various behaviors)	24	Mean±SD Range CV% n=696	9.22 ±3.44 1.34–20.3 37.3%	19.05 ±2.97 9.23–23.89 15.6%	9.83 ±4.03 0.21–22.20 41%	9.57±3.76 1.34–21.93 39.3%	18.82 ±3.10 9.23–23.75 16.5%	⋯	11.73 ±3.53 3.9–21.7 30.1%	13.85 ±3.58 5.7–23.4 25.8%	15.99 ±3.43 7.4–23.6 21.5%	0.308±0.137 0.038–1.064 44.6%	0.37 ±1.02 0–8 275.7%	⋯
vs Whistle subsample 1		t-test df p-value	9.10 1116 <0.0001*	8.19 1116 0.83	9.45 1116 0.0056*	13.29 1116 <0.0001*	2.43 1116 <0.0001*	⋯	11.54 1116 <0.0001*	6.16 1116 <0.0001*	0.048 1116 0.96	11.88 1116 <0.0001*	1.44 1116 0.15	⋯

TABLE II. (*Continued.*)

Study	Recording bandwidth (kHz)	Stats	MinF	MaxF	DeltaF	StartF	EndF	PeakF	$\frac{1}{4}$ Freq	$\frac{1}{2}$ Freq	$\frac{3}{4}$ Freq	Duration	IP	Harmonics
Pivari and Rosso (2005)	18	Mean ± SD	7.97 ± 2.89	14.46 ± 2.88	6.48 ± 3.13	8.15 ± 3.0	14.35 ± 3.0	⋯	⋯	⋯	⋯	0.229	0.17 ± 0.51	⋯
		Range	1.0–15.80	2.2–17.90	0–16.30	1.0–16.0	2.0–17.9					± 0.110	0–4	
Southeastern Brazil		CV%	36.20%	19.91%	48.29%	36.77%	21.20%					0.038–0.627	294.7%	
(During foraging)		n=3235										47.87%		
vs														
Whistle sub-sample 3[ψ]		t-test	3.38	NS	2.91	5.56	NS	⋯	⋯	⋯	⋯	NS	NS	⋯
(during foraging,)		df	3281		3281	3281								
		p-value	0.0007*		0.0037*	<0.0001*								
This study	200–250	Mean ± SD	12.31	21.21	9.02	13.83	19.51	16.11	15.37	16.60	17.60	0.200	0.440	0.932
Gandoca-Manzanzillo,			± 5.16	± 5.82	± 5.71	± 6.16	± 6.36	± 5.59	± 5.35	± 5.33	± 5.36	± 0.187	± 1.03	± 1.38
Costa Rica		Range	1.38–35.75	3.0–48.40	0.95–29.30	1.13–47.36	1.52–47.36	1.76–39.06	1.10–39.06	1.13–37.60	5.37–39.06	0.007–1.027	0–8	0–13
(During various		CV%	41.90%	27.45%	59.14%	44.59%	32.62%	34.68%	34.88%	32.18%	30.45%	93.52%	234.57%	148.53%
behaviors)		n=422												
Whistle subsample 1	24	Mean ± SD	11.57	19.01	7.63	12.99	17.43	15.33	14.30	15.19	16.00	0.031	0.459	0.856
			± 3.94	± 3.44	± 4.74	± 4.77	± 4.10	± 3.71	± 3.81	± 3.51	± 3.24.	± 0.573	± 0.97	± 1.35
		Range	1.38–19.13	3.0–23.98	0.95–17.80	1.13–23.38	1.52–23.98	1.76–23.24	1.10–21.68	1.13–22.56	5.37–21.97	0.02–1.05	0–7	0–10
		CV%	34.04%	18.07%	53.22%	36.74%	23.56%	25.09%	26.62%	23.06%	20.24%	96.24%	212%	143.80%
		n=335												
Whistle	18	Mean ± SD	9.57	14.93	5.60	10.75	13.91	12.27	11.77	12.27 ± 2.58	12.84	0.097	0.35	0.86
subsample 2			± 3.16	± 2.72	± 2.78	± 3.77	± 2.97	± 2.59	± 2.70	1.13–17.25	± 2.46	± 0.10	± 0.71	± 1.58
		Range	1.46–17.69	3.0–17.99	1.06–14.26	1.13–17.67	3.0–17.99	1.76–16.87	1.10–17.25	21.06%	5.37–16.99	0.02–0.105	0–4	0–10
		CV%	33.1%	18.19%	49.74%	35.06%	21.36%	21.10%	22.95%		19.17%	103.2	203.1%	184.10%
		n-108												

2004; Bazúa-Durán and Au, 2002, 2004; Azevedo and Van Sluys, 2005; Baron *et al.*, 2008). In addition, the contour was divided into four parts equally distributed in time to measure the frequency at $\frac{1}{4}$, $\frac{1}{2}$, and $\frac{3}{4}$ of the contour. These three frequency measurements are standard in Guyana dolphin whistle studies (see Azevedo and Van Sluys, 2005; Erber and Simão, 2004; Rossi-Santos and Podos, 2006). Peak frequency (PF) was also measured and is defined as the frequency at which maximum power occurs (Bazúa-Durán and Au, 2002, 2004; May-Collado and Wartzok, 2007) and number of harmonics (e.g., Wang *et al.*, 1995; Erber and Simão, 2004). Whistle contours were categorized as ascending, descending, ascending-descending, descending-ascending, and constant in frequency, sine, and others (see Azevedo and Van Sluys, 2005; Erber and Simão, 2004).

C. Effect of recording systems' bandwidth

To evaluate the effect of bandwidth limit on understanding Guyana dolphin whistle structure the 422 analyzed dolphin whistles (full data set) were subsampled into a data set containing all whistles with maximum frequency below 18 kHz ($n=108$ whistles, subsample 1) to "mimic" the recording system used by Azevedo and Simão (2002) and Pivari and Rosso (2005) with a bandwidth up to 18 kHz, and a data set containing whistles with maximum frequency below 24 kHz ($n=335$ whistles, subsample 2) to mimic the recording system of Erber and Simão (2004) and Azevedo and Van Sluys (2005) with a bandwidth up to 24 kHz. Whistle frequency variables were then compared for the three data sets using multivariate statistics; see Sec. II E.

Because bandwidth limit has been shown to have an important effect on dolphin whistle correct classification among dolphin species (see Oswald *et al.*, 2004), the effect of bandwidth limits was also evaluated on whistle classification of the sympatric Guyana dolphins and bottlenose dolphins. A total of 77 bottlenose dolphin whistles were obtained from a previous study in the same study area and following the same recording protocol used in this study (see May-Collado and Wartzok, 2008). Because these whistles were recorded at an upper frequency of 192–250 kHz, bottlenose dolphin whistles were also subsampled to mimic the limited bandwidth, 18 and 24 kHz, for comparison purposes.

D. Whistle comparison with other studies (populations)

Comparisons between populations were made using published data on mean values for SF, EF, MinF, MaxF, DF, $\frac{1}{4}F$, $\frac{1}{2}F$, $\frac{3}{4}F$, and duration, and when possible the mean number of inflection points and harmonics were also included (see Table II). To account for differences in bandwidth between studies, the whistle variables reported by Azevedo and Simão (2002) and Pivari and Rosso (2005) were compared to the whistle subsample 1 (whistles with maximum frequency below 18 kHz), and whistle variables reported by Erber and Simão (2004) and Azevedo and Van Sluys (2005) were compared to subsample 2 (whistles with maximum frequency below 24 kHz).

Finally, it is important to emphasize that the Guyana dolphin (*S. guianensis*), an exclusively marine species, was recently recognized as a separate species from the freshwater Tucuxi dolphin (*S. fluviatilis*) (see Monteiro-Filho *et al.*, 2002; Cunha *et al.*, 2005; and Caballero *et al.*, 2007). Because of this recent taxonomic change, the studies above identified the dolphins as *S. fluviatilis*. However, all studies considered in Table II for comparison correspond to the Guyana dolphin (*S. guianensis*) as all of them took place in marine environments.

E. Statistical analyses

The statistical softwares SPSS 16.0, 2007 (SPSS Inc.) and JMP 2007® (SAS Institute Inc.) were used for statistical analyses. Descriptive statistics were performed to provide mean, standard deviation, frequency range, and coefficient of variation values for each whistle. All whistle variables were Box-Cox transformed (except for the number of inflection points and harmonics) to normalize their distribution (Sokal and Rohlf, 1995). Multivariate analyses of variance (MANOVAs) were performed to determine whether whistle variables SF, EF, MinF, MaxF, DF, PF, $\frac{1}{4}F$, $\frac{1}{2}F$, $\frac{3}{4}F$, and duration vary with bandwidth limits, sympatric species, and their interaction. The Box's M-test was used to evaluate homogeneity among covariance matrices. Because MANOVA performs multiple univariate ANOVA analyses, type I error was controlled using a Bonferroni procedure to adjust the level of significance. The same procedure was used for multiple pairwise comparisons of whistle variables among group factors. A scatter plot between starting and ending frequencies was made to visualize the differences in frequency span between bandwidths.

To examine the effect of bandwidth on whistle classification between the two sympatric dolphin species, a discriminant analysis was performed for each whistle data set separately using whistle SF, EF, MinF, MaxF, DF, PF, and duration as predictors. Since the covariance matrices for the species were significantly different (Box's $M=1208.02$, df1 $=140$, $p<0.0001$) we adjusted the prior probabilities by computing classification scores from group size. The canonical correlation (equivalent to Pearson's correlation and proper for two groups) was used to determine the efficacy of the discriminant function. The chi-square statistics test was used to assess how well the discriminant function does versus chance alone at the statistical significance level of 0.05 (Green and Salkind, 2003). The cross-validation method was used to calculate correct classification scores for the discriminant functions and the Kappa index as well as a chi-square test were performed to evaluate the accuracy of the classification at the p-value level of $p=0.05$ (Green and Salkind, 2003). Box plots were generated to compare whistle structure between species.

For comparisons between populations published whistle data on mean (and standard deviation) SF, EF, MinF, MaxF, DF, PF, $\frac{1}{4}F$, $\frac{1}{2}F$, $\frac{3}{4}F$, and duration were used to compare with either subsample 1 (18 kHz) or subsample 2 (24 kHz) depending on the recording system used in the published study. Before performing each pairwise comparison the assumption of equal variance was tested using Levene's F-test. When

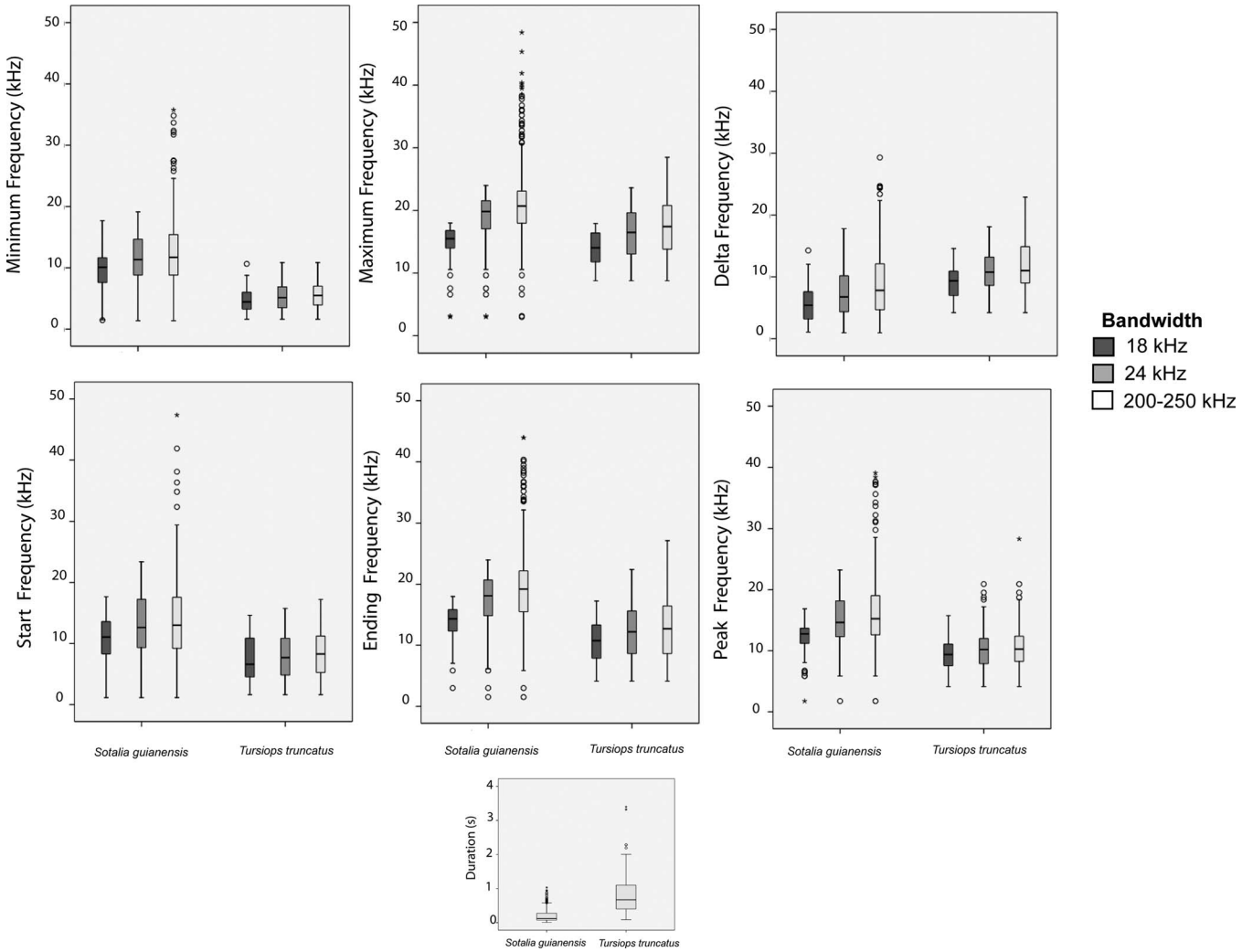


FIG. 4. Whistle variation in frequency and time parameters of both sympatric dolphin species whistles recorded using the broadband recording system (200 Hz–250 kHz).

variances were equal the *t*-test was used and the Welch *t*-test for unequal variances. The Bonferroni procedure was also used here to adjust the level of significance for the multiple comparisons.

III. RESULTS

A. Whistle characterization

Guyana dolphins from Costa Rica emitted whistles with a greater frequency span (1.38 up to 48.40 kHz) than previously reported in Brazilian studies using bandwidth-limited recorded systems (see Table II). The broadband recording system also allowed detection of high order harmonics for 37% of the total analyzed whistles (Table III and Fig. 2). Most of these whistles contained one and two harmonics, and up to 13 harmonics reaching frequencies up to 136 kHz. Guyana dolphin emitted whistles that were mainly ascending in frequency (57.6%) followed by constant (13%), descending (10.2%), ascending-descending (6.6%), descending-ascending (6%), and sine (6.6%).

B. Effect of bandwidth limits on whistle structure understanding

The means for Guyana dolphin whistle SF, EF, MinF, MaxF, DF, PF, and duration were significantly different among bandwidth limits [Wilk's $\Lambda=0.01$, $F(14,2078)=1.24\times10^3$, $p<0.0001$], dolphin species [Wilk's $\Lambda=0.60$, $F(7,1039)=101.26$, $p<0.0001$], and their interaction [Wilk's $\Lambda=0.95$, $F(14,2078)=4.02$, $p<0.0001$]. About 89% of the multivariate variance found in whistle variables was associated with bandwidth limits, 41% to dolphin species, and only 2.6% to their interaction.

Dolphin whistles SF, EF, MinF, MaxF, DF, PF, $\frac{1}{4}F$, $\frac{1}{2}F$, $\frac{3}{4}F$, and duration varied significantly among bandwidths [Wilk's $\Lambda=0.83$, $F(20,1680)=8.1$, $p<0.0001$, Table II]. Whistle maximum (16%), $\frac{3}{4}F$ (11%), and ending (10%) frequencies explained most of the whistle variation between bandwidths. Figure 3 shows how Guyana dolphin whistle frequency span (start-ending frequencies) changes considerably among bandwidths, with bandwidth at 200–250 kHz showing the entire frequency span of Guyana dolphins.

TABLE III. Maximum frequency descriptive statistics for harmonic components of 155 whistle harmonics emitted by Guyana dolphins.

No. of harmonics	No. of whistles	Stats	H1	H2	H3	H4	H5	H6	H7	H8	H9	H10	H11	H12	H13
H1	55	Mean ± SD	35.1 ± 10.1	…	…	…	…	…	…	…	…	…	…	…	…
		Range	19.9–78.6												
H2	57	Mean ± SD	39.2 ± 14.17	53.5 ± 21.1	…	…	…	…	…	…	…	…	…	…	…
		Range	14.1–103.7	13.6–121.2											
H3	18	Mean ± SD	35.6 ± 9.6	52.2 ± 16.5	65.6 ± 21.2	…	…	…	…	…	…	…	…	…	…
		Range	21.8–61.0	35.2–94.8	45.8–121.0										
H4	12	Mean ± SD	37.1 ± 9.6	52.8 ± 13.8	65.3 ± 17.8	77.9 ± 21.1	…	…	…	…	…	…	…	…	…
		Range	21.8–53.4	40.4–78.8	48.5–85.4	62.5–77.9									
H5	6	Mean ± SD	35.8 ± 9.0	55.2 ± 12.7	66.3 ± 15.5	82.2 ± 15.8	96.1 ± 15.8	…	…	…	…	…	…	…	…
		Range	20.2–43.7	34.9–65.0	40.2–66.3	59.4–78.6	78.5–91.2								
H6	1	Value	42.2	57.8	77.3	87.1	92.9	105.8	…	…	…	…	…	…	…
H9	2	Range	23.7–25.2	29.6–36.9	43.2–49.8	59.3–61.5	71.1–79.5	85.4–87.2	91.9–97.3	101.8–114.0	112.5–120.7	…	…	…	…
H10	1	Value	22.1	43.5	50.1	64.1	70.2	80.5	90.5	102.2	110.3	121.3	…	…	…
H11	2	Range	22.6–22.7	29.1–29.7	39.4–39.7	50.5–61.7	60.3–69.7	69.3–78.1	79.1–86.1	89.1–93.1	98.6–101.7	105.9–107.6	115.9–131.6	…	…
H13	1	Value	24.8	29.8	41.0	48.9	62.1	68.6	82.6	90.6	102.6	107.2	117.3	124.3	135.9

FIG. 3. Plot of whistle starting vs ending frequency for the full data set (422 whistles) using the broadband recording system (200 Hz–250 kHz) and subsampled whistle data sets with maximum frequencies below 18 and 24 kHz. The lines represent the mean value for starting and ending frequencies.

C. Effect of bandwidth limits on whistle classification

Whistle variables were significantly different between species (see Fig. 4 for comparison and statistics). The discriminant analyses correctly classified whistles with high success to the respective dolphin species regardless of bandwidth limits (Table IV). However, an increase in bandwidth slightly improved the classification success for Guyana dolphins. The "best" whistle variables to discriminate between dolphin species were whistle minimum frequency and duration for bandwidth limit at 200–250 kHz [MinF: Wilk's $\Lambda = 0.73$, $F(1,497) = 171.8$, $p < 0.0001$; duration: Wilk's $\Lambda = 0.74$, $F(1,497) = 170.8$, $p < 0.0001$] and at 24 kHz [MinF: Wilk's $\Lambda = 0.71$, $F(1,399) = 155.1$, $p < 0.0001$; duration: Wilk's $\Lambda = 0.72$, $F(1,399) = 154$, $p < 0.0001$]. Whistle duration was the most important variable for species discrimina-

TABLE IV. Classification results of the discriminant analyses for the three bandwidth-limited whistle data sets. The percentage for whistles correctly classified for each species is given in bold, all of which were significantly different (χ^2 test $p<0.0001$) than expected by chance alone at the significance level of $p<0.05$. Overall correct classification percentages are given at the bottom with their respective kappa index and χ^2 test statistics.

	Bandwidth 18 kHz Predicted species			Bandwidth 24 kHz Predicted species			Bandwidth 200–250 kHz Predicted species		
Actual species	Bottlenose dolphins	Guyana dolphins	*n*	Bottlenose dolphins	Guyana dolphins	*n*	Bottlenose dolphins	Guyana dolphins	*n*
Bottlenose dolphins	**76.7%**	23.3%	43	**66.7%**	33.3%	66	**63.6%**	36.4%	77
Guyana dolphins	11.1%	**88.9%**	108	3.9%	**96.1%**	335	3.8%	**96.2%**	422
Overall correct classification %	Overall 85.4%, kappa=0.75, $p<0.0001$ χ^2 test $p<0.0001$			Overall 91.3%, kappa=0.72, $p<0.0001$ χ^2 test $p<0.0001$			Overall 91.2%, kappa=0.66, $p<0.0001$ χ^2 test $p<0.0001$		

tion when using the 18 kHz bandwidth [Wilk's $\Lambda=0.58$, $F(1,149)=106.2$, $p<0.0001$]. Overall, Guyana dolphin whistles have a higher minimum frequency and are much shorter in duration than bottlenose dolphins (Fig. 4).

D. Whistle comparison between populations

After accounting for differences in bandwidth, pairwise comparisons between this study and the studies of Erber and Simao (2004), Pivari and Rosso (2005), and Azevedo and Van Sluys (2005) suggest that Brazilian and Costa Rican dolphins vary significantly in whistle structure (Table II for statistics). In general Guyana dolphins from Costa Rica emitted whistles that were higher in almost every frequency parameter, while Brazilian dolphins emitted significantly longer whistles.

IV. DISCUSSION

Guyana dolphin whistles have been described using a variety of bandwidth-limited recording systems (generally 18 kHz and 24 kHz). The most recent study reported a whistle frequency span from 1.34 to 23.89 kHz (Azevedo and Van Sluys, 2005; bandwidth 24 kHz), but as the authors reported some of the observed whistles appeared to extend beyond the limits of the upper frequency of the recording system. Using a broadband recording system (up to 250 kHz) this study provides evidence that Guyana dolphins from Costa Rica can emit whistles beyond 24 kHz, joining a short list of cetacean species known to emit such whistles: the botos (May-Collado and Wartzok, 2007), bottlenose dolphins (Boisseau, 2005; May-Collado and Wartzok, 2008), white-beaked dolphins (Rasmussen and Miller, 2002), spinner dolphins (e.g., Lammers *et al.*, 1997; Lammers *et al.*, 2003; Oswald *et al.*, 2004), spotted, striped, and common dolphins (Oswald *et al.*, 2004). The Guyana dolphin has the widest whistle frequency span ever reported in delphinids (1.38 up to 48.40 kHz). Also of the analyzed whistles 37% contained harmonics, some of which reached frequencies up to 136 kHz. This is the first time high order harmonics have been reported for whistles emitted by Guyana dolphins. High order harmonics in dolphin whistle sounds have been described for only a handful of dolphin species (white-beaked dolphins, Rasmussen *et al.*, 2006; spinner dolphins, Lammers and Au, 2002; killer whales, Miller, 2002). Lammers and Au (2002) showed that in spinner dolphins whistle directionality increased with frequency especially with regard to harmonics. The authors suggested that whistle harmonic structure can potentially carry information on the direction of movement of signaling animal (s) and therefore facilitate group coordination. This would be an interesting hypothesis to test in the future for Guyana dolphins.

A. Bandwidth limit and whistle structure

As shown by Oswald *et al.* (2004) in spinner, spotted, striped, and common dolphins the recording system bandwidth capabilities are a very important consideration when studying dolphin whistle acoustic characteristics. This study shows that limited bandwidth distorts the understanding of Guyana dolphin whistle frequency variables, particularly in whistle maximum, ending, and $\frac{3}{4}$ frequencies. Whistles selected to mimic narrowband recordings systems with bandwidths of 18 and 24 kHz limited the characterization of the whistle frequency span of Guyana dolphins to a portion of the actual frequency range (see Fig. 3). For instance, about 73 whistles (out of 422) had maximum frequencies that extended beyond the 24 kHz limit, and additional 14 whistles had minimum (and starting) frequencies above 24 kHz, and would have been completely missed by narrowband recording systems. In addition, most of the harmonics would have been missed with narrowband recording systems. In order to properly document whistle repertoire of Guyana dolphins (including harmonic components) a recording system with a bandwidth of at least 150 kHz is necessary.

B. Bandwidth limit and dolphin species whistle classification

Although narrowband recording systems obscure Guyana dolphin whistle frequency range the consequences of this for dolphin species whistle classification were minor, presumably because Guyana and bottlenose dolphin whistles are different enough to be discriminated with sparse data. Increase in bandwidth improved slightly correct classification percentages of whistles between species, which were in general high (85%–91%) compared to previous studies (e.g., Oswald *et al.*, 2004; Rendell *et al.*, 1999; Steiner, 1981). Both dolphin species were very distinct in their whistle structure, particularly in whistle minimum frequency and dura-

tion. Bottlenose dolphin whistles were much lower in minimum frequency and longer in duration than whistles emitted by the Guyana dolphins (see Fig. 4).

The clear distinction between Guyana dolphins and bottlenose dolphin whistles may be the result of a combination of factors as follows: (1) Phylogenetic distance, the two species belong to different subfamilies (e.g., LeDuc *et al.*, 1999; May-Collado and Agnarsson, 2006; Agnarsson and May-Collado, 2008); (2) body size, bottlenose dolphins are large (up to 3.0 m) and robust animals, while Guyana dolphins are small (up to 1.79 m) and slender (Rosas and Monteiro-Filho, 2002). Because body size and minimum frequency are negatively correlated in cetaceans (e.g., Matthews *et al.*, 1999; May-Collado *et al.*, 2007a), this intrinsic relationship may largely account for the clear distinction in whistle structure between the two dolphin species; and (3) differences in social structure. Bottlenose dolphins live in complex societies where some individuals sustaining long-term relationships (e.g., Mann *et al.*, 2000) while Guyana dolphins live in relatively simple and fluid societies with no apparent long-term relationships as in bottlenose dolphins (De Oliveira and Rosso, 2008). Interestingly, May-Collado *et al.* (2007b) found a relationship between social elements such as group size, and whistle minimum frequency and duration, where in general social species living in simple societies tended to emit whistles that were higher in frequency and shorter in duration. Previous dolphin species whistle classification studies have not taken into consideration these factors, but these seem to be key particularly when algorithms are being designed to improve classification scores for species identification. For instance, Oswald *et al.* (2004) obtained relatively low correct classification percentages between spinner, spotted, striped, and common dolphins (30%–37%); these species are closely related with relatively similar body size and social structures [but see study by Rendell *et al.* (1999)]. In contrast, Steiner (1981) obtained relatively high correct classification percentages between bottlenose, spotted, Atlantic-white sided dolphins, and pilot whales (57%–80%). These species are not closely related, and all four vary considerably in size and social structures. We proposed that future classification algorithms should take in consideration phylogenetic relationships, body size, and social structure as tools that can guide the algorithm to classify species.

C. Comparison between populations

After accounting for difference in recording equipment bandwidth, comparisons between whistles from the Costa Rican and the Brazilian populations showed significant differences in whistle structure. Brazilian dolphins emit longer whistles than the Costa Rican dolphins (Erber and Simão, 2004; Azevedo and Van Sluys, 2005). However, whistles from the Costa Rican dolphins were consistently higher in almost all whistle frequency variables described for the Brazilian populations (see Table II). These results provide corroborative evidence for the hypothesis proposed by Azevedo and Van Sluys (2005) and Rossi-Santos and Podos (2006) that Guyana dolphins' whistle frequency increases (particularly in minimum and starting frequencies) from south to north. A stronger test of this hypothesis must await a broadband recording study of the Brazilian populations, and other populations in between.

Several factors have been proposed to explain dolphin whistle geographical variations including dispersal capabilities of a species (McGregor *et al.*, 2000; Mundinger, 1982), isolation and genetic divergence between groups or populations (e.g., Ford, 2002; McGregor *et al.*, 2000), and adaptation to ecological conditions (e.g., Brumm, 2006; Gillam and McCracken, 2007; Morisaka *et al.*, 2005; Peters *et al.*, 2007). Rossi-Santos and Podos (2006) noticed in Guyana dolphin whistles a discontinuity particularly in whistle minimum and starting frequencies and suggested that this pattern could reflect dispersal limitations between populations. There appears to be a discontinuity in Guyana dolphin distribution in Central America and Panama, where pockets of Guyana dolphin populations occur along the Caribbean coast, one in the southern part of Panama (May-Collado, 2008), in the northern part of Nicaragua (Cayos Miskito Reserve) (e.g., Carr and Bonde, 2000; Edwards and Schnell, 2001), and the Costa Rican population, which appears to be restricted to the studied area [77.2% of the photoidentified animals are regularly observed in the Refuge year around (Gamboa-Poveda and May-Collado, 2006)].

V. CONCLUSIONS

This study confirms that the whistle repertoire (fundamental and harmonics) of the Guyana dolphin, *Sotalia guianensis*, extends beyond 24 kHz, with a frequency span among the greatest ever reported in delphinid species. The importance of a broadband recording system to study the entire whistle repertoire is demonstrated as prior studies using a narrowband recording system gave only an incomplete understanding of Guyana dolphin whistles. Although the dolphin species studied here are very distinct in their whistle structure, an increase in broadband recording systems slightly improved the whistle classification of Guyana dolphin species. Until broadband recording systems are used for more populations, the potential patterns in whistle geographical variation and factors promoting such variation remain poorly understood. However, this study provides evidence supporting the hypothesis that whistle frequency variables increase with latitude. Future studies on Guyana dolphin whistles should employ recording systems with bandwidth up to 50 kHz (for whistle fundamental) and up to 150 kHz (when considering high order harmonics) to ensure the inclusion of the entire whistle repertoire.

ACKNOWLEDGMENTS

Thanks to Ingi Agnarsson, University of Puerto Rico, and two anonymous reviewers for their suggestions that improved the manuscript. Thanks to Alexandre Azevedo, Universidade do Estado do Rio de Janeiro, and Marcos Rossi-Santos, Instituto Baleia Jubarte, Brazil, for initial guidance on *Sotalia* whistle analyses. Also thanks to the captains Dennis Lucas and Alfonso and whale-watching operators at the Wildlife Refuge of Gandoca-Manzanillo. The following

people assisted in the field: Mónica Gamboa-Poveda, Jose David Palacios, Jose D. Martinez, Evi Taubitz, Jorge May-Barquero, Yadira Collado-Ulloa, and. This study was carried out with permission from the Ministerio de Ambiente y Energía and the National Park System, Area de Conservación Talamanca (Permit No. 137-2005 SINAC) de la República de Costa Rica. Funding for this project came from The Latin American Student Field Research Award by the American Society of Mammalogists, Judith Parker Travel Grant, Lerner-Gray Fund for Marine Research of the American Museum of Natural History, Cetacean International Society, Project Aware, Whale and Dolphin Conservation Society, the Russell E. Train Education Program-WWF, and a Dissertation Year Fellowship, Florida International University to Laura May-Collado.

Acevedo-Gutiérrez, A., DiBerardinis, A., Larkin, S., Larkin, K., and Forestell, P. (**2005**). "Social interactions between tucuxis and bottlenose dolphins in Gandoca-Manzanillo, Costa Rica," LAJAM **4**, 49–54.

Agnarsson, I., and May-Collado, L. J. (**2008**). "The phylogeny of Cetartiodactyla: The importance of dense taxon sampling, missing data, and the remarkable promise of cytochrome b to provide reliable species-level phylogenies," Mol. Phylogenet. Evol. **48**, 964–985.

Au, W. W. L., Lammers, M. O., and Aubauer, R. (**1999**). "A portable broadband data acquisition system for field studies in bioacoustics," Marine Mammal Sci. **15**, 526–530.

Azevedo, A. F., and Simão, S. M. (**2002**). "Whistles produced by marine tucuxi dolphins *Sotalia fluviatilis* in Guanabara Bay, southeastern Brazil," Aquat. Mamm. **28**, 261–266.

Azevedo, A. F., and Van Sluys, M. (**2005**). "Whistles of tucuxi dolphins (*Sotalia fluviatilis*) in Brazil: Comparisons among populations," J. Acoust. Soc. Am. **117**, 1456–1464.

Baron, S. C., Marinez, A., Garrison, L. P., and Keith, E. O. (**2008**). "Differences in acoustic signals from delphinids in the western North Atlantic and northern Gulf of Mexico," Marine Mammal Sci. **24**, 42–56.

Bazúa-Durán, M. C., and Au, W. W. L. (**2002**). "Whistles of Hawaiian spinner dolphins," J. Acoust. Soc. Am. **112**, 3064–3072.

Bazúa-Durán, M. C., and Au, W. W. L. (**2004**). "Geographic variations in the whistles of spinner dolphins (*Stenella longirostris*) of the Main Hawaiian Islands," J. Acoust. Soc. Am. **116**, 3757–3769.

Boisseau, O. (**2005**). "Quantifying the acoustic repertoire of a population: The vocalizations of free-ranging bottlenose dolphins in Fiordland, New Zealand," J. Acoust. Soc. Am. **117**, 2318–2329.

Brumm, H. (**2006**). "Animal communication: City birds have changed their tune," Curr. Biol. **16**, R1003–R1004.

Caballero, S., Trujillo, F., Vianna, J. A., Barrios-Garrido, H., Montiel, M. G., Beltrán-Pedreros, S., Marmontel, M., Santos, M. C., Rossi-Santos, M., Santos, F. R., and Baker, C. S. (**2007**). "Taxonomic status of the genus *Sotalia*: Species level ranking for "Tucuxi" (*Sotalia fluviatilis*) and "Costero" (*Sotalia guianensis*) dolphins," Marine Mammal Sci. **23**, 358–386.

Carr, T., and Bonde, R. K. (**2000**). "Tucuxi (*Sotalia fluviatilis*) occurs in Nicaragua, 800 km north of its previously known range," Marine Mammal Sci. **16**, 447–452.

Cunha, H. A., da Silva, V. M. F., Lailson-Brito, J. Jr., Santos, M. C. O., Flores, P. A. C., Martin, A. R., Azevedo, A. F., Fragoso, A. B. L., Zanelatto, R. C., and Solé-Cava, A. M. (**2005**). "Riverine and marine ecotypes of *Sotalia* dolphins are different species," Mar. Biol. (Berlin) **148**, 1432–1793.

Da Silva, V. M. F., and Best, R. C. (**1996**). "*Sotalia fluviatilis*," Mammalian Species **527**, 1–7.

De Oliveira, S. M. C., and Rosso, S. (**2008**). "Social organization of marine tucuxi dolphins, Sotalia guianensis, in the Cananeia estuary of southeastern Brazil," J. Mammal. **89**, 347–355.

Edwards, H. H., and Schnell, G. D. (**2001**). "Status and ecology of *Sotalia fluviatilis* in the Cayos Miskito Reserve, Nicaragua," Marine Mammal Sci. **17**, 445–472.

Erber, C., and Simão, S. M. (**2004**). "Analysis of whistles produced by the Tucuxi Dolphin, *Sotalia fluviatilis* from Sepetiba Bay, Brazil," An. Acad. Bras. Cienc. **76**, 381–385.

Flores, P. A. C. (**2002**). "Tucuxi-*Sotalia fluviatilis*," in *Encyclopedia of Marine Mammals*, edited by W. F. Perrin, B. Wursig, and J. G. M. Thewissen (Academic, New York), pp. 1267–1269.

Ford, J. K. B. (**2002**). "Dialects," in *Encyclopedia of Marine Mammals*, edited by W. F. Perrin, B. Wursig, and J. G. M. Thewissen (Academic, New York), pp. 322–323.

Forestell, P., Wright, A., DiBerardinis, A., Larkin, S., and Schott, V. (**1999**). "Sex and the single tucuxi: Mating between bottlenose and tucuxi dolphins in Costa Rica," in *Abstracts, 13th Biennial Conference on the Biology of Marine Mammals*, Maui, HI (Society of Marine Mammalogy).

Gamboa-Poveda, M., and May-Collado, L. J. (**2006**). "Insights on the occurrence, residency, and behavior of two coastal dolphins from Gandoca-Manzanillo, Costa Rica: *Sotalia guianensis* and *Tursiops truncatus* (Family Delphinidae)," presented to the *IWC Scientific Committee*, St. Kitts and Nevis, WI, June 2006, available from The International Whaling Commission, The Red House, 135 Station Road, Impington, Cambridge, Cambridgeshire CB24 9NP, UK.

Gillam, E. H., and McCracken, G. F. (**2007**). "Variability in the echolocation of *Tadaria brasiliensis*: Effects of geography and local acoustic environment," Anim. Behav. **74**, 277–286.

Green, S. B., and Salkind, N. J. (**2003**). *Using SPSS for Window and Macintosh Analyzing and Understanding Data* (Prentice-Hall, Englewood Cliff, NJ).

Lammers, M. O., and Au, W. W. L. (**2002**). "Directionality in the whistles of Hawaiian spinner dolphins (*Stenella longirostris*): A signal feature to cue direction of movement?," Marine Mammal Sci. **19**, 249–364.

Lammers, M. O., Au, W. W. L., and Aubauer, R. (**1997**). "Broadband characteristics of spinner dolphin (*Stenella longirostris*) social acoustic signals," J. Acoust. Soc. Am. **102**, 3122.

Lammers, M. O., Au, W. W. L., and Herzing, H. L. (**2003**). "The broadband social acoustic signaling behavior of spinner and spotted dolphins," J. Acoust. Soc. Am. **114**, 1629–1639.

LeDuc, R. G., Perrin, W. F., and Dizon, A. E. (**1999**). "Phylogenetic relationships among the delphinid cetaceans based on full cytochrome *b* sequences," Marine Mammal Sci. **15**, 619–648.

Mann, J., Connor, R. C., Tyack, P. L., and Whitehead, H. (**2000**). *Cetacean Societies: Field Studies of Dolphins and Whales* (University of Chicago Press, Chicago).

Matthews, J. N., Rendell, L. E., Gordon, J. C. D., and MacDonald, D. W. (**1999**). "A review of frequency and time variables of cetacean tonal calls," Bioacoustics **10**, 47–71.

May-Collado, L. J. (**2008**). "Marine mammals," *The Marine Biodiversity of Costa Rica, Central America (Monographiae Biologicae)*, edited by I. S. Wehrtmann and J. Cortés (Springer, New York), pp. 915–936.

May-Collado, L. J., and Agnarsson, I. (**2006**). "Cytochrome b and Bayesian inference of whale phylogeny," Mol. Phylogenet. Evol. **32**, 344–354.

May-Collado, L. J., and Wartzok, D. (**2007**). "The freshwater dolphin *Inia geoffrensis* produce high frequency whistles," J. Acoust. Soc. Am. **121**, 1203–1212.

May-Collado, L. J., and Wartzok, D. (**2008**). "A comparison of bottlenose dolphin whistles in the Atlantic Ocean: Factors promoting whistle variation," J. Mammal. **89**, 1229–1240.

May-Collado, L. J., Agnarsson, I., and Wartzok, D. (**2007a**). "Reexamining the relationship between body size and tonal signals frequency in whales: A comparative approach using a novel phylogeny," Marine Mammal Sci. **23**, 524–552.

May-Collado, L. J., Agnarsson, I., and Wartzok, D. (**2007b**). "Phylogenetic review of tonal sound production in whales in relation to sociality," BMC Evol. Biol. **7**, 136.

McGregor, P. K., Peake, T. M., and Gilbert, G. (**2000**). "Communication, behaviour, and conservation," *in Behaviour and Conservation*, edited by L. M. Gosling and W. J. Sutherland (Cambridge University Press, Cambridge), pp. 261–285.

Miller, P. J. O. (**2002**). "Mixed-directionality of killer whale stereotyped calls: A direction of movement cue?," Behav. Ecol. Sociobiol. **52**, 262–270.

Monteiro-Filho, E. L. A., and Monteiro, K. D. K. A. (**2001**). "Low-frequency sounds emitted by *Sotalia fluvialitis guianensis* (Cetacea: Delphinidae) in an estuarine region in southeastern Brazil," Can. J. Zool. **79**, 59–66.

Monteiro-Filho, E. L. A., Monteiro, L. R., and Dos Reis, S. F. (**2002**). "Skull shape and size divergence in dolphins of the genus *Sotalia*: A tridimensional morphometric analysis," J. Mammal. **83**, 125–134.

Morisaka, T., Shinohara, M., Nakahara, F., and Akamatsu, T. (**2005**). "Effects of ambient noise in the whistles of Indo-Pacific bottlenose dolphin

Tursiops aduncus populations in Japan," J. Mammal. **86**, 541–546.
Mundinger, P. C. (**1982**). "Microgeographic and macrogeographic variation in the acquired vocalizations of birds," in *Acoustic communication in birds: Song learning and its consequences*, edited by Kroodsma, Miller E. H., and H. Ouellet (Academic, New York), pp. 147–208.
Oswald, J. N., Barloy, J., and Norris, T. F. (**2003**). "Acoustic identification of nine delphinids species in the eastern tropical Pacific ocean," Marine Mammal Sci. **19**, 20–37.
Oswald, J. N., Rankin, S., and Barlow, J. (**2004**). "The effect of recording and analysis bandwidth on acoustic identification of delphinids species," J. Acoust. Soc. Am. **116**, 3178–3185.
Peters, R. A., Hemmi, J. M., and Zeil, J. (**2007**). "Signaling against the wind: Modifying motion-signal structure in response to increased noise," Curr. Biol. **17**, 1231–1234.
Pivari, D., and Rosso, S. (**2005**). "Whistles of small groups of *Sotalia fluviatilis* during foraging behavior in southeastern Brazil," J. Acoust. Soc. Am. **118**, 2725–2731.
Rasmussen, M. H., and Miller, L. A. (**2002**). "Whistles and clicks from white-beaked dolphins, (*Lagenorhynchus albirostris* Gray 1846) recorded in Faxafloi Bay, Iceland," Aquat. Mamm. **28**, 78–89.
Rasmussen, M. H., Lammers, M., Beedholm, K., and Miller, L. A. (**2006**). "Source levels and harmonic content of whistles in white-beaked dolphins (*Lagenorhynchus albirostris*)," J. Acoust. Soc. Am. **120**, 510–517.
Rendell, L. E., Matthews, J. N., Gill, A., Gordon, J. C. D., and MacDonald, D. W. (**1999**). "Quantitative analysis of tonal calls from five odontocete species, examining interspecific and intraspecific variation," J. Zool. **249**, 403–410.
Rosas, W. F. C., and Monteiro-Filho, E. L. A. (**2002**). "Reproduction of the Estuarine dolphin (*Sotalia guianensis*) on the coast of Parana, Southern Brazil," J. Mammal. **83**, 507–515.
Rossi-Santos, M. R., and Podos, J. (**2006**). "Latitudinal variation in whistle structure of the estuarine dolphin *Sotalia guianensis*," Behaviour **143**, 347–364.
Sokal, R. R., and Rohlf, F. J. (**1995**). *Biometry* (W. H. Freeman, New York).
Steiner, W. W. (**1981**). "Species-specific differences in pure tonal whistle vocalizations of five western North Atlantic dolphin species," Behav. Ecol. Sociobiol. **9**, 241–246.
Terry, R. P. (**1983**). "Observations on the captive behaviour of *Sotalia fluviatilis guianensis*," Aquat. Mamm. **10**, 95–105.
Wang, D., Würsig, B., and Evans, W. E. (**1995**). "Whistles of bottlenose dolphins: comparisons among populations," Aquat. Mamm. **21**, 65–77.
Wiersma, H. (**1982**). "Investigations on Cetacean Sonar IV, a comparison of wave shapes of odontocete sonar signals," Aquat. Mamm. **9**, 57–66.

Functional bandwidth of an echolocating Atlantic bottlenose dolphin (*Tursiops truncatus*)

Stuart D. Ibsen,[a] Whitlow W. L. Au, Paul E. Nachtigall, and Marlee Breese
Marine Mammal Research Program, Hawaii Institute of Marine Biology, P.O. Box 1106, Kailua, Hawaii 96734-1106

(Received 17 April 2008; revised 9 November 2008; accepted 18 November 2008)

The frequency band that an Atlantic bottlenose dolphin (*Tursiops truncatus*) used to perform an echolocation target discrimination task was determined using computer simulated phantom targets. The dolphin was trained to discriminate frequency filtered phantom targets from unfiltered ones in a go/no-go paradigm. The dolphin's performance indicated perception of echo alteration only when applied filters interfered with the frequency band between 29 and 42 kHz. The dolphin did not behaviorally convey perception of applied filters that affected frequencies outside this functional bandwidth, such as a low pass 43 kHz or a high pass 28 kHz filter. The upper limit of the functional bandwidth at 42 kHz corresponded with the dolphin's upper hearing limit of 45 kHz, as determined through auditory evoked potential measurements. The lower limit of the functional bandwidth corresponded to a drop in intensity below 30 kHz within the dolphin's echolocation clicks. The randomized presentation of different filters showed that the dolphin paid attention to the entire 29–42 kHz band for each trial, not just subsets. The absence of temporal cues between some of the targets the dolphin could discriminate indicated that in these cases the target discrimination cues were based solely on the frequency content.
[DOI: 10.1121/1.3050274]

PACS number(s): 43.80.Ka, 43.80.Ev, 43.80.Jz [JAS] Pages: 1214–1221

I. INTRODUCTION

The underwater hearing ability and frequency perception of marine mammals has become a subject of particular interest due to increased marine mammal interactions with human generated sounds in the ocean (Frisk, 2003). It is essential to understand how these sounds might interfere with dolphin hearing and echolocation so that ways can be developed to minimize this interference. Much work has been done to determine the audiograms of several species of passively listening cetaceans using both behavioral (Au, 2000; Nachtigall *et al.*, 2000) and auditory evoked potential (AEP) methods (Nachtigall *et al.*, 2007; Supin *et al.*, 2001). In a behavioral audiogram the dolphin passively listens for tonal signals of different frequencies and behaviorally indicates perception of the tone. In a common AEP method the dolphin passively listens for amplitude modulated signals and "perception" is determined by monitoring the dolphin's brain activity for repeatable patterns evoked by the amplitude modulation of the carrier frequency (Dolphin *et al.*, 1995; Supin *et al.*, 2001).

However, the active process of echolocation appears to have special requirements for the hearing system that the passive listening system is not necessarily well suited for (Dubrovskiy, 1990). For example, the passive hearing system is evolutionarily attuned to detecting and processing sounds where there is little a priori knowledge about where the sound originated from, what frequencies the sound may contain, or when to anticipate the reception of the sound. However, during active echolocation, the hearing system must be attuned to signals where much more information is known. The spatial origin of an echo can be anticipated since the echolocation beam is highly directional. The general frequency content of the echo can be anticipated since the frequency content of the outgoing click is known. After the first echoes return from the object the time arrival of future echoes can also be anticipated. This led to the concept that dolphins have two listening modes: an active listening mode that depends on the dolphin's own echolocation click production and is attuned to the pulselike nature of echoes that result from those clicks, and a passive mode to handle sounds that are generated from the external environment that can be of much longer duration than an echolocation click (Dubrovskiy, 1990). The perception of sound pressure level provides a good example as to how the two hearing modes can be different. Experiments were conducted with a false killer whale where AEPs were collected during active echolocation at aluminum cylinder targets. Different echo sound pressure levels arriving at the false killer whale's position were created by using a series of targets that varied in target strength and by using different target distances. Even though the echo sound pressures varied, the AEPs had very similar peak heights, indicating that the false killer whale's perception of the sound pressure level could be independent of the actual sound pressure level (Supin *et al.*, 2005). The AEPs of dolphins passively listening to stimulus click signals do not show this sort of constant independence from the signal sound pressure level, but rather follow increases and decreases in the sound pressure level (Ridgway *et al.*, 1981).

[a]Present address: University of California San Diego, Serf Bldg. Room 295 0435, 9500 Gilman Drive, La Jolla, CA 92093. Electronic mail: sibsen@ucsd.edu

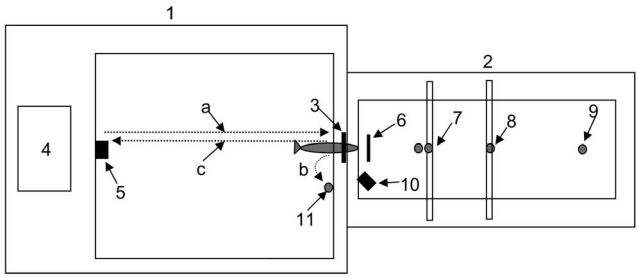


FIG. 1. The experimental setup for the phantom system. The different components are designated in the text.

The perception of echo frequency content by dolphins during active echolocation, as well as how that might differ from the well known frequency perception of dolphins during passive listening, has not yet been studied directly. During active echolocation it is not known if the dolphin pays attention to just a narrow band of frequencies within the spectrum of the pulse, or to the entire available pulse spectrum, within the limits of its hearing range. Determining what frequency regions the dolphin pays attention to while actively echolocating can provide insight into what cues might be used for echolocation discrimination tasks.

The use of computer generated phantom targets in this study provides a unique opportunity to explore these questions. One of the major advantages of the phantom targets over the use of physically real targets is that the frequency content of the echo returned to the dolphin is precisely known for each click. The frequency content of each returned echo is also under experimental control. By using high and low pass filters, the phantom targets can be made very similar or extremely different. The echolocation clicks made by the dolphin while interacting with these targets can be monitored for changes in frequency content between different targets and over time. Using behavioral techniques it can be determined if the dolphin perceived the application of the filters to the echoes. This information can also be used to find the band, or bands, of frequencies the dolphin paid attention to during echolocation (the functional bandwidth).

II. MATERIALS AND METHODS

A. Subject and experimental conditions

These experiments were conducted at the Hawaii Institute for Marine Biology, Marine Mammal Research Program. The subject was an adult female Atlantic bottlenose dolphin (*Tursiops truncatus*) named BJ, who at the time of the experiments in 2004 was 19 years of age. She was housed in a floating wire-net enclosure in Kaneohe Bay, HI. She was fed a daily ration of approximately 6.3 kg total of herring and silver smelt. In 1998 she participated in experiments with the prototype version of the phantom echo generator used here (Aubauer *et al.*, 2000).

The experimental setup was laid out as shown in Fig. 1. The experimental enclosure consisted of two parts. The first was an 8×10 m floating pen frame that had a wire-net bottom (1) and was used to house BJ. The frame consisted of a wooden framework supported on 55 gal barrel floats. The second was a target pen (2) that did not have a wire bottom and was used to house the reception and projection equipment for the phantom target. The wire-net bottom was omitted from the target pen to prevent the production of extra confounding echoes from the wire during echolocation. BJ stationed in a hoop (3) so that her head location was known and so she would have unobstructed echolocation access to the target pen (2). The main computer for the phantom system was set up in an electronic shack (4). This was where the experimenters stayed for the entire length of the session. Its location was chosen to be behind BJ while she was in the hoop to help prevent unintentional visual cues from the researchers. When not in the hoop BJ stationed at a foam pad (5) in front of the electronic shack (4). From this vantage point BJ could see the experimenter. A movable baffle (6) made of 3.2 mm aluminum was used to interfere with BJ's ability to echolocate into the target pen between target presentations. A visual screen made of thin plastic sheet was placed behind the baffle to prevent visual cues from the target pen but allow unobstructed acoustical access. Each click that BJ made was recorded by a hydrophone (7) that was located 2.5 m from the hoop. Each resulting phantom echo was played back to BJ using a transducer (8) that was 5 m from the hoop. An underwater camera (10) was set up to monitor BJ's stationing in the hoop to assure uniformity of her body orientation with respect to the hoop across different trials. The time delay of phantom target echo production was chosen to simulate echoes coming from a target (9) that would have been located 7.6 m from the hoop. Real targets were placed at this location during the training phase of the experiment. The hoop (3), recording hydrophone (7), transducer (8), and targets (9) were all at a depth of 1 m and were aligned to ensure all clicks were recorded along the axis of BJ's transmission beam.

B. Experimental procedure

A trial was initiated with a hand signal to cue BJ to swim from the stationing pad (5) to the hoop (3) along the dotted path (a). BJ swam into the hoop and was cued to begin echolocation when the movable baffle (6) was lowered out of the way giving her free echolocation access to the target pen (2). The baffle remained in the lowered position for 3 s and was then raised. The clicks were recorded on axis by the phantom system and the stimulus phantom echoes were sent back using the transducer (8). A go/no-go paradigm was used where the go response was required when the standard unaltered phantom target was presented and the no-go response was required when the comparison phantom targets were presented. For the go response BJ would back out of the hoop and follow path (b) to touch the response paddle (11) with her rostrum. BJ received a bridge whistle if her response was correct. BJ would then follow path (c) back to the stationing pad (5), receive the fish reward, and would wait for the hand signal to begin the next trial. For the no-go response instead of touching the response paddle BJ was required to stay in the hoop for additional 3 s after the baffle was raised back into place. If her response was correct BJ received a bridge whistle and followed path (c) back to the stationing pad (5), received her fish reward, and waited for

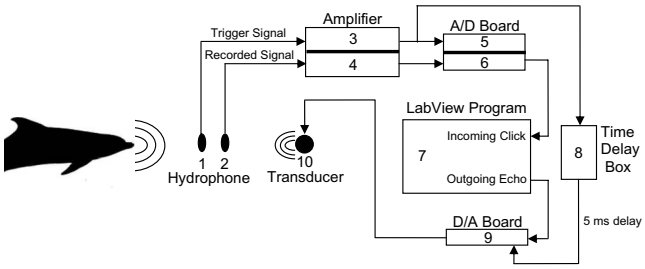


FIG. 2. Schematic of the phantom system. The different components are designated in the text.

the hand signal to begin the next trial. If BJ was incorrect in either a go or no-go trial she was not given a fish reward and was required to return to the stationing pad (5).

One session was typically conducted per day. Each session consisted of 50 trials, 25 of which were the standard unaltered phantom targets and the other 25 trials were comparison phantom targets. Within these 25 comparison trials up to five different types of comparison targets were presented to BJ. The order of stimulus presentation was determined using a modified Gellermann series (Gellermann, 1933).

BJ's baseline performance was maintained during each session because half of the targets were the standard unaltered phantoms. If her baseline performance with the unaltered phantom targets was less than 95% correct within a session it was determined she was not performing reliably. The data for that session were removed from the final data set. This happened for only two sessions.

C. Click recording hardware and software

The phantom system was designed as shown in Fig. 2. The dolphin made a click that was first recorded by an omnidirectional B&K 8103 hydrophone (1), which acted as a triggering hydrophone for the entire system. The triggering signal was amplified 50 dB by a custom built two channel electronic amplifier (3) and sent to a Measurement Computing Corporation PCI-DAS4020/12 analog-to-digital (A/D) board, which digitized the signal at a sample rate of 1 MHz. This triggered the system to begin recording with the second omnidirectional B&K 8103 hydrophone (2). The time delay of click propagation across the 5 cm distance between the triggering and recording hydrophones allowed the system to initialize the recording hydrophone in time to collect the entire click. The recorded click signal was then amplified 36 dB in the second channel of the electronic amplifier (4) and recorded by the second channel on the A/D board using a sample rate of 1 MHz (6). A custom written LABVIEW program convolved the click signal with the transfer function of the phantom target (7) (Aubauer and Au, 1998). The resulting phantom echo was then stored in the onboard memory of a Strategic Test UF6011 arbitrary waveform generator board (9) where it was held until triggered by a custom built time delay box (8) to output the echo to an International Transducer Corporation ITC-1042 transducer (10) with a sample rate of 1 MHz. The signal collected by the triggering hydrophone (1) also triggered the time delay box to produce its

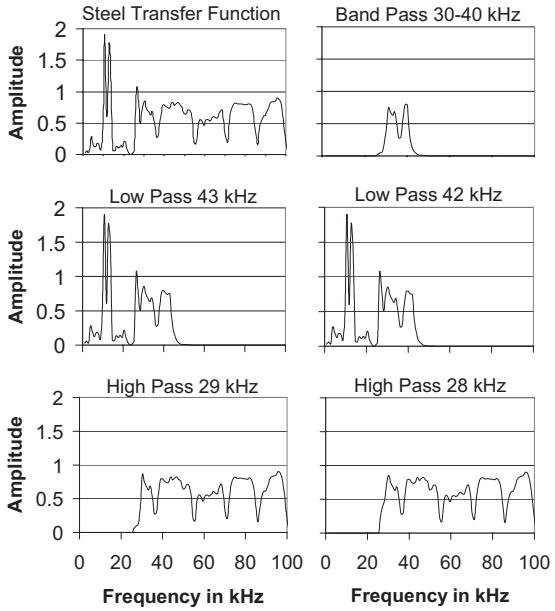


FIG. 3. The frequency spectra of the phantom steel target transfer functions that were modified with different filter types as labeled. The standard phantom steel target is shown in the upper left corner.

trigger pulse 5 ms later. The 5 ms timing was set to replicate the arrival time to BJ of an echo produced by a real target located 7.6 m in front of the hoop. The time delay box had a more consistent time delay of 5 ms than could be achieved using the computer's own clock. This ensured greater uniformity of the time delay between echo production and the initial click trigger. The resulting echo then propagated back from the transducer (10) to the dolphin.

The system could handle up to 30 clicks/s using this algorithm without significantly skipping clicks. The dolphin usually made 20–25 clicks/s. The convolution of each click individually with the target transfer function allowed the system to produce phantom targets that were as flexible as the real targets in response to click to click changes in intensity and frequency content.

D. Phantom targets

During the training phase of the experiment only real targets were used. The standard was a 7.6 cm diameter solid stainless steel sphere and the comparison was a 7.6 cm diameter solid brass sphere. The stimuli that were presented to BJ during the actual data collection sessions consisted entirely of phantom targets that simulated echoes from the 7.6 cm diameter solid stainless steel sphere. These were the same spherical targets used in the phantom experiment of Aubauer and Au (1998) where it was demonstrated that the echo parameters of the spheres were sufficient for BJ to perceive and discriminate. The radial symmetry of the sphere eliminated variations in the echo due to rotations. The phantom system was designed to reproduce the sphere echoes in both frequency content and intensity. The standard stimulus was an unfiltered version of the phantom steel target. The comparison stimuli were filtered versions of the phantom steel target and an unfiltered version of the brass target. Figure 3 shows the amplitude spectra of the selected phantom

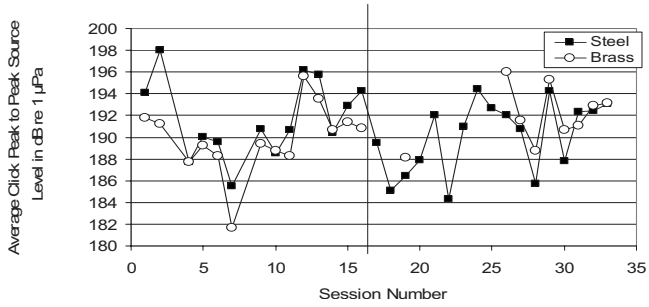


FIG. 4. The average of BJ's click peak to peak source levels for each session. Comparing the source levels made when BJ was exposed only to real targets (left of the black line) with source levels made when she was exposed only to phantom targets (right of the black line) indicated that she used similar click intensities between the real and phantom targets. Not all sessions used the brass phantom target.

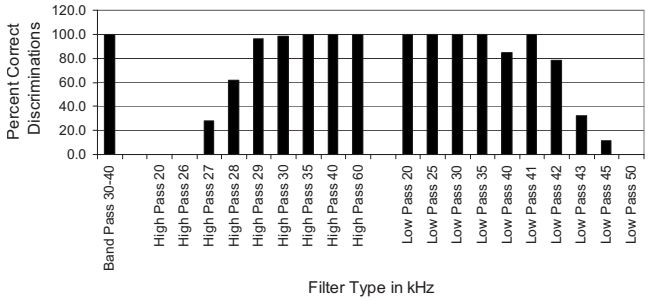


FIG. 5. BJ's behavioral performance with the different filtered phantom targets. The threshold for perception of the applied filter was 70%. Based on that criterion, BJ's functional bandwidth was between 29 and 42 kHz. She paid attention to this entire frequency band while echolocating.

steel target transfer functions that were altered with the filters as labeled. The complete set of filters used included high pass filters at 60, 40, 35, 30, 28, 27, 26, and 20 kHz and low pass filters at 50, 45, 42, 41, 40, 35, 30, 25, and 20 kHz and a bandpass filter from 30 to 40 kHz. The rolloff slopes of both the high pass and low pass filters were 87 dB/octave. These transfer functions were convolved with the incoming clicks to create the comparison filtered phantom steel target echoes. The transfer function in the upper left hand corner of Fig. 3 was the unfiltered stainless steel sphere transfer function and was used to create the standard unfiltered phantom target echoes. These filters were appropriate for the types of clicks that BJ was producing because the average peak frequency of her clicks was 40 kHz, see Fig. 8(a). This average peak frequency is lower than what is typically observed for *Tursiops* and is discussed further in Ibsen *et al.*, 2007.

III. RESULTS

A. The dolphin's interaction with the system

In the phantom echo experiment of Aubauer and Au (1998) it was demonstrated that BJ accepted the phantom targets and classified them as she did the real targets. It was essential to confirm that BJ accepted the phantom targets produced by the new upgraded phantom echo system created for this experiment. A comparison was made between BJ's clicks produced during training sessions where only real targets were presented and sessions where only the phantom versions of the targets were presented.

The first parameter considered was the click peak to peak source level. The average source level used in each session is shown in Fig. 4. The first 16 sessions had only real steel and brass target presentations. Sessions 17–33 consisted of phantom steel and phantom brass target presentations. The clicks made during exposure to the steel targets were considered separately from those made during exposure to the brass targets to prevent complications from different target types. These two groups of real and phantom sessions were considered two populations. The following statistical analyses all used a nonparametric Wilcoxon rank sum test performed at a 0.05 significance level. A nonparametric test was chosen because the populations did not have normal distributions. Not all sessions used the brass phantom target. The average click intensities for real and phantom brass targets were not significantly different ($p=0.18$) and neither were the average click intensities for the real and phantom steel targets ($p=0.35$).

It is interesting to note that BJ used similar click intensities for both the steel and brass targets within the same session, despite the variation in average click intensity between sessions. This indicated that in addition to the natural variation in the dolphin's behavior the click intensity levels might have been influenced more by the ambient background noise level, which could change from day to day, than by the type of target that was presented. This pattern was consistent for both real and phantom target sessions.

The second parameter considered between the two groups was the click peak frequency. The distributions of peak frequencies from each session were considered rather than the averages due to the high variability in between clicks. The maxima of the distributions were between 40 and 50 kHz for all the sessions and targets. The width of the distributions was characterized by calculating the rms value. Using the same Wilcoxon rank sum test as before, the rms values of the real and phantom brass distributions were not significantly different ($p=0.055$) and neither were the real and phantom steel distributions ($p=0.097$).

The similarities of click intensity and peak frequency distribution between real and phantom targets showed that BJ did not change the types of clicks used when interacting with the phantom system. This indicated that BJ was doing a real echolocation task when interacting with the phantom system.

It was important that the phantom echo generator process incident clicks quickly enough to ensure that each click had a corresponding phantom echo. BJ usually made 20–25 clicks/s and the system could handle up to 30 clicks/s without skipping.

B. Functional bandwidth

BJ's behavioral performance with the different filters is summarized in Fig. 5. The bar graph shows the percentage of discriminations BJ got correct with the different filtered targets. The cutoff criterion for successful identification of the filter was 70% correct. Based on that criterion, BJ's functional bandwidth was between 29 and 42 kHz.

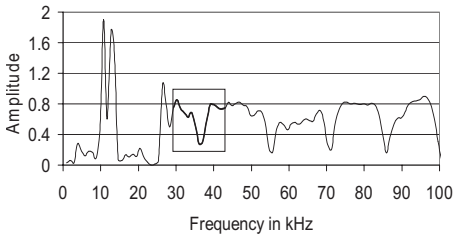


FIG. 6. The steel sphere transfer function showing the standard target's reflection characteristics. BJ's functional bandwidth is shown inside the box. BJ paid attention only to this frequency band while echolocating and paid attention to the entire band during each target presentation.

Random guessing would yield a 50% correct level. Those filtered targets that had less than 50% correct meant that BJ could not perceive the presence of the filter and mostly identified it as the unfiltered standard. All filters were presented a minimum of 50 times except for the high pass 20 kHz, high pass 26 kHz, low pass 45 kHz, and low pass 50 kHz. These filters were presented less than 50 times each because continued presentation of these filters to BJ resulted in behavioral breakdown. It was clear that BJ did not perceive frequency changes in these regions based on her behavior, performance, and inability to perceive the high pass 27 kHz and low pass 43 kHz filters.

The low pass 40 kHz filtered target had a lower than 100% correctness as shown in Fig. 5 because BJ incorrectly accepted it as the standard target for the first nine presentations. After that she was able to correctly identify this filtered phantom target as a comparison for the next 51 presentations. It was this target that might have taught BJ to make her discriminations based on fine frequency differences.

IV. DISCUSSION

A. Functional bandwidth

It was not known before this study what actual band of frequencies any odontocete used during echolocation. It was also not known if the entire passive hearing range was utilized for echolocation. The information in Fig. 5 indicates that during echolocation BJ paid attention only to frequencies between 29 and 42 kHz making this region her functional bandwidth. This limited frequency range was a small portion of the reflection from the phantom target as shown within the boxed region of Fig. 6.

The upper limit of BJ's functional bandwidth at 42 kHz corresponds well with the upper limit of her hearing range. AEP audiograms for BJ were collected in 2001 and in 2005 and are shown in Fig. 7. BJ's hearing began to rapidly decline above 40 kHz. The AEP audiogram was measured in Kaneohe Bay, where BJ was masked by the background snapping shrimp noise. Therefore, the apparent overall sensitivity was much lower than the sensitivity measured for young *Tursiops* in a quiet tank by Johnson (1967). This establishes that BJ used the same upper hearing limit during the active echolocation process and also during passive hearing. This has not been demonstrated before since the relationship between the active and passive hearing is not well known.

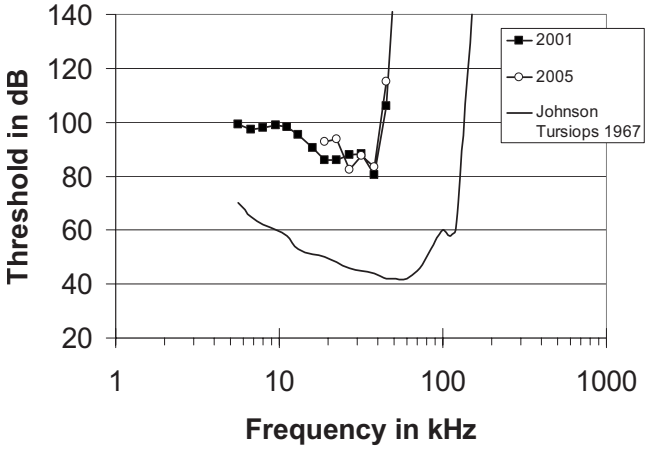


FIG. 7. BJ's AEP audiograms collected in both 2001 and 2005. BJ's hearing rapidly decreased after 45 kHz. Johnson's (1967) bottlenose dolphin audiogram is included for comparison.

The lower limit of BJ's functional bandwidth seems to be a function of several different factors including background noise level, target reflection, and the frequency content of the clicks themselves. Figure 8 shows normalized graphs of the spectrum from the clicks, the steel target transfer function, and the background noise. The background noise was collected at random times from 1 m in front of the hoop using the same B&K 8103 signal collecting hydrophone used in the phantom system. BJ's functional bandwidth was the frequency region between the two black lines. The lower limit of the functional bandwidth occurs in a region where the frequency content of the click begins to drop

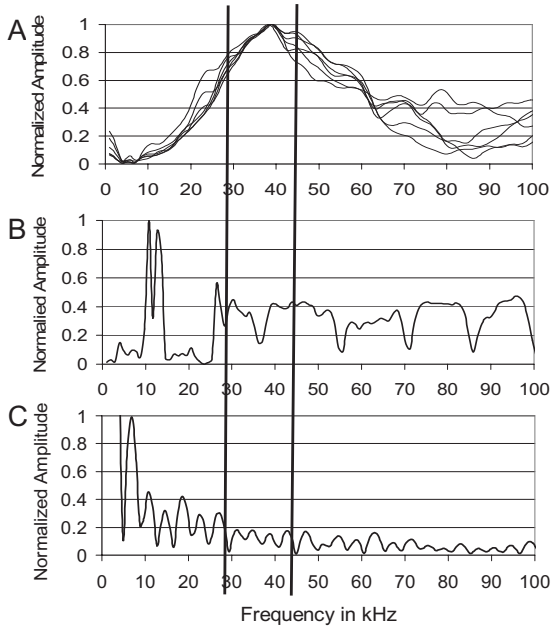


FIG. 8. Spectral analysis of (a) six characteristic click spectra, (b) the steel transfer function, and (c) the background noise spectra of the bay. The region between the two black lines was BJ's functional bandwidth, the frequency band she paid attention to while echolocating. BJ's click was designed to maximize frequency intensity in her functional bandwidth, reflect well off the phantom target, and take advantage of a region of frequencies that had the lowest background noise level she could perceive. All these factors were optimized aspects of her echolocation system to maximize the perceived signal to noise ratio of the target echoes.

off and the background noise level begins to rise. The frequency reflection off the target itself begins to drop off significantly in the region as well. These three factors all contribute to create the lower limit of the functional bandwidth.

It can also be concluded from this study that BJ paid attention to the entire 29–42 kHz band and not a subset of this band. BJ could not anticipate the frequency content of the target she would be presented since the filters were presented in a random fashion. This means that BJ had to be paying attention to the entire 29–42 kHz functional bandwidth during each target presentation. For example, if BJ could anticipate that the target was a low pass 30 kHz filter, she would only have to pay attention to frequencies above 30 kHz. However, since the presentation order was random, she had to pay attention to the presence or absence of all the frequencies in the 29–42 kHz band for each trial. To achieve nearly 100% accuracy with the discrimination task, BJ could not have consistently concentrated on a subrange of the 29–42 kHz band because the high and low pass filters covered the whole band.

With BJ's restricted hearing range the functional bandwidth was pretty well determined by the upper hearing limit and the lower frequency content of the clicks. When studying a dolphin with a much larger hearing range it would be considerably more difficult to estimate the functional bandwidth. The phantom techniques described here could be used to determine a dolphin's functional bandwidth as a general phenomenon.

The functional bandwidth encompassed the frequencies in her clicks that had the highest intensities and that were reflected by the phantom target with high intensity. In addition, the functional bandwidth had relatively the lowest background noise level within BJ's hearing range. These patterns indicate that BJ's functional bandwidth was a frequency region that maximized the signal to noise ratio for the system, which included her click production, target reflectance, background noise, and her upper hearing limit.

Frequencies outside the functional bandwidth did not contribute to higher echo signal to noise ratios. For example, the transfer function between 25 and 28 kHz reflected 18% more energy than any 3 kHz band between 29 and 42 kHz but this frequency region had lower intensity from the click and had a 50% higher background noise level. The region between 10 and 15 kHz had very high reflectance from the phantom target, but the intensity of the dolphin's click was very low in that region causing the resulting echoes to have negligible frequency content in that band.

The average rms bandwidth of the most commonly used clicks during this experiment was 67 kHz. This was far wider than BJ's 13 kHz functional bandwidth. BJ could better optimize her energy expenditure by having both her click rms bandwidth and functional bandwidth the same size and perfectly overlapping in spectrum. However, this sort of spectral matching may not be possible. Although BJ has considerable control over the frequency content of her echolocation clicks she is still bound by the physical limitations of click production, which may not make it possible for a click

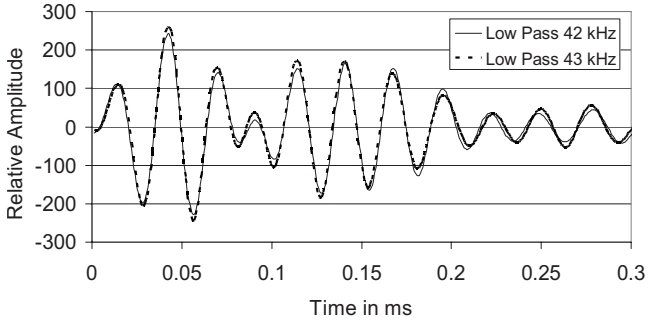


FIG. 9. The time domain waveform of the echoes from the low pass 42 kHz and low pass 43 kHz filtered phantom steel targets. There was a difference in these two signals that BJ was able to use as a cue to discriminate between them. That cue could not have been the timing between the highlights because the timing was the same.

to be produced that has a 13 kHz bandwidth. The clicks BJ made were probably the best she could produce in terms of spectral matching.

The lower limit of the audiogram in Fig. 7 indicates that BJ could still hear frequencies between 10 and 20 kHz despite this being a region she did not use for echolocation. BJ probably could not use this region for echolocation because no click ever recorded from BJ contained significant energy there, possibly due to physical limitations in the click production mechanisms themselves. BJ's ability to hear below 20 kHz probably played a greater role in her passive hearing than it did in her active echolocation.

B. Possible discrimination cues

Establishing BJ's functional bandwidth provides clues about the cues she used to make the target discriminations. BJ's ability to discriminate clearly between phantom targets that differed in frequency content by as little as 1 kHz indicates that frequency content was a powerful cue for these discrimination tasks. Although a dolphin will probably use any available cue to perform a discrimination task, including time cues such as time separation pitch, there were certain cases during this experiment where frequency content was the main difference between targets. This was demonstrated with the low pass 43 kHz filtered and low pass 42 kHz filtered targets. BJ accepted the low pass 43 kHz target as be-

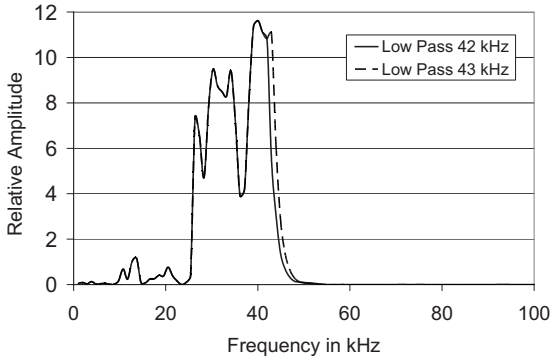


FIG. 10. The amplitude spectra of the echoes from the low pass 42 kHz and low pass 43 kHz filtered phantom steel targets. BJ was able to discriminate between these two targets despite their very similar amplitude spectra profiles.

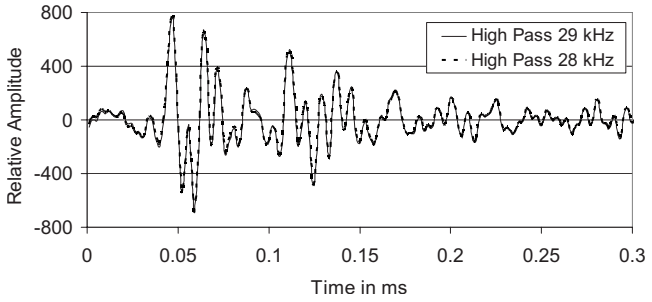


FIG. 11. The time domain waveform of the echoes from the high pass 29 kHz and high pass 28 kHz filtered phantom steel targets. There was a difference in these two signals that BJ was able to use as a cue to discriminate between them. That cue could not have been the timing between the highlights because the timing was the same.

ing the standard but rejected the low pass 42 kHz indicating that these two were perceived as being different targets. A characteristic click was used to calculate the time waveforms of echoes produced from these filters. These echoes are plotted together in Fig. 9 showing that the time separation between the highlights of these two targets was the same. The amplitude spectra of the same two echoes are graphed together in Fig. 10. They show slight changes due to the different degrees of filtering. These frequency differences provided the largest cues for this discrimination.

A similar trend was found when looking at the high pass 28 kHz and high pass 29 kHz filters. BJ accepted the high pass 28 kHz filter as the standard but rejected the high pass 29 kHz filter indicating she could perceive a difference between them. Time domain waveforms of echoes from each target are shown in Fig. 11 and their amplitude spectra in Fig. 12. As seen before, the time separation of the highlights was the same between the targets causing the main cue to come from the small difference in frequency content.

V. CONCLUSION

The functional bandwidth of an echolocating bottlenose dolphin was determined using phantom echo techniques. The dolphin was trained to discriminate computer generated phantom targets that had been frequency filtered to different degrees. The results of the behavioral experiments indicated that the dolphin paid attention only to the frequencies between 29 and 42 kHz and paid attention to that entire frequency band for each trial. The upper limit of the functional bandwidth corresponded with the dolphin's upper hearing limit. The lower limit of the functional bandwidth corresponded to a drop in the intensity of the echolocation click itself around 30 kHz. It appears that the dolphin was optimizing her echolocation click to maximize signal to noise ratio considering the background noise levels, the reflection characteristics of the targets, and her upper hearing limit. This optimization was done within the physical constraints of her click production system. It was also found that the dolphin's passive hearing range was greater than the functional bandwidth.

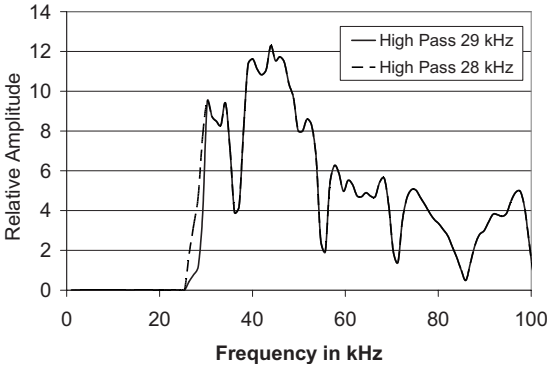


FIG. 12. The amplitude spectra of the echoes from the high pass 29 kHz and high pass 28 kHz filtered phantom steel targets. BJ was able to discriminate between these two targets despite their very similar frequency spectra profiles.

The dolphin was able to discriminate between phantom targets that differed by as little as 1 kHz in frequency content. In several of these discriminations the time separation pitch was the same between these targets and so could not have been used as a cue. The slight difference in frequency content alone was the most likely cue.

ACKNOWLEDGMENTS

The authors are grateful for the support of Dera Look, Vincent De Paolo, Kristen A. Taylor, Michelle Yuen, T. Aran Mooney, and the Marine Mammal Research Program of the Hawaii Institute of Marine Biology at the University of Hawaii. The study was supported by the Office of Naval Research Grant No. N00014-98-1-0687, for which the authors thank Robert Gisiner. This work was conducted under Marine Mammal Permit No. 978-1567 issued to P.E.N. by the NMFS NOAA Office of Protected Resources.

Au, W. W. L. (**2000**). "Hearing in whales and dolphins: An overview," in *Hearing by Whales and Dolphins*, edited by W. W. L. Au, A. N. Popper, and R. R. Fay (Springer-Verlag, New York), pp. 1–42.

Aubauer, R., and Au, W. W. L. (**1998**). "Phantom echo generation: A new technique for investigating dolphin echolocation," J. Acoust. Soc. Am. **104**, 1165–1170.

Aubauer, R., Au, W. W. L., Nachtigall, P. E., Pawloski, D. A., and DeLong, C. M. (**2000**). "Classification of electronically generated phantom targets by an Atlantic bottlenose dolphin (*Tursiops truncatus*)," J. Acoust. Soc. Am. **107**, 2750–2754.

Dolphin, W. F., Au, W. W. L., Nachtigall, P. E., and Pawloski, J. (**1995**). "Modulated rate transfer functions to low-frequency carriers in three species of cetacean," J. Comp. Physiol., A **177**, 235–245.

Dubrovskiy, N. A. (**1990**). "On the two auditory subsystems of dolphins," in *Sensory Ability of Cetaceans: Laboratory and Field Evidence*, edited by J. A. Thomas and R. Kastelein (Plenum, New York), pp. 233–254.

Frisk, G. (**2003**). "Ocean noise and marine mammals," Ocean Studies Board The National Research Council.

Gellermann, L. W. (**1933**). "Chance orders of alternating stimuli in visual discrimination experiments," J. Gen. Psychol. **42**, 206–208.

Ibsen, S., Au, W., Nachtigall, P., Delong, C., and Breese, M. (**2007**). "Changes in signal parameters over time for an echolocating Atlantic bottlenose dolphin performing the same target discrimination task," J. Acoust. Soc. Am. **122**, 2446–2450.

Johnson, S. (**1967**). "Sound detection thresholds in marine mammals," in *Marine Bioacoustics*, edited by W. N. Tavolga (Pergamon, New York), Vol. **2**, pp. 247–260.

Nachtigall, P. E., Lemonds, D. W., and Roitblat, H. L. (**2000**). "Psychoacoustic studies of whale and dolphin hearing," in *Hearing by Whales and Dolphins*, edited by W. W. L. Au, A. N. Popper, and R. R. Fay (Springer-Verlag, New York), pp. 330–364.

Nachtigall, P. E., Mooney, T. A., Taylor, K. A., and Yuen, M. L. (**2007**). "Hearing and auditory evoked potential methods applied to odontocete cetaceans," Aquat. Mamm. **33**, 6–13.

Ridgway, S., Bullock, T., Carder, D., Seeley, R., Woods, D., and Galambos, R. (**1981**). "Auditory brainstem response in dolphins," Proc. Natl. Acad. Sci. U.S.A. **78**, 1943–1947.
Supin, A., Nachtigall, P., Au, W., and Breese, M. (**2005**). "Invariance of evoked-potential echo-responses to target strength and distance in an echolocating false killer whale," J. Acoust. Soc. Am. **117**, 3928–3935.
Supin, A., Popov, V. V., and Mass, A. M. (**2001**). *The Sensory Physiology of Aquatic Mammals* (Kluwer Academic, Boston, MA).

Underwater detection of tonal signals between 0.125 and 100 kHz by harbor seals (*Phoca vitulina*)

Ronald A. Kastelein,[a] Paul J. Wensveen, and Lean Hoek
Sea Mammal Research Company (SEAMARCO), Julianalaan 46, 3843 CC Harderwijk, The Netherlands

Willem C. Verboom
Acoustic Consultancy, Junostraat 10, 2402 BH, Alphen a/d Rijn, The Netherlands

John M. Terhune
Department of Biology, University of New Brunswick, P.O. Box 5050, Saint John, New Brunswick E2L 4L5, Canada

(Received 19 May 2008; revised 30 October 2008; accepted 20 November 2008)

The underwater hearing sensitivities of two 1-year-old female harbor seals were quantified in a pool built for acoustic research, using a behavioral psychoacoustic technique. The animals were trained to respond when they detected an acoustic signal and not to respond when they did not (go/no-go response). Pure tones (0.125–0.25 kHz) and narrowband frequency modulated (tonal) signals (center frequencies 0.5–100 kHz) of 900 ms duration were tested. Thresholds at each frequency were measured using the up-down staircase method and defined as the stimulus level resulting in a 50% detection rate. The audiograms of the two seals did not differ statistically: both plots showed the typical mammalian U-shape, but with a wide and flat bottom. Maximum sensitivity (54 dB *re* 1 μPa, rms) occurred at 1 kHz. The frequency range of best hearing (within 10 dB of maximum sensitivity) was from 0.5 to 40 kHz ($6\frac{1}{3}$ octaves). Higher hearing thresholds (indicating poorer sensitivity) were observed below 1 and above 40 kHz. Thresholds below 4 kHz were lower than those previously described for harbor seals, which demonstrates the importance of using quiet facilities, built specifically for acoustic research, for hearing studies in marine mammals. The results suggest that under unmasked conditions many anthropogenic noise sources and sounds from conspecifics are audible to harbor seals at greater ranges than formerly believed.
 [DOI: 10.1121/1.3050283]

PACS number(s): 43.80.Lb, 43.80.Nd [WWA] Pages: 1222–1229

I. INTRODUCTION

The harbor seal (*Phoca vitulina*) has the most extensive geographic distribution of any seal species. It inhabits the eastern Baltic Sea as well as both eastern and western coasts of the Atlantic (30° to 80° north) and Pacific (28° to 62° north) Oceans. It leads an amphibious life, resting and pupping on land, while migration, foraging, and courtship occur underwater (Burns, 2002). During the breeding season, male harbor seals produce underwater broadband pulsed vocalizations, termed roars, that have their major sound energy between 0.4 and 2 kHz and durations of 1–10 s (Schusterman *et al.*, 1970; Van Parijs and Kovacs, 2002).

To determine the importance of sound for harbor seals during activities such as communication, reproduction, predator avoidance, and navigation, and the potential for disturbance by anthropogenic noise, information is needed on the species' underwater hearing sensitivity. This has been tested for pure tones (Møhl, 1968; Terhune, 1988, 1989; Turnbull and Terhune, 1993; Kastak and Schusterman, 1998; Southall *et al.*, 2005) and frequency swept tones (Turnbull and Terhune, 1994). However, in each of the seven studies, only the sensitivity of a single harbor seal over a part of the frequency range of hearing was investigated. Moreover, in each study different equipment, methodology, and signal parameters were used and the animals were of different ages (but all were males). In addition, some of the hearing thresholds may have been influenced (masked) by background noise in the research pool. Because of the differences in these studies it is difficult to construct an appropriate composite audiogram to describe the hearing capabilities of harbor seals.

Many human activities occur in the coastal waters where harbor seals are found. To assess potential disturbance by anthropogenic noises, it is important to obtain robust underwater hearing threshold curves for this pinniped species. For this, a quiet testing environment and multiple representative study animals are needed. Therefore, a pool and filtration system with special acoustic features designed for hearing studies was built at a quiet location in the Netherlands. Two young healthy female harbor seals were obtained specifically for this hearing study. Our aim was to determine absolute (unmasked) underwater hearing thresholds for both seals over their entire hearing range.

[a]Author to whom correspondence should be addressed. Electronic mail: researchteam@zonnet.nl

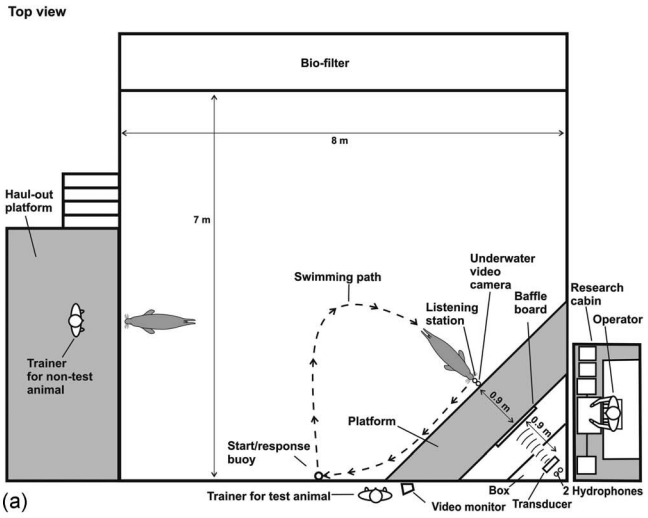


FIG. 1. The study area, showing the test harbor seal in position at the underwater listening station, and the nontest animal with the other trainer; (a) top view and (b) side view, both to scale.

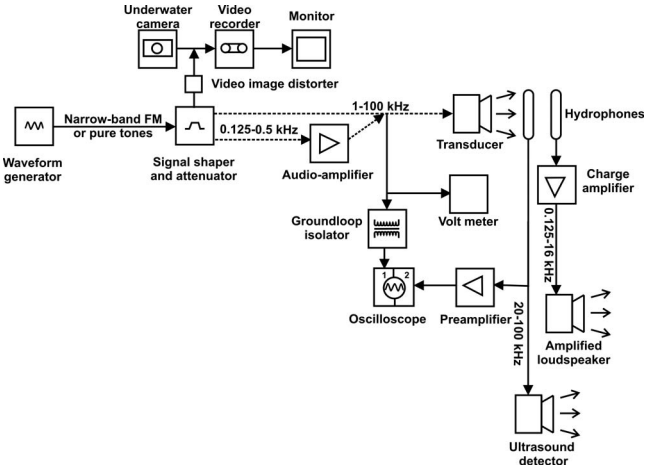


FIG. 2. Block diagram of the transmitting and listening systems.

II. MATERIALS AND METHODS

A. Study animals

The study animals were two female harbor seals (identified as SM.Pv.01 and SM.Pv.02), which were born at Ecomare, Texel, The Netherlands. The animals were moved to the research facility soon after they had been weaned. Throughout the study, the animals were healthy. They were not exposed to ototoxic medication prior to or during the study period. During the study they aged from 14 to 18 month old and their body weight increased from around 34 to around 42 kg. The seals consumed between 1.4 and 1.8 kg of thawed fish (herring, *Clupea harengus*, mackerel, *Scomber scombrus*, and sprat, *Sprattus sprattus*) divided into four meals per day. In general, the seals received most of their daily ration during research sessions.

B. Study area and staff

The study was conducted at SEAMARCO's Research Institute, The Netherlands, which is in a remote area that was specifically selected for acoustic research. The measurements were conducted in an outdoor pool [8 m(l)×7 m(w), 2 m deep], with an adjacent haul-out platform (Fig. 1). The pool walls and floor were made of plywood covered with polyester. The pool floor was 1 m below ground level. To reduce sound reverberation in the pool, the inner walls were covered with 3-cm-thick mats of coconut fiber embedded in 4-mm-thick rubber. The coconut mats extended 10 cm above the water level to reduce splashing noises caused by waves. The bottom of the pool was covered with approximately 20 cm of sand. Skimmers kept the water level constant. Seawater was pumped in directly from the nearby Oosterschelde, a lagoon of the North Sea. Most of the water (80%) was recirculated daily through a biological filter system to ensure year-round water clarity, so that the animals' behavior could be observed via an underwater camera during the test sessions.

To limit the amount of noise that the seals were exposed to on a regular basis (to prevent a temporary hearing threshold shift before the hearing tests), the water circulation system and aeration system for the adjacent biofilter were designed to be as quiet as possible. This was done by choosing "whisper flow" pumps, mounting the pumps on rubber blocks, and connecting the pumps to the circulation pipes with very flexible rubber hoses. There was no current in the pool during the experiments, as the water circulation pump and the air pump of the biofilter were switched off for 10 min before and during test sessions. This also prevented flow noise from the skimmers. The water temperature varied between 20 °C in August and 12 °C in November, and the salinity was around 3.4%.

During the 15 min test sessions the animal not being tested was trained to keep very still in the water next to the haul-out area (this was quieter than staying on land, where a scratch of a flipper nail could trigger a prestimulus response in the animal being tested). The signal operator and the equipment used to produce the stimuli and listen to underwater sounds were in a research cabin next to the pool, out of sight of the animals (Fig. 1).

C. Test stimuli

A schematic of the equipment used to configure and emit outgoing signals is shown in Fig. 2. All stimuli were produced by a waveform generator (Hewlett Packard, model 33120A). Two types of tonal signals were used. Between 0.125 and 0.25 kHz, pure tones were used because due to the size of the pool the chances of standing waves occurring were small for these long wavelengths. Between 0.5 and

100 kHz, narrowband sinusoidal frequency-modulated (FM) tonal signals (with center frequencies of 0.5, 1, 2, 4, 8, 16, 25, 31.5, 40, 50, 63, 80, and 100 kHz) were used. The modulation range of the signals was ±1% of the center frequency (the frequency around which the signal fluctuated symmetrically), and the modulation frequency was 100 Hz (for example, if the center frequency was 10 kHz, the frequency fluctuated 100 times per second between 9.9 and 10.1 kHz). Narrowband FM signals were used above 0.25 kHz because such signals produce fewer constructive and destructive interference effects (standing waves) in a reverberant pool than pure tones (Kastelein *et al.*, 2002, 2005; Finneran and Schlund, 2007).

A modified audiometer for testing human aerial hearing (Midimate, model 602) was used to control the duration and amplitude of signals. The stationary portion of all signals was 900 ms in duration; the rise and fall times were 50 ms to prevent transients. Hearing thresholds depend on signal duration, and integration time is also frequency dependent, decreasing with increasing frequencies (Terhune, 1988). The 900 ms signal duration used in the present study is probably above the integration time of the harbor seal's hearing system. The sound pressure level (SPL) at the seal's head while it was at the listening station could be varied in 5 dB increments (this step size was determined by the audiometer: 5 dB steps are generally used in human audiometry). The 0.125–0.5 kHz signals from the audiometer were amplified by means of an audio amplifier (Sony TA-F335 R).

A directional transducer (Ocean Engineering Enterprise, model DRS-12; 30 cm diameter) was used to project the signals into the water (an impedance matching transducer was not used, in order to eliminate harmonics). Multipath arrivals and standing waves can introduce both temporal and spatial variations in the observed SPL at the listening station. Therefore, the transducer was placed in a corner of the pool in a protective wooden box lined with sound-absorbing rubber. The transducer was hung with four nylon cords from the cover of the box and made no contact with the box. A stainless steel weight was fixed to the lower part of the transducer to compensate for its buoyancy. The transducer was 1.85 m from the tip of the L-shaped listening station (Fig. 1), and was positioned so that the acoustic axis of the projected sound beam pointed at the center of the listening station (i.e., the center of the study animal's head while it was at the listening station). To reduce reflections from the bottom of the pool and water surface reaching the listening station, a baffle board was placed halfway between the transducer and the animal. The board consisted of 2.4 m high, 1.2 m wide 4 cm thick plywood, covered with a 2 cm thick closed cell rubber mat on the side facing the transducer. A 30-cm-diameter hole was made in the board with its center at the same level as the seal's head and the transducer (1 m below the water surface). As an indicator of the condition of the transducer, its capacitance was checked once a week with a capacity meter (SkyTronic 600.103). During the study period the capacity remained constant.

D. Stimuli level calibration and background noise measurement

Audiograms are easily influenced by background noise in the test area. Therefore, great care was taken to make the seal's listening environment as quiet as possible. Nobody was allowed to move within 15 m of the pool during sessions. Underwater background noise levels were measured monthly during the study period, under the same conditions as during the test sessions (i.e., in various weather conditions but without rain and with wind speed below Beaufort 3).

The equipment used to measure the background noise in the pool consisted of a hydrophone [Bruel & Kjær (B&K) 8101], a voltage amplifier system (TNO TPD, 0–300 kHz), and a dual spectrum analyzer system (0.025–160 kHz). The system was calibrated with a pistonphone (B&K 4223) and a white noise signal (0.025–40 kHz) which was inserted into the hydrophone preamplifier. Measurement results were corrected for the frequency sensitivity of the hydrophone and the frequency response of the measurement equipment. The customized analyzer consisted of an A/D-converter (Avisoft UltraSoundGate 116; 0–250 kHz) coupled to a notebook computer (sampling rate: 500 kHz). The digitized recordings were analyzed by two parallel analysis systems: (1) a fast Fourier transform narrow-band analyzer (0.025–160 kHz) and (2) a 1/3-octave band analyzer (0.025–160 kHz).

1/3-octave band background noise levels were determined in the range 0.025–100 kHz and converted to 'equivalent sound pressure spectrum levels' (L_{eq} method, Hassall and Zaveri, 1988), expressed in dB *re* 1 μPa/$\sqrt{\text{Hz}}$. Figure 3(a) shows the power averaged ($n=5$) background noise in the pool, alongside the self-noise of the measuring system.

The received SPL (dB *re* 1 μPa, rms) of each stimulus was measured approximately once each month at the seals' head position (Fig. 1). During trials, the seals' head positions (while at the listening station) were carefully monitored and were consistent to within a few cm. The received SPL variation between calibration sessions was frequency dependent. The deviation from the mean was generally around 2 dB for all frequencies, except 31.5 kHz, where for unknown reasons the deviation from the mean was around 5.5 dB. No harmonic distortions were present in the test frequencies at the SPLs used in the hearing tests. The linear averaged received SPL per test frequency was calculated from all five calibration sessions. The means were used to determine the session thresholds.

The received SPLs were calibrated at a level of 14–65 dB (depending on frequency) above the threshold levels found in the present study. The linearity of the transmitter system was checked during the study by measuring levels around 15 dB above the thresholds found in this study, and it was consistent within a few decibels. At one frequency (0.125 kHz) the stimulus tone levels could not be amplified sufficiently above background noise to ensure that the measurements were not influenced by the background noise. The SPL was also measured 10 cm in all directions from the listening station (the location of the seal's head), and varied

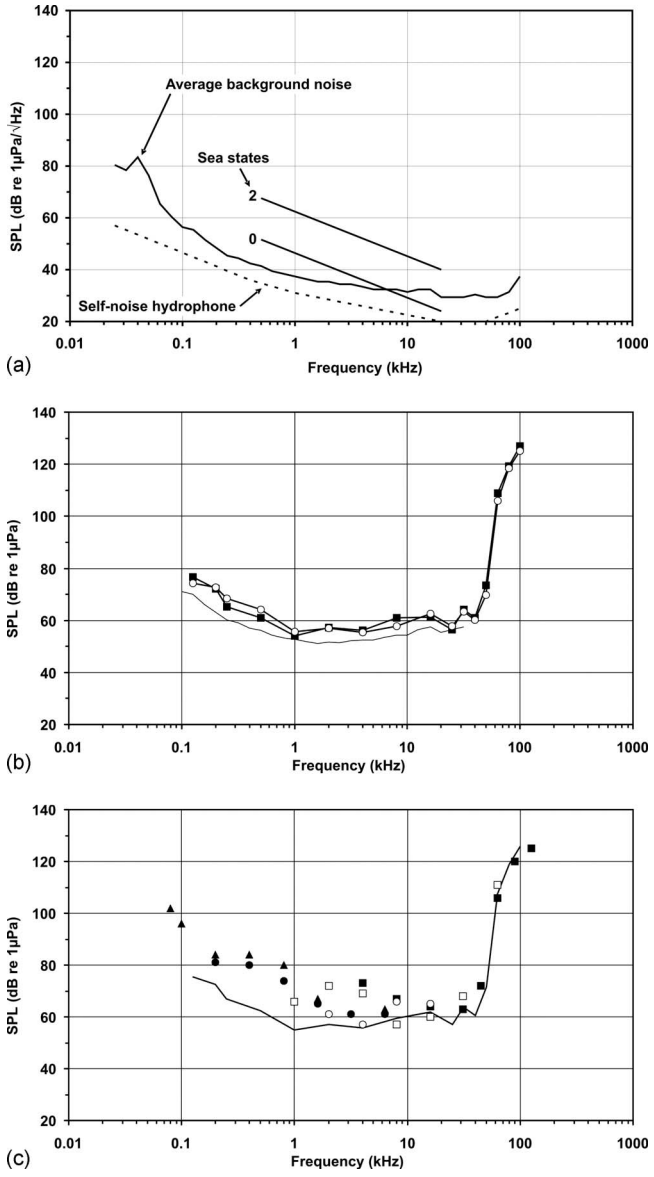


FIG. 3. (a) Power averaged background noise level in the pool in dB re 1 μPa/$\sqrt{}$Hz (derived from 1/3-octave band levels; $n=5$) and the self-noise of the B&K 8101 hydrophone (amplifier). Also shown are noise levels at sea measured at sea states 0 and 2 (Knudsen *et al.*, 1948). (b) The mean 50% detection thresholds (dB re 1 μPa, rms, line level) for pure tone and narrowband FM (900 ms) signals obtained for female harbor seals 01 (■) and 02 (○) in the present study (for details see Table I). The thin line shows the calculated noise-limited theoretical detection threshold (dB re 1 μPa) based on the background noise levels of Fig. 3(a) and CRs of harbor seal hearing. (c) The average underwater hearing threshold (in dB re 1 μPa, rms) of the two study animals in the present study, shown as a line, and the underwater hearing thresholds found for harbor seals in previous studies [Møhl, 1968 (±500 ms, ■); Terhune, 1988 (500 ms, □); Turnbull and Terhune, 1993 (repeated signals, 50 ms, 10/s, ○); Kastak and Schusterman, 1998 (500 ms, ▲); Southall *et al.*, 2005 (500 ms, ●)]. The numbers between brackets indicate the signal durations used in the studies

between 0 and 4 dB (depending on the frequency) from the average SPL used to calculate the thresholds.

E. Experimental procedure

The seals were trained to respond ("go") in the presence of a signal and to withhold the response ("no-go") in the absence of a signal. A trial began when the nonstudy animal was near the platform with one trainer and the study animal positioned with its head at the start/response buoy at the edge of the pool next to the research trainer [Fig. 1(a)]. When the trainer gave the animal a vocal command accompanied by a gesture (pointing downwards), the animal descended to the listening station (an L-shaped, 32-mm-diameter, water-filled polyvinylchloride tube with an end cap), so that its external auditory meatus was 200 cm from the sound source and 100 cm below the water surface [i.e., midwater; Fig. 1(b)]. Each animal was trained to position its nose against the listening station so that its head axis was in line with the projected beam axis of the transducer. The listening station was not connected to the sound box, and the transducer was suspended within the box by four thin ropes, so the animals were not able to use vibration via contact conduction to the nose to detect the signals. The animals' positions could be viewed from above by means of an underwater camera (Mariscope, Micro), which was attached to the listening station. The images were visible to the trainer near the start/response buoy (but out of the study animal's view when it was at the listening station) and to the operator in the research cabin.

Two trial types were conducted during each experimental session: signal-present trials and signal-absent trials. In signal-present trials, the stimulus was presented unpredictably between 4 and 10 s after the animal was positioned correctly at the listening station. A minimum waiting time of 4 s was chosen because it took about 4 s for the waves, created by the animal's descent, to dissipate. If the animal detected the sound, it responded by leaving the listening station (go response) at any time during the signal's duration and returning to the start/response buoy [Fig. 1(a)]. The signal operator then indicated to the trainer that the response was correct (a hit), after which the trainer gave a vocal signal and the seal received a fish reward. If the animal did not respond to the signal, the signal operator signaled to the trainer that the animal had failed to detect the signal (a miss). The trainer then indicated to the animal (by tapping softly on the side of the pool) that the trial had ended, thus calling the animal back to the start/response buoy. No reward was given following a miss. If the animal moved away from the listening station to the start/response buoy before a signal was produced (a prestimulus response), the signal operator indicated to the trainer to end the trial without rewarding the animal. After a prestimulus response, the animal was ignored for 8–10 s by the trainer.

In signal-absent, or catch, trials the signal operator hand signaled to the trainer to end the trial after a random interval of 4–10 s from when the seal had stationed (determined by a random number generator). The trial was terminated when the trainer blew very softly on a whistle. The tapping on the pool wall and whistle blowing were done softly to reduce the seal's exposure level difference between the test signals and the acoustic signals from the trainer. We believe this helped the animal to focus on very faint sounds throughout the sessions. If the animal responded correctly by remaining at the listening station until the whistle was blown (a correct rejection), it then returned to the start/response buoy and received a fish reward. If the seal left the listening station before the

whistle was blown (a prestimulus response), the signal operator indicated to the trainer to end the trial without rewarding the animal. The same amount of fish was given as a reward for correct go and no-go responses. In both signal-present and signal-absent trials, the trainer was unaware of the trial type when she sent the animal to the listening station. After sending the animal to the listening station, the trainer stepped out of the seal's view.

A session generally consisted of 30 trials per animal and lasted for about 15 min per animal. The seals were always tested in the same order, one immediately after the other. Sessions consisted of 70% signal-present and 30% signal-absent trials presented in random order, and only one signal frequency was presented each day. For each session, one of four data collection sheets was used. Each sheet comprised a different random series of trial types. Each seal had its own set of four data collection sheets. In each session, the signal amplitude was varied according to the simple up-down staircase procedure, a conventional psychometric technique (Robinson and Watson, 1973). This is a variant of the method of limits, which results in a 50% correct detection threshold (Levitt, 1971). During preliminary sessions, a rough threshold per test frequency was determined. During subsequent experimental sessions, the starting SPL of the signal was 10–15 dB above the estimated threshold. Following each hit, the signal amplitude on the next signal-present trial was reduced by 5 dB. Following each miss, the signal level was increased on the next signal-present trial by 5 dB. Prestimulus responses did not lead to a change in signal amplitude for the next trial. A switch in the seal's response from a detected signal (a hit) to an undetected signal (a miss), or *vice versa*, is called a reversal.

Thresholds were determined for 16 tonal signals (three pure tones and 13 narrowband FM signals). To prevent the animals' learning process from affecting the threshold levels, the test frequency was varied each day and adjacent frequencies were usually tested on successive days (going from low to high and from high to low frequencies, and so forth). This way the difference in frequency between days was limited, reducing the potential need for the study animals to adapt to a frequency. During the study we learned that the thresholds obtained for higher frequencies (>40 kHz) were not influenced by the wind force. Therefore, those frequencies were tested under relatively high wind force conditions, whereas the 0.125–0.5 kHz signals were only tested under wind force conditions below 2 Beaufort, because they required a quieter environment. Usually four experimental sessions were conducted on 5 days per week (at 0900, 1100, 1400, and 1600 h). Data were collected between August and November 2007.

Before each session, the acoustic equipment producing the stimuli was checked to ensure that it was functional and the stimuli were produced accurately (Fig. 2). Also the background noise level was checked to make sure it was not too high for testing. This was done in the following ways:

(1) To test the sound generating and amplifying equipment, the voltage output toward the underwater transducer was measured with an oscilloscope (Dynatek 8120, 20 MHz; Channel I) and a voltmeter (Hewlett Packard 3478A). This was done with the stimulus to be used in that session, at the amplitude at which the stimuli were calibrated.
(2) To test the sound level produced by the underwater transducer, the voltage output of a hydrophone (Labforce 1 BV), which was always placed in a fixed position 20 cm from the transducer and connected to a preamplifier (100×), was checked with the same oscilloscope (Channel II) and voltmeter when the stimulus for that session was produced.
(3) Audio stimuli (at sufficient SPLs) were checked aurally by the signal operator via another hydrophone (Labforce 1 BV), a charge amplifier (Bruel & Kjær, 2635), and an amplified loudspeaker. The operator also used this setup to monitor the background noise aurally before and during each session.
(4) Ultrasonic stimuli (>16 kHz); at sufficient SPLs) were checked for via the hydrophone (Labforce 1 BV). The signals were made audible to the signal operator by means of a modified ultrasound detector (Stag Electronics, Batbox III).

F. Analysis

Sessions with more than 20% prestimulus responses (i.e., more than six of the usual 30 trials per session) were not included in the analysis. These sessions occurred only four times per animal during the entire study, and usually coincided with obvious transient background noises.

For each session, the mean session hearing threshold was calculated by taking the mean of all reversal pairs in that session. Because no warm-up trials were used, it sometimes took several reversals before a stable threshold was reached. In these cases, the first one to four reversal pairs were not included in the analysis. The data included in the final analysis were from sessions carried out after the mean session threshold had leveled off. This usually occurred within four sessions (depending on the frequency and the animal). The reported thresholds for each seal were based on the mean of all remaining reversal pair values per frequency; approximately 110 reversal pairs per frequency obtained in about 11 sessions. The hearing thresholds of the two seals at each frequency were compared using a paired t-test.

III. RESULTS

The seals' sensitivity for each test frequency was stable over the 4 month study period. The mean prestimulus response rate (for both signal-present and signal-absent trials) varied between 3% and 13%, depending on the frequency (Table I). Most prestimulus responses occurred during tests with low-frequency signals.

The thresholds of the two seals were similar ($t=0.73$, $d.f.=15$, $P=0.48$). The underwater audiograms (50% detection thresholds) for the two seals showed the typical mammalian U-shape. However, the bottom part of the U was very flat and wide, the low frequency sensitivity decreased gradually, and the high frequency cutoff was steep [Fig. 3(b) and Table I]. The range of best hearing (10 dB from the maxi-

TABLE I. The mean 50% detection thresholds, standard deviation (SD), and total number of reversal pairs of 1-year-old female harbor seals 01 and 02 for three pure tones (0.125–0.25 kHz) and 13 narrowband FM (0.5–100 kHz) signals, and their prestimulus response rate based on the number of prestimulus responses in all trials (signal-present and signal-absent trials). Also shown are the CRs used to calculate the theoretical detection threshold in Fig. 3(b).

			Harbor seal 01			Harbor seal 02		
Center frequency (kHz)	Frequency modulation range 1% of center frequency (kHz)	Critical ratio (dB)	Total No. of reversal pairs	Mean 50% detection threshold (SPL in dB *re* 1 μPa, rms) $\pm$SD	Prestimulus response rate (%)	Total No. of reversal pairs	Mean 50% detection threshold (SPL in dB *re* 1 μPa, rms) $\pm$SD	Prestimulus response rate (%)
0.125	Pure tone	14.6	126	77 $\pm$ 4	9	124	74 $\pm$ 4	9
0.200	Pure tone	14.6	112	72 $\pm$ 4	6	90	73 $\pm$ 4	10
0.25	Pure tone	14.7	116	65 $\pm$ 4	11	96	69 $\pm$ 5	7
0.5	0.495–0.505	14.8	93	61 $\pm$ 4	11	90	64 $\pm$ 5	12
1	0.99–1.01	15.2	128	54 $\pm$ 4	6	103	56 $\pm$ 4	7
2	1.98–2.02	16.3	100	57 $\pm$ 4	5	98	57 $\pm$ 5	9
4	3.96–4.04	19.0	126	56 $\pm$ 5	5	113	55 $\pm$ 4	7
8	7.92–8.08	22.0	120	61 $\pm$ 4	10	105	58 $\pm$ 4	3
16	15.84–16.16	25.0	118	61 $\pm$ 4	9	109	63 $\pm$ 4	7
25	24.75–25.25	27.0	109	57 $\pm$ 4	8	109	58 $\pm$ 4	3
31.5	31.19–31.82	28.0	108	64 $\pm$ 4	8	124	63 $\pm$ 4	5
40	39.60–40.40	⋯	102	61 $\pm$ 4	13	103	60 $\pm$ 4	8
50	49.50–50.50	⋯	120	73 $\pm$ 3	5	117	70 $\pm$ 4	6
63	62.37–63.63	⋯	112	109 $\pm$ 3	8	100	106 $\pm$ 5	5
80	79.20–80.80	⋯	106	119 $\pm$ 4	5	106	119 $\pm$ 4	3
100	99.00–101.00	⋯	114	127 $\pm$ 4	5	120	125 $\pm$ 4	4

mum sensitivity at 1 kHz which was 54 dB *re* 1 μPa, rms) was very wide: from 0.5 to 40 kHz ($6\frac{1}{3}$ octaves), and sensitivity fell below 1 kHz and above 40 kHz.

IV. DISCUSSION AND CONCLUSIONS

A. Evaluation of the data

The biggest challenge in hearing studies is to maintain a low background noise level; we took great care to do this in the present study. The main factor influencing the low-frequency part of the background noise spectrum in the pool was the wind (which, when increased, caused airborne wind noise and increased soil vibrations). During the 4 month study period, the wind speed was low compared to other years, which resulted in very low background noise levels in the pool [even partly below sea state 0, see Fig. 3(a)].

It is important to know whether the audiograms of the present study are absolute audiograms or if the signals were influenced by the background noise in the pool. Theoretical masked detection thresholds (MDTs) [Fig. 3(b)] were calculated based on the mean background noise levels [Fig. 3(a)] and the harbor seal's critical ratio (CR) (from a smooth line through the data points of Turnbull and Terhune, 1990 and Southall *et al.*, 2000; Table I). The noise-limited theoretical MDT is calculated as MDT=background noise (spectrum level)+CR. The theoretical MDTs lie below the hearing thresholds found in the present study, suggesting that the thresholds were not masked by the ambient noise. Also, the background noise measurements averaged the sound pressure over time, whereas the level certainly fluctuated temporally. Most masking studies have been conducted with random Gaussian noise, but in the real world, the acoustic environment consists of noise where the energy across the frequency regions is coherently modulated in time. Bottlenose dolphins (*Tursiops truncatus*) have lower masked thresholds in temporally fluctuating comodulated noise than in Gaussian noise with the same spectral density level (Branstetter and Finneran, 2008). Thus the thresholds found in the present study were probably not being masked by the background noise in the pool.

The 0.125–0.5 kHz signals were only tested under wind force conditions below 2 Beaufort, because more prestimulus responses occurred at higher wind speeds. This was probably because the animals reacted to elements of background noise which resembled the test signals. Still, the prestimulus response rate was generally highest for frequencies below 1 kHz. Most transient background noise signals, which may well trigger prestimulus responses, are in this part of the spectrum. Because both seals were tested within the same sessions, any differences between the thresholds obtained for the two animals must have been due to differences in their hearing sensitivity and/or individual differences in their response criteria, motivational state, or behavior. Differences could not have been caused by differences in equipment, equipment settings, methodology, personnel, or background noise. Any changes (increase or decrease) in wind force influencing the background noise level during the 30 min in which the two seals were tested will probably have balanced each other over the very high number of sessions on which the thresholds were based.

Did the use of FM signals instead of pure tones influence the hearing thresholds found in the present study? In most previous studies of pinniped hearing, except in two experiments, pure tones were used as stimuli. Only the hearing of a Pacific walrus (*Odobenus rosmarus divergens*) and two Steller sea lions (*Eumetopias jubatus*) have been tested with narrowband FM tonal signals (Kastelein *et al.*, 2002, 2005) to create a more stable received SPL. In humans, FM signals tend to have a slightly higher arousal effect than pure tones, and therefore slightly lower hearing thresholds (<5 dB depending on center frequency and modulation frequency; Morgan *et al.*, 1979). However, the use of FM signals instead of pure tones probably had little effect on the thresholds found in the present study. This assumption is based on a hearing test with 0.25 kHz signals on a Pacific walrus (Kastelein *et al.*, 2002). In that study no difference was found between thresholds derived with narrowband FM signals (exactly like those used in the present study; frequency modulation only 1% of the center frequency) and those derived with pure tone signals.

B. Comparison with previous hearing studies in harbor seals

Comparing the hearing of the study seals with that of the other four harbor seals of which the underwater hearing sensitivities have been studied is not straightforward. In the various studies there are differences in the calibration methodology and threshold calculation, and variation in the threshold data between sessions. Also the background noise measurements are often limited, and lower frequency thresholds are not always free from masking influences. Researchers have used various methods and stimulus parameters such as signal type (pure tone versus FM signal) and signal duration (50–500 ms versus 900 ms). Also the SPL calculation method is not specified for all the studies (peak-to-peak or rms, causing a 9 dB difference). Despite these complications, general comparisons can be made between the underwater audiograms of the harbor seals in the present study and those in previous studies (Møhl, 1968; Terhune, 1988, 1989; Turnbull and Terhune, 1993, 1994; Terhune and Turnbull, 1995; Kastak and Schusterman, 1998; Southall *et al.*, 2005). Above 4 kHz, the thresholds found in the previous hearing studies and those found in the present study are similar. However, below 4 kHz the thresholds found in the present study were up to 20 dB lower than those found in the previous studies. Differences between the hearing sensitivity of the animals in the present studies and those in previous studies below 4 kHz may occur because of the following.

(1) Low-frequency signals were masked by background noise in previous studies.
(2) Animals in previous studies may have had temporary hearing loss due to the high background noise levels from pumps before the hearing tests were conducted.
(3) The signal duration in most previous studies was shorter than the one used in the present study, possibly causing an increase in the hearing threshold (not necessarily because of the integration time, but probably because it is difficult for the animals to distinguish between transient signals in the background noise and the test signals).
(4) There may have been individual, gender, health condition, or age-related differences in hearing sensitivity between the test animals.

Based on the small minimum audible angles for low frequencies, Bodson *et al.* (2007) concluded that harbor seals are low-frequency hearing specialists. The present study shows that harbor seals have a very wide frequency range of best hearing, and in quiet conditions are able to hear lower frequencies better than previously thought.

C. Ecological significance

The most important finding of this study is that harbor seal hearing is more sensitive below 4 kHz than found in previous studies [Fig. 3(c)]. The hearing range of harbor seals overlaps in frequency with the loudest and most common anthropogenic noise sources. The effect of anthropogenic noise on marine mammals is highly variable in type and magnitude (Severinsen, 1990; Cosens and Dueck, 1993; Richardson *et al.*, 1995), and harbor seals show avoidance behavior to certain sounds in certain contexts (Kastelein *et al.*, 2006a, 2006b). Anthropogenic noise might reduce the time harbor seals forage in particular areas, thus reducing their physiological condition and their reproductive success. In addition to the hearing sensitivity of the harbor seal, the radii of avoidance and disturbance zones around sound sources depend on several other factors such as the general background noise level, water depth, ocean floor sediment properties, and the spectrum, source level, and duration of the anthropogenic noise. In general, based on the findings of the present study, under unmasked conditions, many anthropogenic noise sources are audible at greater ranges than formerly believed. When ambient noise levels in nature are higher than those of our testing facility, the auditory thresholds will be masked, however.

The dominant energy of harbor seal underwater sound production is below 2 kHz (Van Parijs and Kovacs, 2002). The low detection thresholds in this frequency range found in the present study means that under unmasked conditions, harbor seals can communicate with each other underwater over greater ranges than formerly believed.

ACKNOWLEDGMENTS

We thank students Aniek van den Berg, Krista Krijger and Tess van der Drift, and Alejandra Vargas, and volunteers Menno van den Berg, Jesse Dijkhuizen, Petra van der Marel, and Sofie Vandermaele for their help with training the seals and collecting the data, and Rob Triesscheijn for making the figures. We thank Dick de Haan (Wageningen IMARES, the Netherlands) for his technical assistance and Veenhuis Medical Audio (Marco Veenhuis and Herman Walstra) for donating the audiometer. We thank Bert Meijering (director of Topsy Baits, Wilhelminadorp, the Netherlands) for providing space for SEAMARCO's Institute, and Hein Hermans for providing technical support to run the facility. We also thank Charles Greene (Greenridge Sciences, USA), Nancy Jen-

nings (Dotmoth.co.uk, Bristol, UK), and two anonymous reviewers for their valuable constructive comments on this manuscript. This study was conducted by SEAMARCO subcontracted to IMARES (contacts Han Lindeboom and Reinier Hille Ris Lambers). The study was funded by We@Sea, Noordzee Wind EIA (wind generator parks at sea), and RIKZ Middelburg, The Netherlands (contacts Belinda Kater and Martine van den Heuvel-Greve; acoustic disturbance of harbor seals in the Westerscheldt). We thank director Just van den Broek and curator of animals Henk Brugge (both from Ecomare, Texel) for making the harbor seals available for this project. The seals' training and testing were conducted under authorization of the Netherlands Ministry of Agriculture, Nature and Food Quality, Department of Nature Management, with Endangered Species Permit No. FF/75A/2005/048.

Bodson, A., Miersch, L., and Dehnhardt, G. (**2007**). "Underwater localization of pure tones by harbor seals (*Phoca vitulina*)," J. Acoust. Soc. Am. **122**, 2263–2269.

Branstetter, B. K., and Finneran, J. J. (**2008**). "Comodulation masking release in bottlenose dolphins (*Tursiops truncatus*)," J. Acoust. Soc. Am. **124**, 625–633.

Burns, J. J. (**2002**). "Harbor seal and spotted seal," in *Encyclopedia of Marine Mammals*, edited by W. F. Perrin, B. Würsig, and J. G. M. Thewissen (Academic, San Diego), pp. 552–560.

Cosens, S. E., and Dueck, L. P. (**1993**). "Icebreaker noise in Lancester sound, N.W.T., Canada: Implications for marine mammal behavior," Marine Mammal Sci. **9**, 285–300.

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.

Hassall, J. R., and Zaveri, K. (**1988**). "Acoustic noise measurements," Brüel & Kjær documentation.

Kastak, D., and Schusterman, R. J. (**1998**). "Low-frequency amphibious hearing in pinnipeds: Methods, measurements, noise, and ecology," J. Acoust. Soc. Am. **103**, 2216–2228.

Kastelein, R. A., Mosterd, P., van Santen, B., Hagedoorn, M., and de Haan, D. (**2002**). "Underwater audiogram of a Pacific walrus (*Odobenus rosmarus divergens*) measured with narrow-band frequency-modulated signals," J. Acoust. Soc. Am. **112**, 2173–2182.

Kastelein, R. A., van Schie, R., Verboom, W. C., and de Haan, D. (**2005**). "Underwater hearing sensitivity of a male and a female Steller sea lion (*Eumetopias jubatus*)," J. Acoust. Soc. Am. **118**, 1820–1829.

Kastelein, R. A., van der Heul, S., Verboom, W. C., Triesscheijn, R. J. V., and Vaughan Jennings, N., (**2006a**). "The influence of underwater data transmission sounds on the displacement of captive harbour seals (*Phoca vitulina*)," Mar. Environ. Res. **61**, 19–39.

Kastelein, R. A., van der Heul, S., Terhune, J. M., Verboom, W. C., and Triesscheijn, R. J. V., (**2006b**). "Deterring effects of 8–45 kHz tone pulses on harbor seals (*Phoca vitulina*) in a large pool," Mar. Environ. Res. **62**, 356–373.

Knudsen, V. O., Alford, R. S., and Emling, J. W., (**1948**). "Underwater ambient noise," J. Mar. Res. **7**, 410–429.

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

Møhl, B. (**1968**). "Auditory sensitivity of the common seal in air and water," J. Aud Res. **8**, 27–38.

Morgan, D. E., Dirks, D. D., and Bower, D. R. (**1979**). "Suggested threshold sound pressure levels for frequency-modulated (warble) tones in the sound field," J. Speech Hear Disord. **44**, 37–54.

Richardson, W. J., Greene, C. R., Malme, C. I., and Thomson, D. H. (**1995**). *Marine Mammals and Noise* (Academic, San Diego).

Robinson, D. E., and Watson, C. S. (**1973**). "Psychophysical methods in modern Psychoacoustics," in *Foundations of Modern Auditory Theory*, edited by J. V. Tobias (Academic, New York), Vol. **2**, pp. 99–131.

Schusterman, R. J., Balliet, R. F., and St. John, S. (**1970**). "Vocal displays under water by the gray seal, the harbor seal, and Stellar sea lion," Psychonomic Sci. **18**, 303–305.

Severinsen, T. (**1990**). "Effects of disturbance on marine mammals," in: *Environmental Atlas Gipsdalen, Svalbard*, edited by T. Severinsen, and R. Hansson (Norwegian Polar Research Institute, Tromsø), Report No. 66, Vol. **3**, pp. 41–63.

Southall, B. L., Schusterman, R. J., and Kastak, D. (**2000**). "Masking in three pinnipeds: Underwater, low-frequency critical ratios," J. Acoust. Soc. Am. **108**, 1322–1326.

Southall, B. L., Schusterman, R. J., Kastak, D., and Reichmuth Kastak, C. (**2005**). "Reliability of underwater hearing thresholds in pinnipeds," ARLO **6**, 243–249.

Terhune, J. M. (**1988**). "Detection thresholds of a harbor seal to repeated underwater high-frequency, short duration sinusoidal pulses," Can. J. Zool. **66**, 1578–1582.

Terhune, J. M. (**1989**). "Underwater click hearing thresholds of a harbor seal," Aquat. Mamm. **15**, 22–26.

Terhune, J., and Turnbull, S. (**1995**). "Variation in the psychometric functions and hearing thresholds of a harbour seal," in *Sensory Systems of Aquatic Mammals*, edited by R. A. Kastelein, J. A. Thomas, and P. E. Nachtigall (De Spil, Woerden), pp. 81–93.

Turnbull, S. D., and Terhune, J. M. (**1990**). "White noise and pure tone masking of pure tone thresholds of a harbour seal listening in air and underwater," Can. J. Zool. **68**, 2090–2097.

Turnbull, S. D., and Terhune, J. M. (**1993**). "Repetition enhances hearing detection thresholds in a harbour seal (*Phoca vitulina*)," Can. J. Zool. **71**, 926–932.

Turnbull, S. D., and Terhune, J. M. (**1994**). "Descending frequency swept tones have lower thresholds than ascending frequency swept tones for a harbor seal and human listeners," J. Acoust. Soc. Am. **96**, 2631–2636.

Van Parijs, S. M., and Kovacs, K. M. (**2002**). "In-air and underwater vocalizations of the eastern Canadian harbour seals, *Phoca vitulina*," Can. J. Zool. **80**, 1173–1179.

Variability in ambient noise levels and call parameters of North Atlantic right whales in three habitat areas

Susan E. Parks
Applied Research Laboratory, The Pennsylvania State University, P.O. Box 30, State College, Pennsylvania 16804

Ildar Urazghildiiev and Christopher W. Clark
Bioacoustics Research Program, Laboratory of Ornithology, Cornell University, 159 Sapsucker Woods Road, Ithaca, New York 14850

(Received 20 June 2008; revised 31 October 2008; accepted 20 November 2008)

The North Atlantic right whale inhabits the coastal waters off the east coasts of the United States and Canada, areas characterized by high levels of shipping and fishing activities. Acoustic communication plays an important role in the social behavior of these whales and increases in low-frequency noise may be leading to changes in their calling behavior. This study characterizes the ambient noise levels, including both natural and anthropogenic sources, and right whale upcall parameters in three right whale habitat areas. Continuous recordings were made seasonally using autonomous bottom-mounted recorders in the Bay of Fundy, Canada (2004, 2005), Cape Cod Bay, (2005, 2006), and off the coast of Georgia (2004–2005, 2006–2007). Consistent interannual trends in noise parameters were found for each habitat area, with both the band level and spectrum level measurements higher in the Bay of Fundy than in the other areas. Measured call parameters varied between habitats and between years within the same habitat area, indicating that habitat area and noise levels alone are not sufficient to predict variability in call parameters. These results suggest that right whales may be responding to the peak frequency of noise, rather than the absolute noise level in their environment. © *2009 Acoustical Society of America.* [DOI: 10.1121/1.3050282]

PACS number(s): 43.80.Nd, 43.80.Ka, 43.50.Rq [WWA] Pages: 1230–1239

I. INTRODUCTION

The North Atlantic right whale (*Eubalaena glacialis*) is a highly endangered species with a remaining population estimated at fewer than 400 individuals in the North Atlantic Ocean (Kraus *et al.*, 2005). Being the "right" whale to hunt, right whales were one of the first cetacean species hunted to extremely low numbers. North Atlantic right whales are found in the coastal waters of the United States and Canada, ranging from the coasts of Florida and Georgia in the winter months to the Bay of Fundy (BOF) in Canada in the summer (Kraus and Kenney, 1991; Brown *et al.*, 1995; Kraus and Rolland, 2007). Their habitat areas overlap with areas of high human use from shipping traffic and fishing activities off the east coast of North America (Kraus and Rolland, 2007). North Atlantic right whales were declining at approximately 2% year in the 1990s (Kraus *et al.*, 2005), while the southern right whale, a closely related species, has shown population growth levels approaching 7%–9% (Best *et al.*, 2001). Potential causes of this disparity between the two regions include differences in reproductive output, anthropogenic mortality from vessel collision and fishing gear entanglement, and prey availability (Kraus *et al.*, 2005). Another potential concern is that increased levels of anthropogenic noise in the whales' urban environment may impact their ability to communicate or increase their stress levels (Parks and Clark, 2007). Ocean noise has been a topic of interest for over 50 years, with early studies describing the main contributions to ambient noise at different frequencies (Knudsen *et al.*, 1948; Wenz, 1962) and raising the possibility that higher ambient noise levels might impact whale communication (Payne and Webb, 1971). More recently, extensive research efforts have been undertaken to determine the effects of human noise sources on marine life (see reviews in Richardson *et al.*, 1995 and Nowacek *et al.*, 2007). Recent studies indicate that low-frequency ocean ambient noise levels are increasing. Two studies comparing contemporary sound levels to recordings from the 1960s indicate that there has been approximately an 8–12 dB increase in sound levels below 100 Hz at two Californian sites over 30–40 years (Andrew *et al.*, 2002; McDonald *et al.*, 2006). Distant shipping traffic has been suggested as the most likely source for these observed increases in low-frequency ambient noise. A similar increase in low-frequency ambient noise between 1958 and 1975 was described for the Western North Atlantic (Ross, 1993). A study of noise levels in the North Atlantic on the Canadian continental shelf noted a peak in low-frequency sound levels around 80 Hz (Zakarauskas *et al.*, 1990).

North Atlantic right whales are found in areas with relatively high levels of shipping activity, suggesting that similar increases of ambient noise levels should be occurring in their environment. The frequency range of North Atlantic right whale "upcalls," a stereotyped signal used by right whales for maintaining contact between individuals, is generally between 50 and 350 Hz (Clark *et al.*, 2007). Although the close passage of a ship can increase noise at frequencies above 1 kHz (Aguilar Soto *et al.*, 2006), most ship noise is below 1 kHz and energy from distant shipping is primarily below 100 Hz (Wenz, 1962). A study of ambient noise levels on the

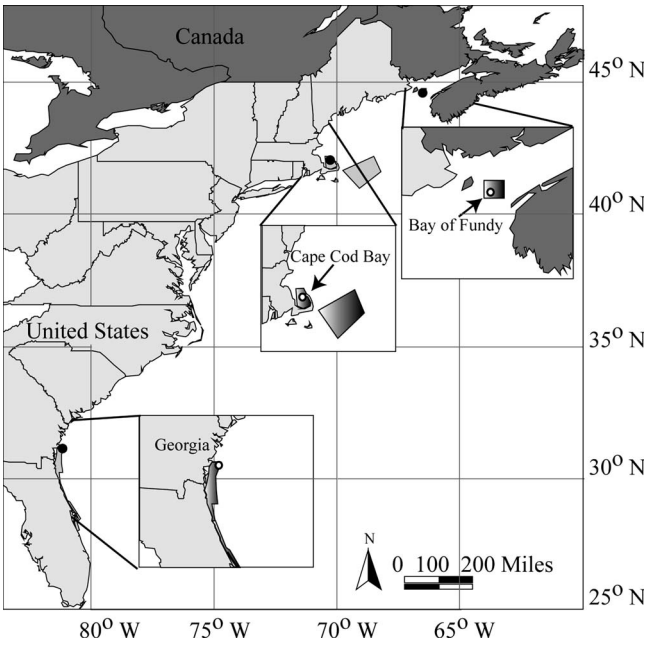


FIG. 1. Map of recording locations. Right whale habitat areas are shown in shaded gray. Enlargements of the right whale habitat areas show the position of the recordings with a white circle.

Stellwagen Bank National Marine Sanctuary, a known habitat area of North Atlantic right whales in the Gulf of Maine, measured the contribution of vessels to increases in low-frequency noise (Hatch *et al.*, 2008). Therefore, there is concern that increased low-frequency noise from shipping may cause masking of important communication signals for this species.

A recent study of right whales indicates that there may be long-term changes in calling behavior in response to increasing ambient noise (Parks *et al.*, 2007b). By collecting upcalls recorded from right whale populations of the North Atlantic and South Atlantic, comparisons were made between call parameters in low and high noise environments. The low noise recordings were from 1956 in the North Atlantic and 1977 in Argentina, while the higher noise recordings were collected recently from both regions. The prediction was that Southern right whale recordings would be from a lower noise environment given the lower level of commercial shipping in the South Atlantic. The second prediction was that the noise levels in both regions would have increased through time with the global increase in shipping activity. The results of the comparisons indicated that in both regions calls were consistently shifted into higher frequency bands under the higher noise conditions (Parks *et al.*, 2007b).

The goals of this study are to quantify the ambient noise levels, including both anthropogenic and natural sound sources, in three habitat areas of the North Atlantic right whale to assess the levels of noise that right whales are exposed to and what percentage of the time their habitats are characterized by high levels of noise. Noise levels were measured and compared from three right whale habitat areas: the BOF, Canada; Cape Cod Bay, Massachusetts; and off Brunswick Harbor, Georgia (Fig. 1). The frequency content and duration of the right whale upcalls in each habitat were measured to determine if there were detectable differences that correlate with variations in ambient noise levels between habitat areas.

II. METHODS

A. Acoustic recordings

All recordings were made using archival bottom-mounted acoustic recorders designed by the Cornell University Bioacoustics Research Program (Clark *et al.*, 2002). These units are referred to as "Pop-ups" because they are released from the bottom by an acoustically triggered release mechanisms that allows them to float to the surface for recovery. Each unit consisted of an HTI-94-SSQ hydrophone with a sensitivity of −168 dB re 1 V/μPa, an amplifier with a gain of 23.5 dB, and A/D converter with a sensitivity of 10^3 bit/V. An additional normalization of the A/D converter output by a factor of 1/2048 was implemented when storing the digitized data on a hard drive. The final transformation coefficient used for calibration of the digitized data was $\tilde{C} = -168.5 + 23.5 + 20\log(10^3) - 20\log(2048) = -151.2$ dB re 1 μPa. The systems had a flat ($\pm$1.0 dB) frequency response between 10 and 585 Hz, which included the bandwidth of right whale upcalls (50–350 Hz) used for the measurements reported here.

Pop-up units were deployed in each of the three right whale habitat areas for several weeks to months in each of 2 years. The three habitat areas included the BOF, Canada, a summer feeding and nursery ground; Cape Cod Bay, Massachusetts, a late winter and early spring feeding ground; and the coastal waters of Georgia at the northern end of the calving grounds (Kraus and Rolland, 2007). The BOF recordings were made in 2004 and 2005, the Cape Cod Bay recordings were made in 2005 and 2006, and the Georgia recordings were made in 2004 and 2006. The locations, dates, and duration for each data set are summarized in Table I. The locations of the recording units relative to the relevant designated right whale critical habitat and conservation areas are shown in Fig. 1. The distance from major shipping lanes varied for each of the recording units. The recording units in the BOF were deployed ~5 nm from a major shipping lane into Saint John, New Brunswick (Knowlton and Brown, 2007). The Cape Cod Bay recording units were deployed ~20 nm from the major shipping lane into Boston, MA. The Georgia recording units were deployed ~12 nm from a moderate shipping lane into Brunswick, GA (Ward-Geiger *et al.*, 2005).

B. Data analysis

1. Ambient noise characteristics

Many different types of noise sources, for example, ships, industrial activities, wind, and rain, and seismic exploration operations contribute to the low-frequency ambient noise in the marine environment (Wenz, 1962; Urick, 1983; Etter, 1991). Spatial distribution of noise radiating vessels, wind speed, and other major factors affecting the ambient noise characteristics change relatively slowly over time. As a result, over short-time intervals of several minutes, ambient noise can be represented as a Gaussian stationary process

TABLE I. Summary of data used in this analysis, including dates, sampling rates, depth, and positions of recording units. BOF=Bay of Fundy, CCB=Cape Cod Bay, and GA=Georgia coast.

Data set	Dates of recording	Sampling rate (kHz)	Approximate depth (m)	Coordinates (D-D)	
				Latitude	Longitude
BOF 2004	August 7, 2004–August 25, 2004	8	200	44.5915°	−66.5003°
BOF 2005	July 29, 2005–August 18, 2005	8	200	44.5915°	−66.5003°
CCB 2005	January 27, 2005–April 25, 2005	2	30	41.99015°	−70.3078°
CCB 2006	February 24, 2006–May 11, 2006	2	30	42.01611°	−70.31095°
GA 2004	November 29, 2004–February 18, 2005	2	15	31.78942°	−80.82567°
GA 2006	December 19 2006–February 25 2007	2	15	31.12051°	−81.13308°

(Urick, 1977; Urazghildiiev and Clark, 2006). However, the long-term variability of the ambient noise characteristics can be important. Therefore, a sliding short-time window, $\Omega(t)=[t,t+1,\ldots,t+N_\Omega-1]$, 300 s in length and consisting of N_Ω samples was used to calculate the ambient noise characteristics. The adjacent time windows were overlapped by 150 s. Measures included the short-time spectrum level and the band level (BL) in the 50–350 Hz frequency band.

In the presence of transients in the acoustic environment, computations of the short-time spectrum level and the BL estimates are not trivial. Since different methods of estimation may produce different results, we represent the short-time spectrum level and BL estimates in their closed forms. In the presence of transients, the standard Bartlett or Welch methods cannot provide an acceptable spectrum level estimation accuracy (Maronna *et al.*, 2006), and methods robust to outliers should be implemented. Here we employ a median-based approach. Similar to the Bartlett method, this algorithm uses N_1 sliding time windows starting at times $t_i \in \Omega(t)$, $i=1,2,\ldots,N_1$, with $N \ll N_\Omega$ samples each and formed within the window $\Omega(t)$. Let us introduce a function

$$p(f,t_i)=\frac{C}{F_S N}\left|\sum_{n=0}^{N-1} x(t_i+n)\exp\{-j2\pi fn/F_S\}\right|^2, \tag{1}$$

where $C=10^{0.1\tilde{C}}$ is a calibration coefficient and F_S is the sampling frequency. We assume that the series of functions $p(f,t_i)$ are computed for the discrete frequency grid $f \in [0,F_S/N,\ldots,(N-1)F_S/N]$ using the fast Fourier transform algorithm. In this case, the function $p(f,t_i)$ is scaled in units of intensity re 1 μPa in a frequency band 1 Hz wide (Marple, 1987). According to the Bartlett method, the short-term spectrum level estimate is a mean value of the functions $p(f,t_i)$. The proposed robust estimate is calculated as a median of the L functions $p(f,t_i)$ on a frequency-by-frequency basis:

$$G(f,t)=\text{med}\{p(f,t_1),p(f,t_2),\ldots,p(f,t_L)\}. \tag{2}$$

In contrast to the Bartlett estimate, the median spectrum level estimate is less sensitive to outliers arising from transient impulsive noise sources. However, it differs from the Bartlett estimates even when there are no outliers present. Observations show that the bandpassed, zero-mean process $x(t)$ has a symmetric probability distribution. Equation (1) represents a quadratic transformation of the input process $x(t)$. Hence, the distribution of the random variables $p(f,t_i)$ is nonsymmetric and close to a Rayleigh distribution if the distribution of $x(t)$ is close to Gaussian. As such, for any f, the median value of $p(f,t_i)$, $i=1,2,\ldots$, is less than the mean value. Test results conducted with sections of data with no detected transients show that the difference between the proposed median and the Bartlett spectrum level estimates is between 0.5 and 2.0 dB. The long-term variations of the ambient noise intensity are much greater than 2 dB, so the magnitude of this estimation error is negligible compared to the estimation error provided by the Bartlett estimate in the presence of outliers.

The short-time spectrum level can be computed from (2) as $S(f,t)=10\log_{10}G(f,t)$ (Urick, 1983). The corresponding BL can be calculated as (Urick, 1983)

$$I(t)=10\log_{10}P(t), \tag{3}$$

where

$$P(t)=\frac{F_S}{N}\sum_{f=f_{50}}^{f_{350}} G(f,t), \tag{4}$$

where f_{50}, f_{350} are the frequency indices corresponding to the frequencies 50 and 350 Hz, respectively.

The data in this study indicate that for any f, the empirical distribution of the values $S(f,t)$ is different from Gaussian. For non-Gaussian distributions, the mean value may not represent the most probable value. Therefore, the mode spectrum level is presented here as the averaged characteristic of the ambient noise in the frequency domain. For each f, the mode spectrum level is calculated as an argument of the peak value of the histogram:

$$S_m(f) = \arg\max_S H(S,f), \tag{5}$$

where $H(S,f)$ is the two-dimensional (2D) histogram of short-time spectrum levels $S(f,t)$. The resultant mode spectrum level $S_m(f)$ was smoothed using the moving average filter with the span 9 implemented in MATLAB 7.4.0.

To characterize ambient noise conditions in which North Atlantic right whales actually vocalized, we also computed the median spectrum level estimate, Eq. (2), for a 1 min period before the time of arrival (TOA) of each detected call. The peak frequency of ambient noise was computed as

$$f_{\text{peak}}(t_j) = \arg\max_{f\in[f_{50},f_{350}]} G(f,t_j), \tag{6}$$

where t_j is the TOA of the jth detected call.

2. Upcall measurements

A single right whale call type, the upcall, was selected for measurement because it is a species specific stereotyped call that functions as a contact call in right whales and is known to be produced by both sexes of right whales in all habitat areas (Clark *et al.*, 2007). The upcalls were automatically detected and their TOAs were estimated using the method described in Urazghildiiev *et al.* (2008). This technique is based on a multiple-stage hypothesis testing process involving a spectrogram-based detector (Urazghildiiev and Clark, 2007), spectrogram testing, and feature testing algorithms. Only signals with high signal-to-noise ratio were used in our tests. Selection of high signal-to-noise ratio calls allowed for more accurate estimation of the measured call parameters, and the whales generating these calls were likely only a short distance from the recording unit. As such, these whales vocalized under approximately the same noise characteristics as those measured by the sensor, and we can use the noise conditions at the sensor to represent the noise at the calling whale.

Each automatically detected signal was checked by experienced human operators, and the signals not clearly visible on the spectrogram and/or that overlapped with transients were removed from the tested data set. The signal parameters measured from the spectrogram included minimum frequency, peak frequency, and duration. On a gray-scaled spectrogram, the distribution of signal energy in a time-frequency plane can be viewed as a dark area on a light noise background. Hence, for each detected signal, the human operator selected a rectangular area on a spectrogram as

$$B(f,t) = \begin{cases} 1, & f_{\min} \le f \le f_{\max}, \quad t_{\text{start}} \le t \le t_{\text{end}} \\ 0 & \text{otherwise} \end{cases} \tag{7}$$

where $f_{\min}, f_{\max}, t_{\text{start}}, t_{\text{end}}$ are the minimum and maximum frequencies, and start and end times, respectively, encompassing the annotated area so as to represent the "signal image" (i.e., the dark area) on the spectrogram. Within that area, the signal spectrogram is computed as

$$S_{\text{sig}}(f,t) = \begin{cases} S(f,t), & S(f,t) > 4S_{\text{med}} \\ 0 & \text{otherwise} \end{cases} \tag{8}$$

where S_{med} is the median value of the data spectrogram within the area $B(f,t)$. The instantaneous frequency of the signal is computed as

$$\hat{f}(t) = \arg\max_f S_{\text{sig}}(f,t). \tag{9}$$

The minimum frequency of the jth detected call is calculated as

$$F_{\min} = \min \hat{f}(t). \tag{10}$$

We define the peak frequency as the frequency corresponding to the peak value in the signal spectrogram. The peak frequency is calculated as

$$F_{\text{peak}} = \arg\max_{f,t} S_{\text{sig}}(f,t). \tag{11}$$

Signal duration is measured as

$$\tau_s = t_{\text{end}}^{\text{sig}} - t_{\text{start}}^{\text{sig}}, \tag{12}$$

where $t_{\text{start}}^{\text{sig}}, t_{\text{end}}^{\text{sig}}$ are the start and end times of the signal spectrogram $S_{\text{sig}}(f,t)$.

3. Statistics

SYSTAT 12 was used to compare the duration and minimum frequency values of measured calls among different habitat areas and across different years using a two-way analysis of variance (ANOVA). Levene's test for homogenous variance was used to test for the equality of variance among the data sets being tested. The differences between the peak frequency of the call and the peak frequency of the noise during the minute before the call were compared using a Wilcoxon Sign Ranked Test. We tested for correlations between noise levels one minute before a call and the duration and peak frequency of that call.

III. RESULTS

A. Background noise parameters

The typical values and variations of ambient noise BL in each of the three habitat areas on a day to day basis over a two week period are shown in Fig. 2. Comparisons of the empirical cumulative density function (ECDF) of the BL for the three habitats in the frequency band of right whale upcalls (50–350 Hz) are shown in Fig. 3. The BL of the ambient noise was highest in the BOF and the lowest in Georgia (Fig. 3). The ECDF illustrates the percentage of time each habitat experienced BL noise above a particular value. For example, the BL in the BOF in 2004 was below 105 dB re 1 μPa only 0.04 or 4% of the time, while the BL of Georgia in 2004 was below 105 dB re 1 μPa 0.8 or 80% of the time. These data in Figs. 2 and 3 are from the two recorders located closest to calling North Atlantic right whales for each recording session. The ambient noise BL in each of the three areas varied by over 30 dB (Fig. 3), and this range of variability occurred within 24 h periods in each area (Fig. 2). The 2D histograms of the spectrum level for each habitat

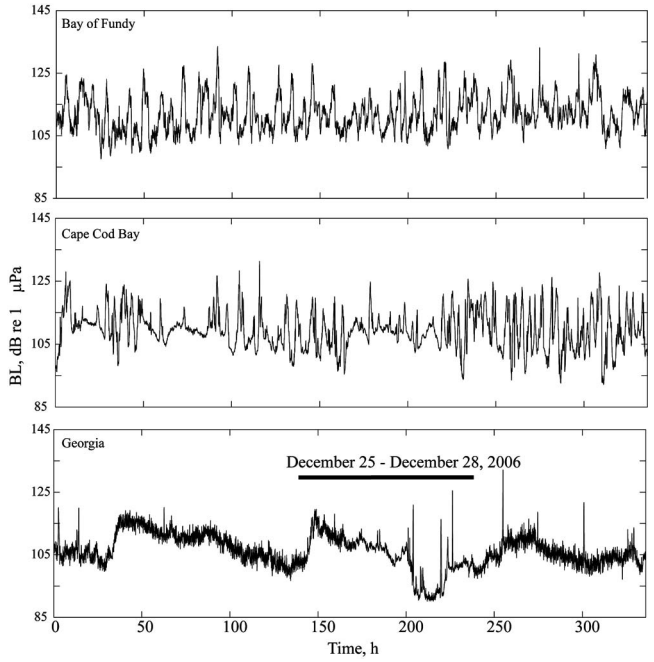


FIG. 2. Ambient BL noise in the 50–350 Hz band over a 2 week (336 h) period in each of the three habitat areas (a) BOF, August 7–20, 2004, (b) Cape Cod Bay, February 24, 2006–March, 9 2006, and (c) Georgia, December 19, 2006–January 1, 2007.

area are shown in Fig. 4. The smoothed mode spectrum level estimates of the ambient noise for both years in each habitat area are shown in Fig. 5. The spectrum level and BL of the ambient noise were highest in the BOF and the lowest in Georgia (Figs. 3–5).

B. Upcall signal parameters

A total of 8671 upcalls were detected and measured for this analysis. A summary of calls per recording session, the duration, minimum, and peak frequency of upcalls in each habitat are summarized in Table II and the distribution of these values are shown in Fig. 6.

For duration, the two-way ANOVA found a main effect

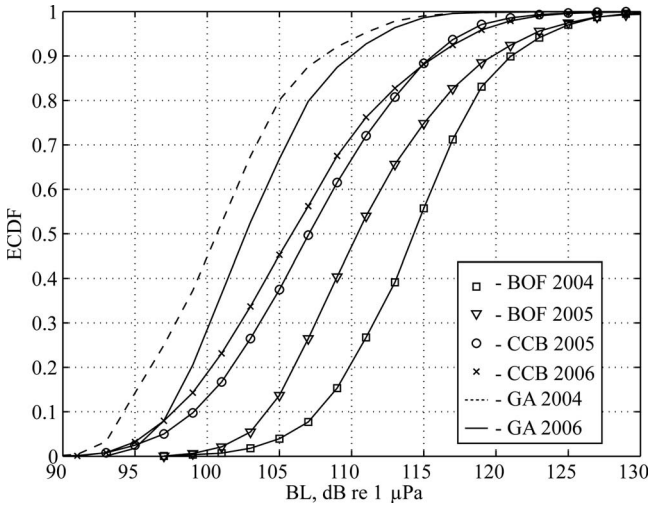


FIG. 3. ECDF of the BL of ambient noise (50–350 Hz) in each recording session. BOF=Bay of Fundy, CCB=Cape Cod Bay, and GA=Georgia.

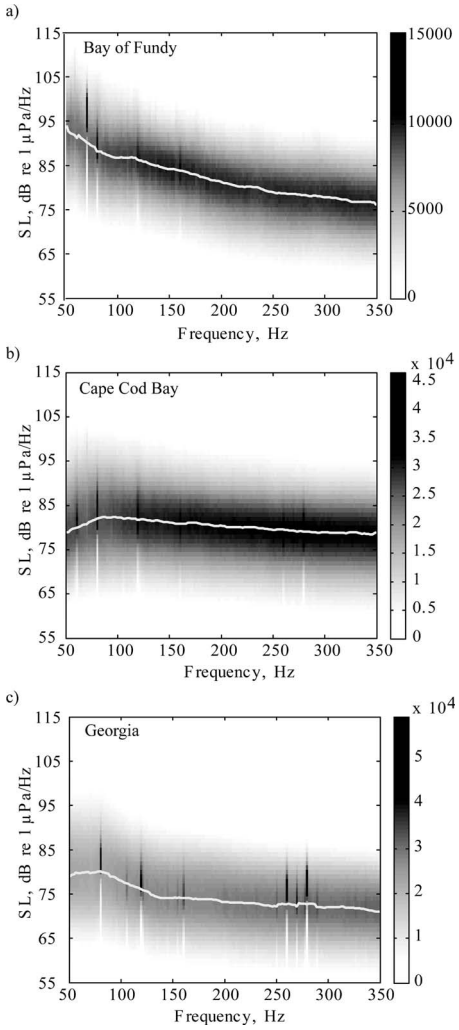


FIG. 4. The 2D histogram of the ambient noise spectrum level measured in (a) the BOF, (b) Cape Cod Bay, and (c) the Georgia coast. The white line in each panel represents the mode value.

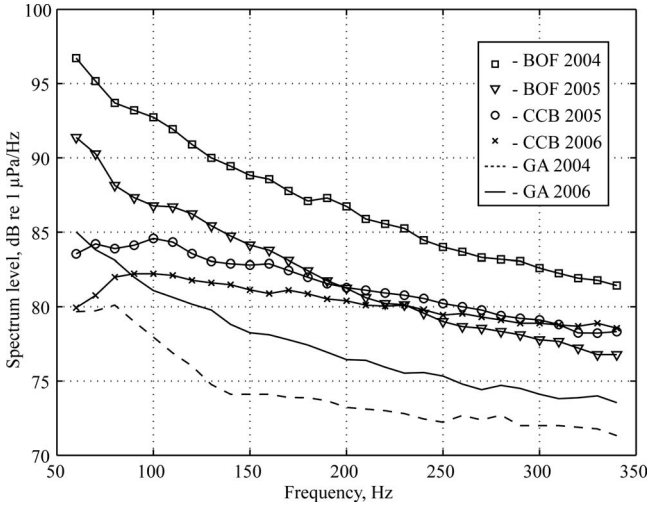


FIG. 5. Mode spectrum level for each recording session in the 50–350 Hz band. BOF=Bay of Fundy, Canada; CCB=Cape Code Bay, Massachusetts; and GA=Georgia. Note that although there are differences in level between years, the relative distribution of noise spectrum level is consistent by habitat.

TABLE II. Summary of upcall parameters measured from each data/set.

Data set	N	Duration (s) range mean ± SD	Minimum frequency (Hz) range mean ± SD	Peak frequency (Hz) range mean ± SD	Maximum frequency (Hz) range mean ± SD
BOF 2004	723	0.25–1.4 0.71 ± 0.20	49–161 101 ± 18	49–250 121 ± 28	106–348 177 ± 40
BOF 2005	935	0.19–2.9 0.75 ± 0.31	54–172 108 ± 17	63–234 121 ± 24	108–348 170 ± 33
CCB 2005	2604	0.29–2.5 0.85 ± 0.24	47–195 103 ± 18	66–266 124 ± 23	101–359 185 ± 36
CCB 2006	1865	0.25–2.0 0.83 ± 0.24	54–164 100 ± 18	63–214 121 ± 23	94–344 183 ± 38
GA 2004	1662	0.28–2.0 0.87 ± 0.23	47–180 94 ± 18	47–273 117 ± 21	98–355 199 ± 42
GA 2006	882	0.25–2.0 0.75 ± 0.23	63–172 107 ± 18	74–238 122 ± 22	113–402 196 ± 38

of region $[F(2,8664)=122.8,\ p\leqslant 0.0001]$. *Post hoc* analysis to evaluate pairwise differences using a Tukey HSD test indicated that the duration of calls in the BOF were shorter than in Georgia, and calls in both these regions were shorter than calls recorded in Cape Cod Bay. The year (1 versus 2) was found to be a main effect $[F(1,8664)=33.9,\ p\leqslant 0.001]$ as well as the interaction between region and year $[F(2,8664)=63.9,\ p\leqslant 0.0001]$. Call duration increased between 2004 and 2005 in the BOF, decreased in Georgia, and remained the same in Cape Cod Bay between 2004/2005 and 2006, respectively.

For minimum frequency, the two-way ANOVA found a main effect of region $[F(2,8664)=24.7,\ p\leqslant 0.0001]$. *Post hoc* analysis to evaluate pairwise difference using a Tukey HSD test indicated that the minimum upcall frequency was higher in the BOF in 2005 than in both years from Cape Cod Bay, though no difference was apparent from the 2004 BOF recordings. The minimum upcall frequency in Cape Cod Bay and the BOF was higher than in Georgia. The year of recording (1 versus 2) was found to be a main effect $[F(1,8664)=163.5,\ p\leqslant 0.0001]$, as well as the interaction between region and year $[F(2,8664)=145.8,\ p\leqslant 0.0001]$, indicating that the minimum frequency increased in the BOF and Georgia between the two recording years, while the minimum frequency decreased in Cape Cod Bay.

Measurements of upcall peak frequencies and noise peak frequency for the 1 min period before calls indicated that calls were higher in peak frequency than the noise peak frequency in almost all cases [BOF (2004, $n=739$); Cape Cod Bay (2006, $n=816$); Georgia (2006, $n=570$), BOF mean call $F_{\text{peak}}=118$ Hz, mean noise $F_{\text{peak}}=52$ Hz, mean difference, 66 Hz; Wilcoxon sign rank test, $Z=23.2$, $p<0.0001$, Cape Cod Bay mean call $F_{\text{peak}}=129$ Hz, mean noise $F_{\text{peak}}=107$ Hz, mean difference, 22 Hz; Wilcoxon sign rank test, $Z=11.5$, $p<0.0001$, Georgia mean call $F_{\text{peak}}=128$ Hz, mean noise $F_{\text{peak}}=40$ Hz, mean difference, 88 Hz; Wilcoxon sign rank test, $Z=33.1$, $p<0.000$. The overall mean call $F_{\text{peak}}=125.0$ Hz, mean noise $F_{\text{peak}}=69.9$ Hz, mean difference, 55.1 Hz; Wilcoxon sign rank test, $Z=33.1$, $p<0.0001$.]

A 2D histogram showing the upcall peak frequency (Hz) plotted against the peak frequency (Hz) of the BL noise 1 min before the call for each of the three habitat areas is shown in Fig. 7. In the BOF and Georgia, the peak frequency of the upcalls was always equal to or greater than the peak frequency of the noise. In Cape Cod Bay, some upcalls had their peak frequency below the peak frequency of the noise, though the majority of calls had higher peak frequency than the peak frequency of the noise. No correlation was found either between the duration of the upcall and the noise intensity peak in the 1 min before a detected call or between the duration of the upcall and the noise BL in the 1 min before a detected call ($r^2=0.007$ and 0.03, respectively).

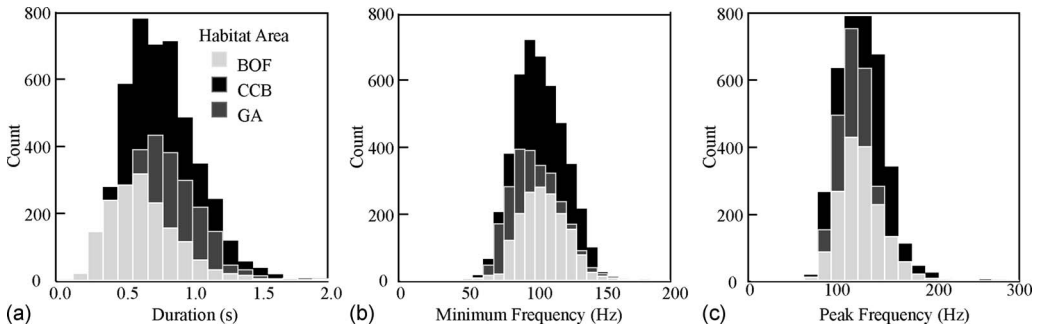


FIG. 6. Distribution of upcall parameters, (a) duration, (b) minimum frequency, and (c) peak frequency as measured for calls from each of the three habitat areas. BOF=Bay of Fundy, Canada; CCB=Cape Cod Bay, MA; and GA=Georgia. These figures combine data from both years in each habitat area.

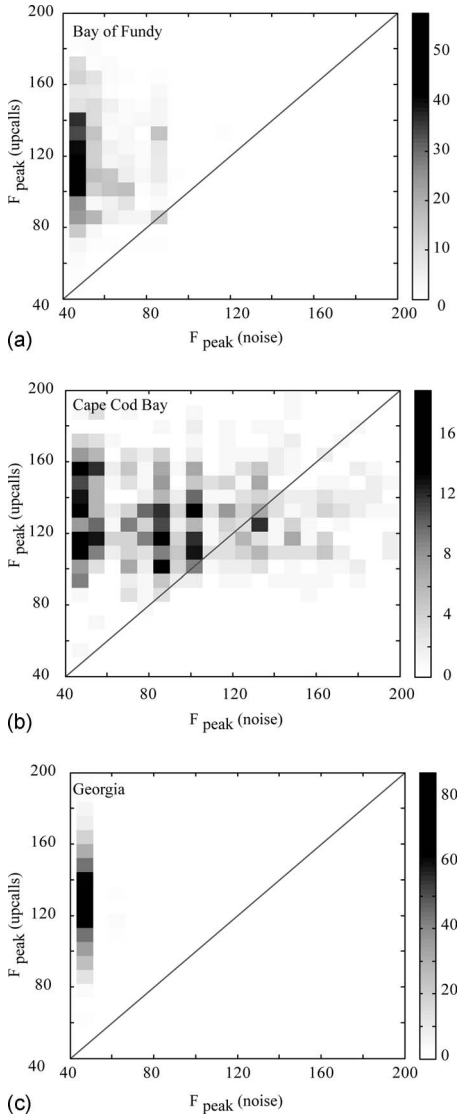


FIG. 7. A 2D histogram showing upcall peak frequency (Hz) plotted against the peak frequency (Hz) of the BL noise 1 min before the call for each of the three habitat areas; (a) BOF (2004), (b) Cape Cod Bay (2006), and (c) Georgia (2006). The number of values in each bin are indicated by the intensity of the shading shown on the scale to the right of the figures.

IV. DISCUSSION

There have been numerous studies investigating the impact of anthropogenic noise on marine mammals (see reviews in Richardson *et al.*, 1995; Nowacek *et al.*, 2007). The goal of this study was to investigate the typical levels of ambient noise to which North Atlantic right whales are exposed in three known habitat areas in order to quantify how "noisy" their environments are during different periods of the year. These measurements include both environmental and man-made sources of noise. Call parameters were measured for upcalls to investigate whether there were any detectable changes in call features related to the noise in the habitat. The results of this study demonstrate that there are regional variations in the noise levels to which right whales are exposed, with the highest average levels recorded in the BOF in the summer months, followed by Cape Cod Bay in the late winter and early spring, with the coast of Georgia having the lowest average noise levels. The detection of right whale calls in each habitat indicate the presence of right whales during the recording period but these recordings represent a small fraction of the known right whale habitat area throughout the year. There were only subtle differences in the distribution of upcall features among habitats, with shorter duration and higher average minimum frequencies of upcalls in the BOF than in the other two habitats, though it remains unclear whether these differences are a result of noise levels, habitat propagation characteristics, or variation in the particular individuals calling in the habitat. Overall, the minimum frequencies of upcalls in all habitat areas were higher than upcalls from Southern right whales in Argentina (Parks *et al.* 2007b).

A. Variability in ambient noise levels

There were clear differences among habitats in both the short-time spectrum levels and the BL of noise. Each habitat experienced a wide variation in BL noise, with changes of up to 30 dB re 1 μPa over a 24 h period (Fig. 2). Similarly, the short-time spectrum levels varied over 30 dB re 1 μPa/Hz (Fig. 4). The BOF consistently had higher noise levels in the frequency range of right whale contact calls (50–350 Hz) than the other two habitats (Fig. 3). The peak frequency of noise varied by habitat area, with the BOF and Georgia sites having lower overall average peak frequencies ($\sim<50$ Hz and between 50 and 75 Hz, respectively) than Cape Cod Bay (around 100 Hz) (Fig. 4). By investigating noise parameters in 2 years for each habitat area, it can be seen that these same trends have held between years. The BL of ambient noise in the BOF was higher than in Cape Cod Bay for both years, and Cape Cod Bay was higher than Georgia in both years. Similarly relative frequency distributions of the spectrum levels were consistent between years.

We suggest that between-habitat variability in human noise producing activities and propagation characteristics are primarily responsible for these observed differences in mode noise values. Environmental factors, including wind and precipitation events, also contribute to the overall noise levels in these recordings but were not monitored in our study. The level of commercial shipping traffic and the distance of the recording units from major shipping lanes varied considerably among these locations, with the BOF units being positioned closest to a major shipping lane and the Cape Cod Bay units being the furthest (Ward-Geiger *et al.*, 2005; Knowlton and Brown, 2007, Hatch *et al.* 2008). Therefore the highest noise levels documented in the BOF could be predicted by the closer proximity of the recording units to a high use commercial shipping lane.

Propagation characteristics likely vary significantly among the three recording areas in this study, which will affect the noise characteristics described here. Although all recordings occurred in relatively shallow water, the recordings in the BOF were made in significantly deeper water ($\sim$150 m) than either the Cape Cod Bay ($\sim$30 m) or Georgia ($\sim$15 m) recordings. The bottom types varied among the BOF (clay and sand/clay-silt) (Desharnais *et al.*, 2000), Cape Cod Bay (sand-clay/silt) and Georgia (sand) (U.S. Geological Survey, 2005). Modal dispersion predicts limited propa-

gation of sounds below 60 Hz in the Georgia site, 30 Hz in Cape Cod Bay, and 10 Hz in the BOF (Jensen *et al.*, 2000).

This study indicates that human activity may be dominating the average noise levels in the right whale upcall frequency range in each of these three areas. This is best demonstrated by the sizable reduction in the ambient noise levels in the Georgia location (Fig. 2), where the lowest noise levels (~200–240 h) coincide with December 27 and December 28. These dates span the period between the Christmas and New Year holiday and it is likely that there was an associated drop in vessel activity at this time. This is similar to previous observations of decreased ambient noise levels from a summer "weekend effect" of decreased commercial shipping activities (Ross, 1993) or an increase in ambient noise levels on weekends versus weekdays off Long Island, NY due to an increase in recreational vessel traffic (Samuel *et al.*, 2005).

B. Variability in call parameters

Differences were found in the values of all measured upcall characteristics by habitat area and by year. However, unlike the sizable differences in minimum frequency between North Atlantic and Southern right whale upcalls described in Parks *et al.* (2007b), this study found only subtle interhabitat differences in call characteristics. These results may partially reflect the large sample size of individual vocalizations measured here, allowing us to statistically distinguish small changes in the mean values of the measured parameters. Some of the measured calls are likely not truly independent events as they may have been produced by the same individual. Right whales are known to produce calls in bouts (Matthews *et al.*, 2001). Another factor to consider is that a statistically significant difference, in duration, for example, may not be biologically significant. A difference in mean value of duration on of the order of 5% between habitats is relatively minor given the variability and distribution of durations of the calls within each habitat (Fig. 6).

The mean peak frequency of calls was higher than the peak frequency of noise in the 1 min prior to the calls, indicating that right whales may be adjusting the frequency characteristics of their calls to accommodate their local ambient noise environment. This suggestion has been made previously regarding the structure of baleen whale vocalizations (Clark and Ellison, 2004). The measurements here demonstrate an upper limit to the peak frequency of the upcalls of $<$300 Hz for all three habitat areas. It is clear that right whales can produce tonal calls at frequencies well above 300 Hz (Parks and Tyack, 2005), so the explanation for this apparent upper frequency limit for the upcall cannot be based on the limitations of the whale's sound production mechanism. Potential explanations include, for example, a selective advantage in maintaining particular aspects of the upcall for individual recognition or for optimizing long-range communication. Given the similarity of the mean upcall peak frequencies of upcalls in all three habitats that have variable propagation environments and noise levels, it seems most likely that there is a social behavioral benefit conferred to whales that produce upcalls with peak frequencies below approximately 300 Hz. Under present ambient noise conditions, the upper frequency range of upcalls still provides whales a low noise window in which to produce calls that are higher in peak frequency than the local peak frequency of the ambient noise.

It is interesting to note that the trends in upcall and noise parameters described here are similar to the results from Parks *et al.* (2007b). The calls from all three North Atlantic right whale habitats described here were higher in frequency than those described for Southern right whales in Parks *et al.* (2007b). Based on the measurements of ambient noise levels, if right whales are modifying their calls in response to the BL of the ambient noise in the 50–350 Hz range, then the minimum frequency of calls would be expected to be highest in the BOF and lowest in Georgia. When data from both years are combined for each habitat area, the highest average minimum frequency across both recording years (105 Hz) was documented in the BOF, and the lowest average minimum frequency (99 Hz) was documented in Georgia (Fig. 6). However, the average minimum frequency in Georgia in the 2005–2006 recordings was 107 Hz and there was substantial overlap in the range of minimum frequencies of upcalls in all habitat areas (Fig. 6).

The clearest characteristics of calling behavior that could be related to the ambient noise trends came from a comparison of upcall peak frequency with the peak frequency of the noise in the 1 min period before the call. These results consistently indicated that the upcall peak frequency was higher than the peak frequency in the BOF and Georgia and for most upcalls recorded in Cape Cod Bay (Fig. 7). These results suggest that if whales are modifying calls as a result of noise, it is the peak frequency of the local ambient noise, rather than the absolute noise level that right whales are responding to. Another possibility is that there is a threshold of received low-frequency, ambient noise BL to which right whales are adapted and to which they respond by shifting their calls into a higher frequency band. If that is the case, it is quite possible that this BL noise threshold to which they are adapted is already exceeded in all three monitored habitat areas; a conclusion that is supported by the data showing that the average start frequency of calls in all three northern areas is relatively high compared to calls from southern right whales (Clark, 1982; Parks *et al.*, 2007b).

Differences in the propagation characteristics of these three western North Atlantic environments need to be taken into account. Different propagation conditions affect the levels and distributions of the ambient noise coming from distant sources (including whales) and as recorded on the receivers. It is possible that right whales can modify their calls to match the propagation conditions of their local environment. Further studies relating calling behavior to the propagation characteristics of the environment and quantification of the detection range for the calls in different habitat areas would be valuable in determining the influence of both noise levels and signal propagation conditions on baleen whale communication.

C. Significance of this study

The main goal of this study was to describe trends in the ambient noise to which North Atlantic right whales are exposed in their urban environment. The results here show that there is variability both within and between habitats in the levels of noise to which these whales are exposed. The mode noise value of 87 dB re 1 μPa/Hz at 100 Hz in the BOF is consistent with the Wenz (1962) ambient noise curves for heavy traffic noise (Fig. 4) and similar to measurements made in 1999 by Desharnais *et al.* (2000). Cape Cod Bay's values are higher than Wenz's (1962) reported usual shallow-water-traffic-noise value by 3–5 dB at 100 Hz, while Georgia's values are consistent with these values (Fig. 4).

In terms of the percentage of time each habitat is "loud," the ambient noise levels in the frequency range of right whale contact calls (50–350 Hz) in the BOF are below 105 dB re 1 μPa only between 5% (2004) and 15% (2005) of the time (Fig. 3). For Cape Cod Bay, these percentages are 37% (2005) and 47% (2006) and in Georgia 70% (2004) and 80% (2006) (Fig. 3). These results indicate that right whales in the BOF are dealing with higher levels of noise, and more often than in the other two habitat areas. This is of particular concern because the observed number of social groups is known to increase during the summer in the BOF (Parks *et al.*, 2007a). Right whales form and find these groups by using social acoustic communication signals (Kraus and Hatch, 2001; Parks and Tyack, 2005). Locating the mating grounds for right whales and quantifying the noise occurring in their breeding areas may be crucial in understanding how increases in ambient noise may limit the range of communication signals that are vital for successful reproduction.

ACKNOWLEDGMENTS

Special thanks go to the New England Aquarium right whale research group and the crew of the R/V Nereid; Michael Moore and the crew of the S/V Rostia; Alex Loer and the crew of the R/V Stellwagen; Christopher Tremblay, Chris Tessaglia-Hymes, Ward Krkoska, and Ingrid Biedron (Bioacoustics Research Program); Marc Costa and the crew of the R/V Shearwater (Provincetown Center for Coastal Studies) for assistance in deployment and recovery of the pop-up recording units; Melissa Fowler, Dimitri Ponirakis, and Ann Warde (Bioacoustics Research Program) for extraction and data analysis; and Kyle Becker for advice on shallow water propagation. Funding was provided by the National Oceanic and Atmospheric Administration, the Massachusetts Division of Marine Fisheries, and by Liberty Harbor Associates through a contract with Environmental Sciences Inc.

Aguilar Soto, N., Johnson, M., Madsen, P. T., Tyack, P. L., Bocconcelli, A., and Borsani, J. F. (**2006**). "Does intense ship noise disrupt foraging in deep-diving Cuvier's beaked whales (*Ziphius cavirostris*)?," Marine Mammal Sci. **22**, 690–699.

Andrew, R. K., Howe, B. M., and Mercer, J. A. (**2002**). "Ocean ambient sound: Comparing the 1960s with the 1990s for a receiver off the California coast," ARLO **3**, 65–70.

Best, P. B., Brandao, A., and Butterworth, D. S. (**2001**). "Demographic parameters of southern right whales off South Africa," J. Cetacean Res. Manage.Spec. Iss. **2**, 161–169.

Brown, M. W., Allen, J. M., and Kraus, S. D. (**1995**). "The designation of seasonal right whale conservation zones in the waters of Atlantic Canada," in *Marine Protected Areas and Sustainable Fisheries*, Proceedings of a Symposium on Marine Protected Areas and Sustainable Fisheries Conducted at the Second International Conference on Science and the Management of Protected Areas, edited by N. L. Shackell and M. J. H. Willison (Science and Management of Protected Areas Association, Wolfville, Nova Scotia, Canada).

Clark, C. W. (**1982**). "The acoustic repertoire of the southern right whale: A quantitative analysis," Anim. Behav. **30**, 1060–1071.

Clark, C. W., and Ellison, W. T. (**2004**). "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. A. Thomas, C. F. Moss, and M. Vater (The University of Chicago Press, Chicago, IL), pp. 565–582.

Clark, C. W., Borsani, J. F., and Notarbartolo-di-Sciara, G. (**2002**). "Vocal activity of fin whales, Balaenoptera physalus, in the Ligurian Sea," Marine Mammal Sci. **18**, 286–295.

Clark, C. W., Gillespie, D., Nowacek, D. P., and Parks, S. E. (**2007**). "Listening to their world: Acoustics for monitoring and protecting right whales in an urbanized ocean," in *The Urban Whale: North Atlantic Right Whales at the Crossroads*, edited by S. D. Kraus and R. M. Rolland (Harvard University Press, Cambridge, MA), pp. 333–357.

Desharnais, F., Laurinolli, M., Hay, A., and Theriault, J. A. (**2000**). "A scenario for right whale detection in the Bay of Fundy," in Oceans 2000, Providence, RI, pp. 1735–1742.

Etter, P. C. (**1991**). *Underwater Acoustic Modeling: Principles, Techniques and Applications* (Elsevier, London, UK).

Hatch, L., Clark, C., Merrick, R., Van Parijs, S., Ponirakis, D., Schwehr, K., Thompson, M., and Wiley, D. (**2008**). "Characterizing the relative contributions of large vessels to total ocean noise fields: A case study using the Gerry E. Studds Stellwagen Bank National Marine Sanctuary," Environ. Manage. (N.Y.) **42**, 735–752.

Jensen, F. B., Kuperman, W. A., Porter, M. B., and Schmidt, H. (**2000**). *Computational Ocean Acoustics* (Springer-Verlag, New York).

Knowlton, A. R., and Brown, M. W. (**2007**). "Running the gauntlet: Right whales and vessel strikes," in *The Urban Whale: North Atlantic Right Whales at the Crossroads*, edited by S. Kraus and R. M. Rolland (Harvard University Press, Cambridge, MA), pp. 409–435.

Knudsen, V. O., Alford, R. S., and Emling, J. W. (**1948**). "Underwater Ambient Noise," J. Mar. Res. **7**, 410–429.

Kraus, S. D., and Hatch, J. J. (**2001**). "Mating strategies in the North Atlantic right whale (*Eubalaena glacialis*)," J. Cetacean Res. Manage. , 237–244.

Kraus, S. D., and Kenney, R. D. (**1991**). "Information on right whales (*Eubalaena glacialis*) in three proposed critical habitats in United States waters of the western North Atlantic Ocean," National Technical Information Services, Washington, DC, p. 71.

Kraus, S., and Rolland, R. (**2007**). "Right whales in the Urban Ocean," in *The Urban Whale: North Atlantic Right Whales at the Crossroads*, edited by S. Kraus and R. Rolland (Harvard University Press, Cambridge, MA), pp. 1–38.

Kraus, S. D., Brown, M. W., Caswell, H., Clark, C. W., Fujiwara, M., Hamilton, P. K., Kenney, R. D., Knowlton, A. R., Landry, S., Mayo, C. A., McLellan, W. A., Moore, M. J., Nowacek, D. P., Pabst, D. A., Read, A. J., and Rolland, R. M. (**2005**). "North Atlantic right whales in crisis," Science **309**, 561–562.

Maronna, R. A., Martin, R. D., and Yohai, V. J. (**2006**). Robust Statistics: Theory and Methods (Wiley, Hoboken, NJ).

Marple, S. L. (**1987**). *Digital Spectral Analysis With Application* (Prentice-Hall, Englewood Cliffs, NJ).

Matthews, J. N., Brown, S., Gillespie, D., Johnson, M., McLanaghan, R., Moscrop, A., Nowacek, D., Leaper, R., Lewis, T., and Tyack, P. (**2001**). "Vocalisation rates of the North Atlantic right whale (*Eubalaena glacialis*)," J. Cetacean Res. Manage. **3**, 271–282.

McDonald, M. A., Hildebrand, J. A., and Wiggins, S. M. (**2006**). "Increase in deep ocean ambient noise in the Northeast Pacific west of San Nicolas Island, California.," J. Acoust. Soc. Am. **120**, 711–718.

Nowacek, D. P., Thorne, L. H., Johnston, D. W., and Tyack, P. L. (**2007**). "Responses of cetaceans to anthropogenic noise," Mammal Rev. **37**, 81–115.

Parks, S. E., and Clark, C. W. (**2007**). "Acoustic communication: Social sounds and the potential impacts of noise," in *The Urban Whale: North Atlantic Right Whales at the Crossroads*, edited by S. Kraus and R.

Rolland (Harvard University Press, Cambridge, MA), pp. 310–332.
Parks, S. E., and Tyack, P. L. (**2005**). "Sound production by North Atlantic right whales (*Eubalaena glacialis*) in surface active groups," J. Acoust. Soc. Am. **117**, 3297–3306.
Parks, S. E., Brown, M. W., Conger, L. A., Hamilton, P. K., Knowlton, A. R., Kraus, S. D., Slay, C. K., and Tyack, P. L. (**2007a**). "Occurrence, composition, and potential functions of North Atlantic right whale (Eubalaena glacialis) surface active groups," Marine Mammal Sci. **23**, 868–887.
Parks, S. E., Clark, C. W., and Tyack, P. L. (**2007b**). "Short and long-term changes in right whale calling behavior: The potential effects of noise on communication," J. Acoust. Soc. Am. **122**, 3725–3731.
Payne, R. S., and Webb, D. (**1971**). "Orientation by means of long range acoustic signaling in baleen whales," Ann. N.Y. Acad. Sci. **188**, 110–141.
Richardson, W. J., Greene, C. R.Jr., Malme, C. I., and Thomson, D. H. (**1995**). *Marine Mammals and Noise* (Academic, San Diego, CA).
Ross, D. (**1993**). "On ocean underwater ambient noise," Acoustic. Bull. **18**, 5–8.
Samuel, Y., Morreale, S. J., Clark, C. W., Greene, C. H., and Richmond, M. E. (**2005**). "Underwater, low-frequency noise in a coastal sea turtle habitat," J. Acoust. Soc. Am. **117**, 1465–1472.
Urazghildiiev, I. R., and Clark, C. W. (**2006**). "Acoustic detection of North Atlantic right whale contact calls using the generalized likelihood ratio test," J. Acoust. Soc. Am. **120**, 1956–1963.
Urazghildiiev, I. R., and Clark, C. W. (**2007**). "Acoustic detection of North Atlantic right whale contact calls using the spectrogram-based statistics," J. Acoust. Soc. Am. **122**, 769–776.
Urazghildiiev, I. R., Clark, C. W., and Krein, T. (**2008**). "Detection and recognition of North Atlantic right whale contact calls in the presence of ambient noise," Can. Acoust. **36**, 111–117.
Urick, R. J. (**1977**). "Models for the amplitude fluctuations of narrow-band signals and noise in the sea," J. Acoust. Soc. Am. **62**, 878–887.
Urick, R. J. (**1983**). *Principles of Underwater Sound* (Peninsula, Los Altos, CA).
U.S. Geological Survey (**2005**). Open-File Report No. 2005-1001, Woods Hole Science Center, Woods Hole, MA.
Ward-Geiger, L. I., Silber, G. K., Baumstark, R. D., and Pulfer, T. L. (**2005**). "Characterization of ship traffic in right whale critical habitat," Coastal Manage. **33**, 263–278.
Wenz, G. M. (**1962**). "Acoustic ambient noise in the ocean: Spectra and sources," J. Acoust. Soc. Am. **34**, 1936–1956.
Zakarauskas, P., Chapman, D. M. F., and Stall, P. R. (**1990**). "Underwater acoustic ambient noise levels on the eastern Canadian continental shelf," J. Acoust. Soc. Am. **87**, 2064–2071.

Beamwidth measurement of individual lithotripter shock waves

Wayne Kreider[a] and Michael R. Bailey
Center for Industrial and Medical Ultrasound, Applied Physics Laboratory, University of Washington, 1013 Northeast 40th Street, Seattle, Washington 98105

Jeffrey A. Ketterling
Frederic L. Lizzi Center for Biomedical Engineering, Riverside Research Institute, 156 William Street, New York, New York 10038

(Received 10 January 2008; revised 13 August 2008; accepted 17 November 2008)

New lithotripters with narrower foci and higher peak pressures than the original Dornier HM3 electrohydraulic lithotripter have proven to be less effective and less safe. Hence, accurate measurements of the focal characteristics of lithotripter shock waves are important. The current technique for measuring beamwidth requires a collection of single-point measurements over multiple shock waves, thereby introducing error as a result of any shock-to-shock variability. This work reports on the construction of a hydrophone array sensor and on array measurements of individual lithotripter shock waves. Beamwidths for an electrohydraulic lithotripter with a broad-focus HM3-style reflector and a narrow-focus modified reflector were measured using both new and worn electrodes as well as two different electrical charging potentials. The array measured the waveform, beamwidth, and focal location of individual shock waves. The HM3-style reflector produced repeatable focal waveforms and beam profiles at an 18 kV charging potential with new and worn electrodes. Corresponding measurements suggest a narrower beamwidth than reported previously from averaged point measurements acquired under the same conditions. In addition, a lack of consistency in the measured beam profiles at 23 kV underscores the value of measuring individual shock waves. © *2009 Acoustical Society of America.* [DOI: 10.1121/1.3050272]

PACS number(s): 43.80.Vj [CCC] Pages: 1240–1245

I. INTRODUCTION

Shock-wave lithotripsy (SWL) is the most common treatment for uncomplicated renal stones.[1] Several recent studies have reported that when compared to outcomes with the original Dornier HM3 electrohydraulic lithotripter, new lithotripters have lower stone-free rates, increased retreatment rates, and increased severity and frequency of trauma to surrounding tissue.[2–5] The differences in clinical outcomes may relate to the method of acoustic coupling to the patient (newer machines utilize a water filled pillow, while the HM3 utilizes a water bath).[6] However, it is generally accepted that the tighter focal geometries and higher peak pressures of the newer machines have had great impact on clinical outcomes. More specifically, tissue injury is known to correlate with peak pressure, while the efficacy of targeting stones is likely reduced for shock waves with narrower focal regions.[7] Moreover, experiments have also found that lithotripters with focal regions that are wide relative to the stone (e.g., the HM3) produce more effective stone fragmentation. The utility of a broad focal beamwidth lies in the ability of a shock wave traveling along the stone perimeter to generate shear waves within the stone.[8–10] In response to clinical outcomes and experimental results, many manufacturers have released new lithotripters specifically marketed as "broad-focus."

The inherent shock-to-shock variability of the acoustic field in spark-source electrohydraulic lithotripters makes it very difficult to characterize instantaneous beamwidths. Typically, beamwidths are estimated from the sequential acquisition of point measurements within the focal plane using a single-element hydrophone. However, shock waves change significantly due to variations in spark energy and in the location of the spark between the electrodes.[11] These variabilities can alter the waveform shape in terms of rise time and pulse width. Moreover, the location of the focal region may shift. Accordingly, the combination of single-element hydrophone measurements from a series of different shock waves may yield broader estimates of beamwidth than are actually present. Given the clinical motivation described above for knowing the shock-wave beamwidth in lithotripsy, improved measurement tools are needed.

In this work, a linear array hydrophone was constructed and used to make instantaneous beamwidth measurements of individual lithotripter shock waves. Measurements were made with broad- and narrow-focus electrohydraulic lithotripters as well as with new and worn electrodes at different charging potentials. The different conditions tested were designed to ensure variability in both focal beamwidths and peak pressures.

II. METHODS

A. Hydrophone array

The hydrophone used in this effort was designed to be a linear array that roughly spans the −6 dB beamwidth of HM3-style lithotripters. From averaged single-point measurements in the same APL-UW lithotripter used in this ef-

[a]Electronic mail: wkreider@u.washington.edu

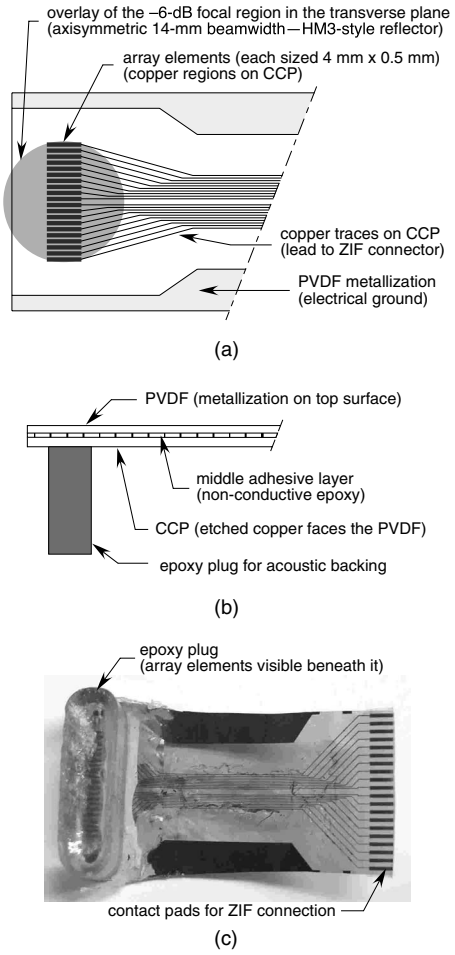


FIG. 1. The hydrophone array as sketched in (a) plan view and (b) elevation view. Part (c) comprises a photograph of the array as viewed from beneath the epoxy plug shown in (b). Note that some residual adhesive used in fixturing the array remains visible in the photograph. The array consisted of 20 PVDF elements that were 0.5×4 mm^2, with a center-to-center element spacing of 0.7 mm. Note that the array elements and the overlay of the focal region for the HM3-style reflector are drawn to scale in (a). For the HM3-style reflector, the overlay represents a previously reported beamwidth that corresponds to the array orientation used in the present measurements.

fort, Cleveland *et al.*[12] reported peak pressures in the focal plane. Linear interpolation of these data implies a −6 dB beamwidth of about 14 mm for the HM3 configuration. As depicted in Fig. 1(a), the array comprised 20 elements, each 4 mm long by 0.5 mm wide. The center-to-center element spacing was 0.7 mm, leading to a total array width of 13.8 mm. The 0.5 mm element widths were chosen to allow high spatial resolution along the length of the array; the larger 4 mm lengths were chosen to maintain higher element surface areas for impedance matching considerations.

This basic array geometry was fabricated using a technique described by Ketterling *et al.*[13] A functional schematic and a photograph of the fabricated device are provided in parts (b) and (c) of Fig. 1. The transduction was performed by a 9 μm polyvinylidene fluoride (PVDF) membrane with one side electroded in gold (Ktech Corp., Albuquerque, NM). The gold electroding provided a common ground for all array elements. Using a thin layer of nonconductive epoxy that typically remained about 1 μm thick after curing, the other side of the PVDF membrane was bonded to a copper-clad polyimide (CCP) film (RFlex 1000L810, Rogers Corp., Chandler, AZ). The actual pattern of array elements and electrical trace lines linking each element to a connection pad was etched onto the CCP using standard printed circuit board techniques. Lastly, the trace pads were spaced to fit into a standard zero-insertion-force (ZIF) flex connector. During the bonding process, the PVDF and CCP layers were clamped between two aluminum plates. After the bonding epoxy cured, a Teflon mold of 15 mm depth was attached to the CCP and filled with additional epoxy. Although not explicitly depicted in the figure, a custom printed circuit board linked the ZIF connector to 20 BNC connectors. Operating characteristics including center frequency and bandwidth have been reported in the literature for a similar device.[13]

Given the design details described above, it is instructive to consider expected performance characteristics of the hydrophone array. Since the bandwidth of PVDF is greater than 70 MHz, shock fronts with rise times of at least 14 ns can theoretically be resolved. However, angular misalignment of the array with the shock front will effectively average the shock over the element surface and increase its apparent rise time. For example, a 5° inclination of the 4-mm-long elements would lead to an apparent 200 ns rise time for a planar shock wave. In addition, spatial variations of the incident acoustic field will be averaged over the area of each element. Although little spatial averaging is expected along the 0.5 mm dimension of each element, the 4 mm dimension represents a nontrivial fraction of the nominal 14 mm beamwidth. Assuming an axisymmetric acoustic field in the plane of the array, averaging along the 4 mm dimension will effectively reduce the spatial resolution relative to that implied by the 0.5 mm element widths.

B. SWL measurements

Measurements were made in degassed water, with the hydrophone array placed at the focus of the APL-UW research lithotripter modeled after the Dornier HM3.[12] The array was oriented at the focus such that the 4×0.5 mm^2 area of each element was approximately perpendicular to the axis of propagation of the shock wave. Moreover, the array was centered such that the middle elements were aligned at the geometric focus of the lithotripter. A charging potential of 15, 18, or 23 kV triggered single shock waves at a rate slower than 1/min. Two ellipsoidal reflectors were used: an HM3-style reflector (semimajor and semiminor axes: $a = 13.80$ cm, $b = 7.75$ cm) and a reflector insert that fit inside the HM3-style reflector ($a = 9.30$, $b = 6.24$). Both reflectors were axisymmetric in that they did not possess the flouroscopy probe cutouts that are typically found in clinical lithotripters. The reflector insert was designed to create a tighter focus, as illustrated in Fig. 2. New electrodes (<200 shock waves) and worn electrodes at the end of their prescribed clinical lifetime (>2000 shock waves) were used. Worn electrodes possessed larger spark gaps and exhibited damage caused by arcing along a broad portion of their surface. As mentioned above, testing with new and worn electrodes ensured variability in the focal pressures.

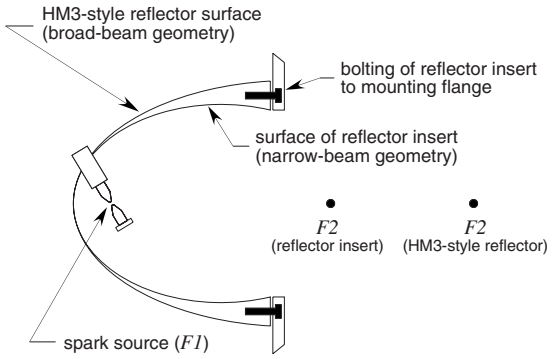


FIG. 2. Scaled drawing of the tested lithotripter geometries. The ellipsoidal reflectors possess two foci, where $F1$ denotes the focus corresponding to the spark source and $F2$ denotes the remote focus corresponding to the treatment site and the central location at which beamwidth measurements were acquired.

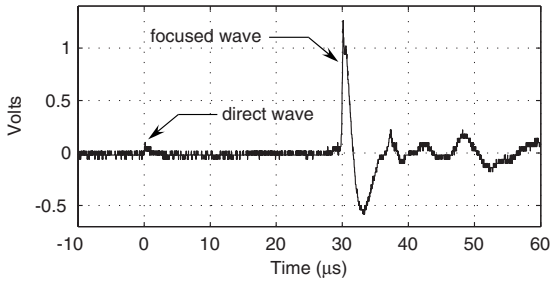


FIG. 3. Measured direct and focal shock waves from a single array element.

Voltage measurements from the array elements were captured on six digital oscilloscopes with sampling rates of at least 50 MHz. Measurements were collected using an input impedance of 1 MΩ on the oscilloscopes and no preamplification. A custom LabVIEW program (National Instruments, Austin, TX) was used to digitally store select waveforms. Oscilloscope measurements corresponding to peak positive pressures of the focal waves were manually recorded for all array elements.

C. Relative sensitivities of array elements

Although array elements were not calibrated for absolute pressure measurements, the relative sensitivity of each element was determined using two independent methods. The first approach utilized the fact that a direct wave diverges spherically from the spark at the $F1$ focus. For the HM3-style reflector, the second focus ($F2$) is 228 mm away from $F1$. Considering the 13.8 mm width of the array and assuming that the direct wave is lossless and axisymmetric, the amplitude of the direct wave should vary by less than 0.1% across the array. Based on such geometric considerations, the amplitude of the direct wave was taken to be constant for all array elements for each incident shock wave. As such, measurements of the direct wave were used to determine the relative sensitivities of each element.

In the second approach, each element of the array was exposed to a controlled low-amplitude sound field. The sound field was produced by a focused transducer with a 0.375 in. diameter and a 2 in. focal length (Panametrics V326, Olympus NDT Inc., Waltham, MA) excited with a 5 MHz, 10 cycle tone burst. The transducer was mounted to motorized translation stages and could be moved precisely from element to element. For each measurement, the transducer was scanned across an element in 200 μm increments, and an average of 100 tone bursts were acquired at a sampling rate of 100 MHz. The maximum peak amplitude was then extracted from the scan data to represent the relative sensitivity of the element. Given that the −6 dB lateral beamwidth of the transducer was 1.6 mm, the 200 μm scan spacing was more than sufficient to capture the relative maximum of each element with negligible positioning error. Accordingly, the maximum voltages from each element were compared to evaluate relative sensitivities. This second approach was pursued after acquisition of beamwidth measurements in order to provide an independent check of element sensitivities.

Ultimately, measurements of focal shock waves were normalized relative to the direct-wave sensitivities. First, varying sensitivity among array elements was addressed by scaling the responses of all elements to match that of the most sensitive element. Then, for each element, the peak voltage from the focal wave was divided by the peak voltage of an "average" direct wave generated with new electrodes at 18 kV. In lieu of an absolute pressure calibration, this normalization approach enables a consistent comparison of results. Moreover, the focal-to-direct-wave ratio can be compared to previous calibrated measurements. From the same APL-UW lithotripter with the HM3-style reflector and a charging potential of 18 kV, Cleveland *et al.*[12] reported peak positive focal pressures of 29.9 MPa with a standard deviation of 4.7 MPa. With an HM3 spark source at a charging potential of 20 kV, Coleman *et al.* measured peak-positive, direct-wave amplitudes between 1 and 3 MPa at $F2$.[11] With a comparable source at the same charging potential, Müller measured the direct wave closer to the source; assuming a $1/r$ scaling, the measured amplitude was equivalent to 0.8 MPa at $F2$ of the HM3-style geometry.[14] From previous measurements, in addition to the somewhat lower 18 kV charging potential in the present work, direct-wave amplitudes near 1 MPa can be expected. Accordingly, focal-to-direct ratios may be interpreted to roughly represent megapascal.

III. RESULTS

A. Array performance

Individual array elements were able to measure waveforms consistent with those previously reported for the HM3. A sample waveform acquired while using the HM3-style reflector and new electrodes is shown in Fig. 3. The sensitivity of the hydrophone array was sufficient to reveal the direct wave seen at $t=0$ as well as the focused wave at $t\approx 30$ μs. Typical waveforms exhibited a peak positive spike of about 1 μs duration followed by a negative trough of about 4 μs, which is the commonly described classic lithotripter waveform.[11,12] The shock front was slightly rounded and not strongly shocked. As discussed above, angular misalignment

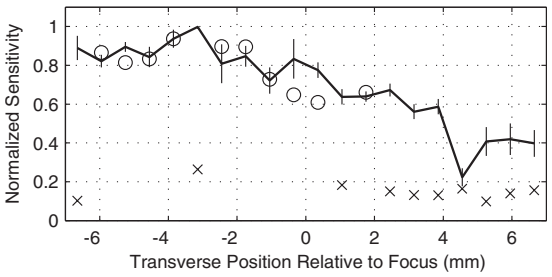


FIG. 4. Normalized element sensitivities. The solid line indicates the element-by-element mean of measurements from six direct waves, while the vertical bars indicate ±1 standard deviation at each element. The x and o markers represent sensitivities determined from exposure of the array to a low-amplitude, 5 MHz source.

of the array such that its elements were not parallel to the shock front is a likely explanation for the slow rise times observed.

As described in the previous section, data were collected to evaluate the relative sensitivities of array elements. Measurements of direct waves were recorded using new electrodes with charging potentials of 18 or 23 kV for the HM3-style reflector and 15 or 18 kV for the reflector insert. Note that direct-wave amplitudes were resolved with different oscilloscope settings from those used for the data in Fig. 3. While all of the direct-wave data exhibited consistent sensitivities for each element, data from the reflector insert were collected while one of the elements was not producing a signal. Hence, the remaining data from six shocks were normalized and averaged to determine relative sensitivities. Each direct-wave measurement was normalized across all elements so that the most sensitive element possessed a value of unity. An average of normalized measurements is plotted as the solid line in Fig. 4, where the vertical bars represent ±1 standard deviation for each element. As implied by the plot, the same element at $x=-3.15$ mm was effectively the most sensitive for all measurements. In addition, even though sensitivity varied significantly among array elements, the relative sensitivity of each element was consistent from shock to shock.

To independently confirm the direct-wave data, relative sensitivity measurements were also acquired using a low-amplitude, 5 MHz source. Using the same normalization described above, these data are plotted in Fig. 4 with "x" and "o" markers. The o measurements show good quantitative agreement with the direct-wave data. In contrast, the elements corresponding to the x measurements exhibited a consistent, but low, sensitivity near 0.2. One of these elements (at $x=+4.55$ mm) had a similarly low sensitivity from the direct-wave measurements, while the others previously had higher sensitivities.

From the evaluation of sensitivity measurements, two qualitative element behaviors can be identified. These behaviors comprise high-sensitivity responses (above about 0.4) and low-sensitivity responses ($\lesssim$0.2). As noted above, one array element exhibited no response during the acquisition of some direct-wave measurements and subsequent measurements of focal shock waves. Interestingly, this same element (at $x=+2.45$ mm) exhibited a low but consistent response

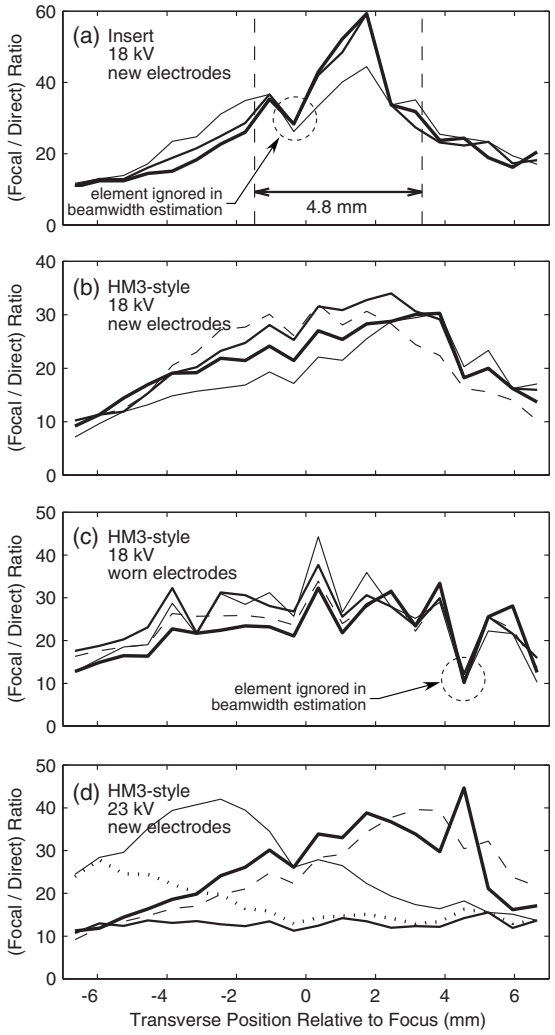


FIG. 5. Beam profiles measured for individual shock waves. Shock waves were generated using (a) the reflector insert, a charging potential of 18 kV, and new electrodes (three shock waves); (b) the HM3-style reflector, 18 kV, and new electrodes (four shock waves); (c) the HM3-style reflector, 18 kV, and worn electrodes (four shock waves); and (d) the HM3-style reflector, 23 kV, and new electrodes (five shock waves). Note that direct-wave measurements at 18 kV were used for all normalizations, thereby allowing a direct comparison of the amplitudes in all plots. The −6 dB beamwidth of the thickest-line profile is marked in (a).

23 months later when additional sensitivity data were acquired at 5 MHz. Based on a visual observation of the PVDF surface, the drops in sensitivity for the x elements appear to be related to cavitation- or stress-related damage to either the bond epoxy between the PVDF and CCP or the epoxy-plug backing. However, the PVDF itself still remained acoustically active for the x elements, as evidenced by their nonzero responses, thereby lending credence to the hypothesis that the PVDF had separated from its CCP backing layer.

B. Beam profiles

After correcting for individual element sensitivities as described above, the beam profiles of individual shock waves are presented in Fig. 5. Because all measurements were normalized relative to the direct wave from the HM3-style reflector at 18 kV, all plotted amplitudes can be directly compared. As shown, measured profiles were sorted based on test

TABLE I. Measurements from individual shock waves.

Conditions		Beamwidth (mm) Mean	Beamwidth (mm) Range	Peak location (mm) Range[a]
(a)	Insert reflector 18 kV, new electrodes (three shock waves)	6.3	4.9	0.0
(b)	HM3-style reflector 18 kV, new electrodes (four shock waves)	10.2	2.0	3.5
(c)	HM3-style reflector 18 kV, worn electrodes (four shock waves)	11.1	4.3	3.5
(d)	HM3-style reflector 23 kV, new electrodes (five shock waves)	···	···	11.2

[a]Denotes the maximum minus the minimum observed value for measurement of either beamwidth or location of peak pressure.

conditions. Focused beams were produced using four distinct combinations of either the HM3-style or the insert reflector, a charging potential of 18 or 23 kV, and new or worn electrodes. In Fig. 5(a), the −6 dB beamwidth of the profile corresponding to the thickest line is explicitly marked and labeled. Note that dashed circles in (a) and (c) denote array elements that consistently measured lower pressures than their neighbors and were ignored in the estimation of beamwidths for these test conditions.

Each of the profiles plotted in Fig. 5 was used to estimate the −6 dB beamwidth and the transverse location of the peak pressure. These data for independently measured shock waves are provided in Table I, in which "range" for either beamwidth or location of peak pressure denotes the maximum minus the minimum observed value. Comparing the four separate groups of shock-wave measurements, several key differences are apparent. First, from test conditions (a) and (b), the insert reflector produced narrower beamwidths, higher peak pressures, and a more consistent localization of the peak pressure than did the HM3-style reflector. In Table I, identical peak-pressure locations were measured for all shock waves from condition (a), thereby yielding a range of zero for peak location. Despite measuring a low number of shock waves, the beamwidths estimated for conditions (a) and (b) are statistically different at a 95% confidence level ($p \approx 0.04$ when beamwidths are assumed to come from normally distributed samples with equal variances). Considering conditions (b) and (c) for the HM3-style reflector, worn electrodes altered the shape of beam profiles and generated slightly higher peak pressures than new electrodes at a charging potential of 18 kV.

At a charging potential of 23 kV with new electrodes, the HM3-style reflector tended to produce higher peak pressures than at 18 kV, as expected. However, shock waves generated at 23 kV also exhibited much greater variability. Given the inconsistency of measured beam profiles shown in Fig. 5(d), corresponding beamwidths were not estimated. As for the observed locations of peak pressure, the greater variability is demonstrated by the 11.2 mm range, inasmuch as peak pressures occurred near both ends of the array. Assuming that peak-pressure locations are normally distributed, the variances of conditions (b) and (d) are statistically different at a confidence level of 89%. Hence, the location of peak pressure was quantitatively less consistent in case (d). This result again demonstrates that the array can be used to discern system variability based on only a few measured shock waves.

In addition to the above comparisons among test conditions used in this effort, measurements with the HM3-style reflector can also be compared to data derived from single-point hydrophone measurements. The −6 dB beamwidth of the HM3-style reflector for the APL-UW lithotripter was previously reported to be about 14 mm.[12] For the relevant conditions tested in this effort, average beamwidths were found to be 10–11 mm. Although this difference may not be statistically significant, these results are consistent with the expectation that averaging of single-point measurements from multiple shock waves leads to an overestimation of beamwidth. Aside from beamwidth, a loose comparison of peak positive pressures can also be made. Single-point data indicated pressures of 29.9 ± 4.7 MPa at 18 kV for the HM3-style lithotripter used in this effort.[12] Such a pressure range is fully consistent with the array profiles shown in Fig. 5(b) if the aforementioned direct-wave amplitude of 1 MPa is assumed.

IV. DISCUSSION AND CONCLUSIONS

A linear hydrophone array constructed and tested in this effort was useful for quickly and precisely measuring the beamwidths of focused shock waves. Data described above represent the first simultaneous acoustic field measurements of individual lithotripter shock waves. Previous data from the literature consisted of point measurements at a single spatial location for each shock wave, thereby requiring averages over multiple shock waves in order to assess the beam characteristics of focal waves. Given the inherent shock-to-shock variabilities of electrohydraulic lithotripters, simultaneous field measurements provide the unique benefit of avoiding any inaccuracies introduced by averaging over multiple shocks. Indeed, even though relatively few shock waves were measured in this effort, the resulting data suggest that beamwidths for the APL-UW lithotripter may be about 25% narrower than previously reported from averaged measurements. This discrepancy may be due to averaging effects and/or the use of a much coarser spatial resolution for the single-point measurements (data were collected on a 5 mm spacing in the focal plane[12]). A more definitive assessment of beamwidths requires further investigation.

In spite of the inherent variabilities associated with spark jitter in electrohydraulic lithotripters, measured focal beams at 18 kV were fairly similar for given conditions [see Figs. 5(a)–5(c)]. Such repeatability is consistent with previous observations of shock-induced fountains.[15] However, the data at 23 kV exhibit much less uniformity. Although spark jitter alone may explain the increased variability at higher charging potentials, Pishchalnikov *et al.*[16] documented the potential for bubbles to distort the acoustic field. In addition, propagation through tissue inhomogeneities may affect the

characteristics of focal shock waves under clinical conditions. More specifically, scattering associated with random variations in the acoustic propagation path has been shown to reduce the amplitude of certain components of focal lithotripter waveforms.[6,17,18] However, it is still not known whether the scattered energy can be refocused elsewhere in the tissue. As demonstrated by the data at 23 kV, a hydrophone array can identify disparate focal locations of individual shocks and is therefore uniquely suited to address such questions of refocusing.

The array used in this effort proved fairly consistent and robust. However, despite the measurement of consistent beam profiles at 18 kV, changes in sensitivity of some elements over time imply that some damage did occur. From the data presented in Fig. 4, it appears that element sensitivities were typically repeatable even though they tended to eventually drop to a low-sensitivity state. While further investigation of manufacturing variabilities and damage modes is warranted, the demonstrated utility of direct-wave characterization suggests a strategy for acquiring future measurements. If array signals are digitized with sufficient bit depth and temporal length to capture both direct and focal waves, element sensitivities could be explicitly monitored for all measurements. Such an approach is appealing from perspectives of simplicity and quality assurance.

Although the hydrophone design implemented in this effort was successful in proving the feasibility of array measurements in SWL, improvements could be made in future designs. With regard to durability, half of the elements suffered a drop in sensitivity after exposure to approximately 100 shock waves and subsequent aging over about two years. Based on experience with other hydrophones in similar acoustic fields, damage may have been caused by bubble collapses at the surface of the PVDF. Hence, modifications of the array to include a protective coating or immersion in oil might be useful to improve consistency and robustness. In addition, fabrication techniques to yield a more uniform epoxy thickness between the PVDF and the CCP can be pursued to potentially improve element-to-element consistency as well as durability. Ultimately, the life expectancy of such hydrophone arrays may be on the order of 1000 shocks; however, even a shorter lifetime on the order of 100 shocks would be reasonable given an approximate fabrication cost of less than $400. As a final comment on array design, the fabrication technique allows adjustment of the number of elements, element size, and element spacing, thereby enabling performance to be tailored to specific lithotripters.

ACKNOWLEDGMENTS

The authors thank our collaborators at the Center for Industrial and Medical Ultrasound and the Consortium for Shock Waves in Medicine. Specifically, we thank Aaron Midkiff (Department of Electrical Engineering, University of Washington) for help in the initial testing of the array and Professor Robin Cleveland (Department of Aerospace and Mechanical Engineering, Boston University) for helpful discussions. This work was supported by the Riverside Research Institute Fund for Biomedical-Engineering Research, the National Institutes of Health (DK 43881), and the National Space Biomedical Research Institute (SMS00402).

[1] M. S. Pearle, E. A. Calhoun, and G. C. Curhan, "Urologic diseases in America project: Urolithiasis," J. Urol. (Baltimore) **173**, 848–857 (2005).
[2] A. P. Evan, J. A. McAteer, J. C. Williams, L. R. Willis, M. R. Bailey, L. A. Crum, J. E. Lingeman, and R. O. Cleveland, "Shock wave physics of lithotripsy: Mechanisms of shock wave action and progress toward improved SWL," in *Textbook of Minimally Invasive Urology*, edited by R. Moore, J. Bishoff, S. Loening, and S. Docimo (Martin Dunitz Limited, London, 2004), pp. 425–438.
[3] K. Kerbl, J. Rehman, J. Landman, D. Lee, C. Sundaram, and R. V. Clayman, "Current management of urolithiasis: Progress or regress?," J. Endourol **16**, 281–288 (2002).
[4] J. E. Lingeman, "Extracorporeal shock wave lithotripsy devices: Are we making progress?," in *New Developments in the Management of Urolithiasis*, edited by J. E. Lingeman and G. M. Preminger (Igaku-Shoin, New York, 1996), pp. 79–96.
[5] J. A. McAteer, M. R. Bailey, J. C. Williams, Jr., R. O. Cleveland, and A. P. Evan, "Strategies for improved shock wave lithotripsy," Minerva Urol. Nefrol **57**, 271–287 (2005).
[6] Y. A. Pishchalnikov, J. S. Neucks, R. J. VonDerHaar, I. V. Pishchalnikova, J. C. Williams, Jr., and J. A. McAteer, "Air pockets trapped during routine coupling in dry head lithotripsy can significantly decrease the delivery of shock wave energy," J. Urol. (Baltimore) **176**, 2706–2710 (2006).
[7] R. O. Cleveland, R. Anglade, and R. K. Babayan, "Effect of stone motion on in vitro comminution efficiency of Storz Modulith SLX," J. Endourol **18**, 629–633 (2004).
[8] R. O. Cleveland and O. A. Sapozhnikov, "Modeling elastic wave propagation in kidney stones with application to shock wave lithotripsy," J. Acoust. Soc. Am. **118**, 2667–2676 (2005).
[9] W. Eisenmenger, "The mechanisms of stone fragmentation in ESWL," Ultrasound Med. Biol. **27**, 683–693 (2001).
[10] O. A. Sapozhnikov, A. D. Maxwell, B. MacConaghy, and M. R. Bailey, "A mechanistic analysis of stone fracture in lithotripsy," J. Acoust. Soc. Am. **121**, 1190–1202 (2007).
[11] A. J. Coleman, J. E. Saunders, R. C. Preston, and D. R. Bacon, "Pressure waveforms generated by a Dornier extra-corporeal shock-wave lithotripter," Ultrasound Med. Biol. **13**, 651–657 (1987).
[12] R. O. Cleveland, M. R. Bailey, N. Fineberg, B. Hartenbaum, M. Lokhandwalla, J. A. McAteer, and B. Sturtevant, "Design and characterization of a research electrohydraulic lithotripter patterned after the Dornier HM3," Rev. Sci. Instrum. **71**, 2514–2525 (2000).
[13] J. A. Ketterling, O. Aristizábal, D. H. Turnbull, and F. L. Lizzi, "Design and fabrication of a 40-MHz annular array transducer," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **52**, 672–681 (2005).
[14] M. Müller, "Dornier Lithotripter im Vergleich Vermessung der Stoßwellenfelder und Fragmentationswirkung (Comparison of Dornier Lithotripters—Measurement of shock wave fields and fragmentation effectiveness)," Biomed. Tech. **35**, 250–262 (1990).
[15] O. A. Sapozhnikov, M. R. Bailey, and L. A. Crum, "Shot-to-shot variability of acoustic axis of a spark-source lithotripter," J. Acoust. Soc. Am. **105**, 1269 (1999).
[16] Y. A. Pishchalnikov, J. A. McAteer, W. Kreider, J. C. Williams, Jr., I. V. Pishchalnikova, and M. R. Bailey, "Influence of cavitation on lithotripter shock waves," in *Proceedings of the 19th International Congress on Acoustics*, Madrid, Spain, September 2–7, 2007; URL: http://www.seaacustica.es/WEB_ICA_07/fchrs/papers/nla-02-005.pdf (Last viewed August 13, 2008).
[17] Y. A. Pishchalnikov, O. A. Sapozhnikov, M. R. Bailey, I. V. Pishchalnikova, J. C. Williams, Jr., and J. A. McAteer, "Cavitation selectively reduces the negative-pressure phase of lithotripter shock pulses," ARLO **6**, 280–286 (2005).
[18] R. O. Cleveland, D. A. Lifshitz, B. A. Connors, A. P. Evan, L. R. Willis, and L. A. Crum, "In vivo pressure measurements of lithotripsy shock waves in pigs," Ultrasound Med. Biol. **24**, 293–306 (1998).

Erratum: "Low-frequency attenuation of acoustic waves in sandy/silty marine sediments" [J. Acoust. Soc. Am., 124, EL308–EL312 (2008)]

Allan D. Pierce and William M. Carey
Department of Mechanical Engineering, Boston University, Boston, Massachusetts 02215

(Received 29 November 2008; accepted 4 December 2008)
[DOI: 10.1121/1.3056567]

PACS number(s): 43.30.Ma, 43.20.Jr, 43.10.Vx

The development in the subject paper on p. EL310 that includes Eq. (6) and the sentence preceding it should be replaced by the following:

Increase in pressure acting on any given material volume causes the volume to decrease, and the appropriate local measure of this tendency for a pure substance is the bulk modulus B, so that, for arbitrary small volume containing fixed mass,

$$\frac{d}{dt}(\Delta V) = -\frac{1}{B}\frac{dp}{dt}(\Delta V),$$

where p is the local pressure. To develop an appropriate analogous constitutive relation for a sediment, one considers a volume V moving such that there is no local mass transport across the confining surface S. The time rate of change of this volume can be alternately written

$$-\int_V \frac{1}{B}\frac{dp}{dt}dV = \int_s \boldsymbol{v}\cdot\boldsymbol{n}\,dS.$$

The volume is regarded as large compared to a grain size and sufficiently large that it is statistically representative of the sediment. In accordance with the arguments given above, and for disturbances that are acoustic, the pressure within the volume can be regarded as spatially uniform and equal to the local averaged pressure within the fluid. Also, the other quantities within the integrands can be replaced by their local averages just as is done in a previous portion of this paper Eq. (1). The velocity in the surface integral is replaced by an appropriately weighted average of the locally averaged velocities in the fluid and the solid, and the bulk modulus in the volume integral can be similarly replaced, with the results that

$$-\int_V \left(\frac{\chi_s}{B_s}+\frac{\chi_f}{B_f}\right)\frac{\partial}{\partial t}\langle p\rangle_f dV = \int_S [\chi_s\langle\boldsymbol{v}\rangle_s + \chi_f\langle\boldsymbol{v}\rangle_f]\cdot\boldsymbol{n}dS.$$

Gauss's theorem (or the divergence theorem) applied to the surface integral, along with the argument that the volume is "macroscopically arbitrary" subsequently yields the partial differential equation

$$\frac{\partial}{\partial t}\langle p\rangle_f = -B_{\text{eff}}[\chi_s\boldsymbol{\nabla}\cdot\langle\boldsymbol{v}\rangle_s + \chi_f\boldsymbol{\nabla}\cdot\langle\boldsymbol{v}\rangle_f]. \quad (6)$$

 0001-4966/2009/125(2)/1246/1/$25.00

Erratum: "Temporal coherence of sound transmissions in deep water revisited" [J. Acoust. Soc. Am., 124, 113-127 (2008)]

T. C. Yang
Navel Research Laboratory, 4555 Overlook Avenue S.W., Washington, DC 20375

(Received 18 November 2008; accepted 21 November 2008)
[DOI: 10.1121/1.3050300]

PACS number(s): 43.30.Re, 43.30.Zk, 43.10.Vx

Two typographical errors appear in the abstract. In lines 7 and 8 of the abstract, referring to the path integral prediction, the $-\frac{1}{2}$ power frequency dependence and the $-\frac{3}{2}$ power range dependence were incorrect. It should read a "-1 power frequency dependence and a $-\frac{1}{2}$ power range dependence were predicted" as stated in the text.

Erratum: "A parametric model of the vocal tract area function for vowel and consonant production" [J. Acoust. Soc. Am., 117, 3231-3254 (2005)]

Brad. H. Story[a)]
Speech Acoustics Laboratory, Department of Speech and Hearing, University of Arizona, Tuscon, Arizona 85721

(Received 31 October 2008; accepted 21 November 2008)
[DOI: 10.1121/1.3050290]

PACS number(s): 43.70.Bk, 43.71.Es, 43.10.Vx

The variable shown in the second column of Table III should read $\Omega(i)$ instead of $\omega(i)$. The coefficient values in Table IV should be replaced with the following values.

TABLE IV. Mode coefficients that reconstruct the indicated vowels when used with Eq. (3).

Vowel	q_1	q_2
i	−5.176	0.981
I	−2.556	1.043
ɛ	−1.092	1.331
æ	0.677	2.315
ʌ	2.561	0.128
ɑ	3.849	1.356
ɔ	3.471	−0.493
ʊ	1.729	−1.91
o	0.102	−2.685
u	−3.565	−2.065

[a)]Electronic mail: bstory@u.arizona.edu

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.*

Preliminary Notice: 157th Meeting of the Acoustical Society of America

The 157th Meeting of the Acoustical Society of America will be held Monday through Friday, 18–22 May 2009 at the Portland Hilton Hotel & Executive Tower in Portland, Oregon. Information about the meeting also appears on the ASA Home Page at ⟨http://asa.aip.org/meetings.html⟩.

Charles E. Schmid
Executive Director

Technical Program

The technical program will consist of lecture and poster sessions. Technical sessions will be scheduled Monday through Friday, 18–22 May 2009.

Special Sessions

Acoustical Oceanography (AO)

Environmental inferences in inhomogeneous ocean environments
(Joint with Underwater Acoustics)
Methods that allow for inferences of probability distributions for values of environmental parameters in ocean waveguides that possess strong spatial and temporal variability
Temporal and spatial field coherence applied to ocean sensing: Measurement, theory and modeling
(Joint with Underwater Acoustics and Signal Processing in Acoustics)
Coherence scale observations, modeling, theoretical predictions, and signal processing uses/impacts

Animal Bioacoustics (AB)

An integration of bioacoustics, neuronal responses, and behavior
Integration of auditory responses and characteristics with neuronal responses and with animal behavior
Autonomous remote monitoring systems for marine animals
(Joint with Acoustical Oceanography)
Use of autonomous remote sensing systems for assessing marine animal populations
Effects of noise on terrestrial animals
(Joint with ASA Committee on Standards)
Effects of anthropogenic noise on the behavior and distribution of terrestrial animals
Fish bioacoustics: Sensory biology, sound production, and behavior of acoustic communication in fishes
Sensory biology, sound production, and behavior of acoustic communication in fishes
Signal processing techniques for subtle or complex acoustic features of animal calls
(Joint with Signal Processing in Acoustics)
Signal processing techniques to quantify acoustic features of animal calls that historically have been difficult to extract, such as the identity of the caller, call variants with "meta-signaling" information on emotional state, or reference to predator type

Architectural Acoustics (AA)

Acoustics of green buildings: A 360° panel discussion
(Joint with Noise)
Panel discussion from green building experts (architects, engineers, owners) covering acoustical issues
Acoustics of health and healing environments
Architectural acoustics, music, sound quality all have an impact on building IEQ and on occupant health and healing with both conscious and subconscious actions on human comfort and performance
Acoustics of mechanical engineering in multifamily buildings
(Joint with Noise)
Acoustical challenges of designing multifamily mechanical systems
Acoustics of mixed use buildings
(Joint with Noise and ASA Committee on Standards)
Acoustics of buildings of mixed use such as retail/commercial with office and/or residential
Computer auralization
Current state of the art in computer model simulation and auralization.
Indoor noise criteria
(Joint with ASA Committee on Standards and Noise)
Current state of knowledge on indoor noise criteria, from research to case studies
Measurements and modeling of scattering effects
Methodologies and results involving experimental measurement, numerical prediction, and computational modeling of scattering effects in architectural acoustics
Multiple channel systems in room acoustics
Microphone and loudspeaker arrays for room acoustics

Biomedical Ultrasound/Bioresponse to Vibration (BB)

Biomedical applications of acoustic radiation force
Topics on theory and applications of acoustic radiation force in biomedicine, including imaging, tissue characterization, therapy, and material manipulation
Biomedical applications of standing waves
Use of, or complications caused by, standing waves in biomedical therapy or imaging
Cardiovascular applications of ultrasound contrast agents
Use of ultrasound contrast agents for the diagnosis and therapy of cardiovascular disease; especially the treatment of stroke and thrombosis, diagnosis of inflammation and vulnerable plaque, high frequency ultrasound (IVUS) and drug delivery
Image enhancement and targeted drug and gene delivery
Research to guide and direct localized or targeted drug and gene delivery with ultrasound
Metrology and calibration of high intensity focused ultrasound
(Joint with Physical Acoustics and ASA Committee on Standards)
Characterzing the acoustic output, cavitation, and heating produced by HIFU medical devices
Shock wave therapy
(Joint with Physical Acoustics)
Research on shock waves for medical therapy

Engineering Acoustics (EA)

Acoustic engineering of wind turbines
(For Topical Meeting on Wind Turbines)
Acoustical engineering design, development, and evaluation of wind turbines
Lasers in underwater acoustics
(Joint with Physical Acoustics)
Application of lasers in underwater acoustics, for example, for the generation of sound through the optoacoustic effect, reception of underwater sound, measurement of acoustic properties, precision alignment of arrays, verification of transducer performance, or other unique applications

 0001-4966/2009/125(2)/1249/6/$25.00

Piezoelectric energy harvesting
Past, present and future of energy harvesting through piezoelectric materials and devices

Education in Acoustics (ED)

Hands-on experiments for high school students
Experiments for high school students
"Project Listen Up"
(Joint with ASA Student Council)
Descriptions of acoustics demonstrations, laboratory experiments or discovery activities for learners of all ages. Apparatus may be shown, but the talks should focus on concepts, explanations, diagrams and drawings with an emphasis on careful scientific approach

Musical Acoustics (MU)

Microphone array techniques in musical acoustics
(Joint with Engineering Acoustics)
Methods and applications of microphone arrays measuring sound fields and vibrations of musical instruments
Wind instruments
Research on the acoustics of wind instruments including model analysis and nonlinear propagation and on the history of wind instruments. Includes a performance by Edinburgh Renaissance Band

Noise (NS)

Bioacoustic metrics and the impact of noise on natural environment
(Joint with Animal Bioacoustics and ASA Committee on Standards)
Current research on the impacts of noise on the natural environment
Hospital noise and health care facilities
(Joint with Architectural Acoustics and ASA Committee on Standards)
Assessment and measurement procedures
Noise/case study of quarry noise and aggregates
Research and results, new developments
Noise litigation
State of the art, efforts, and experiences
Roof design to limit rain noise
Techniques, practices, procedures, and applications
Wilderness and park soundscapes
(Joint with Architectural Acoustics and ASA Committee on Standards)
Measurement, assessment, and protection of the natural and non-natural soundscape
Workshop–Soundscape and community noise
(Joint with Architectural Acoustics, ASA Committee on Standards and Biomedical Ultrasound/Bioresponse to Vibration)
Soundscape and its applications for city planning

Physical Acoustics (PA)

A half-century with the parametric acoustic array
(Joint with Underwater Acoustics and Engineering Acoustics)
The parametric acoustic array is celebrated through this review of theoretical developments, applications, and instruments inspired by P. J. Westervelt's discovery
Influence of temperature on sound in condensed matter
Elastic moduli of solids are fundamental thermodynamic variables that reveal much about the fundamental science. The session will focus on the temperature dependence of the elastic moduli of solids as a probe of fundamental physics
Numerical methods for weak shock propagation
(Joint with Biomedical Ultrasound/Bioresponse to Vibration)
Development, testing, and application of computational techniques for simulating the propagation of finite amplitude waves and weak shocks, including atmospheric, geophysical, and biomedical wave propagation

Psychological and Physiological Acoustics (PP)

Neuroimaging of human spatial hearing
Presentation of neuroimaging data for spatial hearing
Theory construction in the domain of auditory perception
Cooperative effort among auditory scientists who have developed useful theories to teach the process to younger scientists

Signal Processing in Acoustics (SP)

Detection and classification of underwater targets
(Joint with Underwater Acoustics)
Signal processing for detection and classification of underwater targets, particularly in shallow water, as well as buried or partially buried objects
Pattern recognition in acoustic signal processing
Statistical pattern classification/machine learning algorithms, their performance assessment, and their application to the full range of acoustic signal processing problems.
Poroelastic materials: Models, bounds, and parameter estimation
(Joint with Acoustical Oceanography)
Effective medium models for poroelastic materials, bounds on elastic properties, and methods for estimating model parameters. Applications in seismic exploration, underwater acoustics, characterization and design of composite materials, and other fields

Speech Communication (SC)

Articulatory speech synthesis and robotic speech
Articulatory speech and voice synthesis with an emphasis on approaches that make use of mechanical models and analog vocal tracts
Exploring the relationship between cognitive processes and speech perception
(Joint with Psychological and Physiological Acoustics)
Research findings and theoretical questions concerning the effect that various cognitive processes (e.g., memory, attention) have on listeners' ability to process speech and/or spoken language
Source/filter interaction in biological sound production
(Joint with Musical Acoustics and Animal Bioacoustics)
Source-filter interactions that affect vocal fold oscillation on the basis of acoustic loads provided by the airways above and below the sound source (larynx or syrinx). Comparisons between biological and man-made instruments are drawn
Vowel inherent spectral change
Various aspects of vowel inherent spectral change, such as descriptions of dynamic spectral properties in the vowels of particular languages and dialects, theoretical and experimental studies of the perceptually relevant aspects of spectral change, the effects of spectral change on second-language speech learning, and the use of dynamic spectral patterns in forensic phonetics
See ⟨http://www.asa09crosslangspeech.com/⟩ for information about the 2nd Special Workshop on Speech: Cross-language speech perception and variations in linguistic experience

Structural Acoustics and Vibration (SA)

Computational structural acoustics
Computational methods and techniques for sound, vibration and their interactions
Concepts of new vibration sensors
(Joint with Engineering Acoustics)
Development and enhancement of vibration sensors and their implementations
Emerging applications of structural acoustics in new fields
Applications of structural acoustics in various emerging fields, for example, (nuclear) power generation
Vibro-acoustic diagnosis and prognosis of complex structures
Analysis and prediction of sound radiation and structural vibration of complex structures
Wind turbine vibration and sound radiation
(Joint with Engineering Acoustics and Noise)
(For Topical Meeting on Wind Turbines)
Impacts of structural vibrations and sound radiation from wind turbines

Underwater Acoustics (UW)

Monostatic and bistatic detection of elastic objects near boundaries: Methodologies and tradeoffs
(Joint with Structural Acoustics and Vibration)
State of the art in monostatic and bistatic detection of elastic objects
Physics-based undersea clutter model verification and validation
(Joint with Acoustical Oceanography)

Verification through the development of benchmark analytic and numerical clutter solutions. Validation of clutter models using advanced data and signal processor outputs such as clustering and amplitude distributions from normalizer and tracker sonar displays
Session in honor of Ralph Goodman
(Joint with Acoustical Oceanography)
Acoustics of ocean bubbles
Waveguide invariant principles for active and passive sonars
(Joint with Signal Processing in Acoustics)
Recent theoretical development and experimental observations in active and passive sonar of waveguide invarient

Other Technical Events

Hot Topics

A "Hot Topics" session sponsored by the Tutorials Committee will cover the fields of Education in Acoustics, Psychological and Physiological Acoustics and Signal Processing in Acoustics.

Distinguished Lecture

A distinguished lecture titled "A Residual-Potential Boundary for Time-Domain Problems in Computational Acoustics" will be presented by Thomas L. Geers from the University of Colorado.

Workshop on Preparing JASA and JASA Express Letters Articles

This workshop will include presentations concerning the preparation and submittal of papers to JASA. JASA Express Letters, and Proceedings of Meetings on Acoustics.

Workshop on Federal Regulations for Human Subjects Protection

This session will provide an overview of the federal regulations for human subjects protection. Topics will include OHRP, FDA and HIPAA regulations. Regulations regarding informed consent, research with children, research with devices and expedited vs. full board review process will be covered. Guidelines for working with local Institutional Review Boards will also be presented.

Wind Turbine Topical Meeting

A Wind Turbine Topical Meeting, sponsored by Structural Acoustics and Vibration, Engineering Acoustic and Noise, is being organized at the Portland meeting. Sessions will focus on the structural and acoustical design, development, and evaluation of wind turbines for power generation, and on the evaluation, mitigation, and community effects of noise generated by deployed wind turbines.

Workshop on Cross-Language Speech Perception and Variations in Linguistic Experience

This workshop will be held 22–23 May at the World Trade Centre in Portland, OR, which is a short distance from the Portland Hilton. Visit <http://www.asa09crosslangspeech.com/index.html> for full details about the workshop.

This ASA Special Workshop on Speech revolves around the fundamental question of how experience with language systematically shapes perception of even the most basic building blocks of spoken communication—consonants and vowels. The topics will cover current theoretical perspectives and recent research on cross-language speech perception. Specific topics include basic issues and findings, neuropsychological undrepinnings, and language-training applications related to language-tuned speech perception. The populations addressed run the gamut from young first language learners, to bilingual language users and learners, to second-language learners.

An exciting two-day program of invited speakers will be offered, as well as sessions on each day for contributed talks and posters. The keynote address will be given by Winifred Strange, who will describe her newly-developed Automatic Selective Perception (ASP) model of attentional tuning in cross-language speech perception, and recent findings motivated by it.

The ASA and the organizing committee have a strong commitment to training future speech researchers. Toward this goal, we have included several young investigator invited speakers, and have scheduled two contributed poster sessions and two contributed talk sessions.

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 April. Submission procedures and password information will be mailed to authors with the acceptance notices.

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

The URL is <http://scitation.aip.org/asameetingpapers>.

Proceedings of Meetings on Acoustics (POMA)

The upcoming meeting of the Acoustical Society of America will have a published proceedings, and submission is optional. The proceedings will be a separate volume of the online journal, "Proceedings of Meetings on Acoustics" (POMA). This is an open access journal, so that its articles are available in pdf format without charge to anyone in the world for downloading. Authors who are scheduled to present papers at the meeting are encouraged to prepare a suitable version in pdf format that will appear in POMA. The format requirements for POMA are somewhat more stringent than for posting on the ASA Online Meetings Papers Site, but the two versions could be the same. The posting at the Online Meetings Papers site, however, is not archival, and posted papers will be taken down six months after the meeting. The POMA online site for submission of papers from the meeting will be opened at the same time when authors are notified that their papers have been accepted for presentation. It is not necessary to wait until after the meeting to submit one's paper to POMA. Further information regarding POMA can be found at the site http://asa.aip.org/poma.html. Published papers from previous meetings can be seen at the site http://scitation.aip.org/POMA.

Meeting Program

A complete meeting program will be mailed as Part 2 of the April issue of JASA. Abstracts will be available on the ASA Home Page <http://asa.aip.org> in April.

Tutorial Lecture on the Art and Science of Unique Musical Instruments

A tutorial presentation on "The Art and Science of Unique Musical Instruments" will be given by Ela Lamblin on Monday, 18 May, at 7:00 p.m.

Ela Lamblin has created many unique musical instruments and sound sculptures in the Northwest. Together with his partner and wife, Leah Mann, he founded the performance company Lelavision. Ela and Leah will demonstrate a variety of instruments, and with the help of ASA's acousticians, explain the sounds they produce—some of which are not what we usually hear from traditional instruments. For a preview visit www.lelavision.com/

The presentation and performance will be held in the modern Newmark Theater (www.pcpa.com/events/newmark.php), located two blocks from the Hilton Hotel in Antoinette Hatfield Hall, 1111 SW Broadway (at Main Street).

There is no fee for this tutorial which is open to the public as part of ASA's outreach program. However, attendees are asked to register to attend the presentation. Register online at <http://asa.aip.org/> or use the form in the printed call for papers.

Short Course on Outdoor Noise Estimation and Mapping

Outdoor environmental noise is becoming more prominent in the United States as the population and infrastructure continues to grow. This short course will present the current state of knowledge on predicting outdoor noise, and discuss the standards and uncertainties that exist in this area. Examples from a variety of practical applications will be explored. Additionally, an introduction to noise mapping and soundscapes will be provided. The course objective is to introduce the methodologies, uncertainties, and standards regarding outdoor noise estimation. A number of environmental noise applications, ranging from prediction of noise from wind farms to noise and psychoacoustics mapping will also be presented.

The short course will be taught by a team of instructors who cover a wide range of expertise in outdoor noise estimation and mapping. Ken Kaliski is Director of Environmental Services at Resource Systems Group, White River Junction, VT. Mr. Kaliski is a Professional Engineer and INCE

Board Certified, with experience in community noise mapping and modeling such sources as wind farms, quarries, and highways. Robert Putnam is a Senior Acoustical Engineer at Siemens Energy Systems, Orlando, FL. Mr. Putnam has over 40 years of experience in the design and prediction of power plant noise, currently serves as Chair of standards writing committees within ASTM and ASME, and has previously conducted tutorials on environmental noise for ASA and INCE. Dr. Brigitte Schulte-Fortkamp is a Professor at the Technical University of Berlin, Germany, and Dr.-Ing. Klaus Genuit is with HEAD Acoustics in Germany, Both Dr. Schulte-Fortkamp and Dr. Genuit are Fellows of the ASA who have extensive experience on community noise evaluation, sound quality, and soundscapes.

The course schedule is Sunday, 17 May 2009, 1:00 to 5:00 p.m. and Monday, 18 May 2009, 8:30 a.m. to 12:30 p.m.

The registration fee is $250.00 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 27 April will be guaranteed receipt of instruction materials. There will be a $50.00 USD discount for registration made prior to 27 April. Full refunds will be made for cancellations prior to 27 April. Any cancellations after 27 April will be charged a $25.00 USD processing fee. Register online at <http://asa.aip.org> or use the form in the printed call for papers.

ASA Meeting Goes Green

Portland is the most sustainable city in the country. We take pride in it. Green is not only our spirit but also our way of life. At this year's ASA meeting in Portland, many sustainable green elements have been implemented, including picking a green seal certified venue, Hilton Portland & Executive Tower; recycling and reusing meeting materials; and following the guideline of the Green Conference Initiative from the U.S. Environmental Protection Agency.

In an effort to offset the carbon footprint of the meeting in Portland, the ASA will contribute $2000 to a non-profit carbon offset provider which will compensate for some of the carbon emissions generated for the meeting. The local organizing committee will select the organization based on applicability of project types (energy, renewables, etc.) and alignment with ASA tenets. Attendees can contribute to minimizing their own carbon footprint by car pooling when possible, take a train instead of driving, and use the light rail from the airport to the hotel.

Special Meeting Features

Student Transportation Subsidies

A student transportation subsidies fund has been established to provide limited funds to students to partially defray transportation expenses to meetings. Students presenting papers who propose to travel in groups using economical ground transportation will be given first priority to receive subsidies, although these conditions are not mandatory. No reimbursement is intended for the cost of food or housing. The amount granted each student depends on the number of requests received. To apply for a subsidy, submit a proposal (e-mail preferred) to be received by 8 April to: Jolene Ehl, ASA, Suite 1NO1, 2 Huntington Quadrangle, Melville, NY 11747-4502, Tel: 516-576-2359, Fax: 516-576-2377, E-mail: jehl@aip.org. The proposal should include your status as a student; whether you have submitted an abstract; whether you are a member of ASA; method of travel; if traveling by auto; whether you will travel alone or with other students; names of those traveling with you; and approximate cost of transportation.

Young Investigator Travel Grant

The Committee on Women in Acoustics (WIA) is sponsoring a Young Investigator Travel Grant to help with travel costs associated with presenting a paper at the Portland 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 Portland meeting, are not currently students, and have not previously received the award. Each award will be of the order of $400 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 to Jennifer Miksis-Olds at ⟨jlm91@psu.edu⟩. Deadline for receipt of applications is 15 April.

Students Meet Members for Lunch

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

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

Buffet socials with cash bar will be held on Tuesday and Thursday evenings beginning at 6:00 p.m.

The ASA Plenary session will be held on Wednesday afternoon, 20 May, at the Hilton Portland where Society awards will be presented and recognition of newly-elected Fellows will be announced.

A Fellows Luncheon will be held on Thursday, 21 May, at 12:00 noon at the Portland Hilton. This luncheon is open to all attendees and their guests. Register online at ⟨http://asa.aip.org⟩ or use the form in the printed call for papers.

Women in Acoustics Luncheon

The Women in Acoustics luncheon will be held on Wednesday, 20 May. Those who wish to attend this luncheon must register online at ⟨http://asa.aip.org⟩ or use the form in the printed call for papers. The fee is $15 (students $5) for pre-registration by 27 April and $20 (students $5) at the meeting.

Transportation and Hotel Accommodations

Air Transportation

The Portland International Airport, (Airport Code PDX) is served by the following airlines: Air Canada, Alaska Airlines, American Airlines, Continental Airlines, Delta Air Lines, Frontier, Hawaiian Airlines, Horizon Air, JetBlue Airways, Lufthansa, Southwest Airlines, Northwest Airlines, United Airlines, and US Airways. For further information see ⟨http://www.flypdx.com/PDX_home.aspx⟩.

Ground Transportation

The Hilton Portland & Executive Tower is located approximately 12.5 miles from Portland International Airport.

Public Transportation: The Hilton is easily accessible from the airport using the Portland MAX Light Rail. The Red Line from the airport stops one block from the Hilton; disembark at the Pioneer Square stop. Information on the MAX can be found at ⟨http://trimet.org/max/⟩.

Taxicabs: A taxi from Portland International Airport to the Hilton costs approximately $35.00.

Automobile Rental: Portland International Airport is served by all major car rental companies with many companies providing on-site rental opportunities. See ⟨http://www.flypdx.com/Grnd_Trans.aspx⟩ for more information.

For a map of the hotel location visit ⟨http://www1.hilton.com/en_US/hi/hotel/PDXPHHH-Hilton-Portland-Executive-Tower-Oregon/directions.do#localmap⟩.

The Hilton offers self-parking with in/out privileges for $18/day. Valet parking is also available for $27.00 per day.

Room Sharing

ASA will compile a list of those who wish to share an hotel room and its cost. To be listed, send your name, telephone number, e-mail address, gender, smoker or nonsmoker preference, not later than 1 April 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

Weather in May is perfect for enjoying the emerging blossoms in the City of Roses. Portland temperatures average between about 47 to 68 degrees (F) in May with an average rainfall of 2.06 inches. The majority of days, on average, range from clear to partly cloudy/cloudy. Recent observations and forecasts may be found on a number of different web pages (e.g., www.weather.com).

Hotel Reservation Information

Hotel Accommodations

The meeting will be held at the Portland Hilton & Executive Tower ⟨http://www.tourhiltonportland.com⟩ in Portland Oregon. Located in the heart of Downtown Portland—just one block from Portland's 'Living Room' Pioneer Courthouse Square—the Hilton is also located only a block away from the MAX Light Rail system with easy access to the city's many attractions.

A block of guest rooms at discounted rates has been reserved for meeting participants at the Hilton Portland & Executive Tower. **Early reservations are strongly recommended**. Note that the special ASA meeting rates are not guaranteed after **20 April 2009**. You must mention the Acoustical Society of America when making your reservations to obtain the special ASA meeting rates.

Please make your reservation directly with the Hilton Portland & Executive Tower. When making your reservation, you must mention the Acoustical Society of America to obtain the special ASA rates. Alternatively, reservations can be made directly online at the website listed below, which has been set up specifically for the Acoustical Society of America, and incorporates the conference rates and all applicable information.

Hilton Portland & Executive Tower
921 SW Sixth Avenue
Portland, Oregon 97204
Tel: 1-503-226-1611
Fax: 1-503-220-2565
http://www.tourhiltonportland.com
Online: http://www.hilton.com/en/hi/groups/personalized/PDXPHHH-ASA-20090513/index.jhtml
ROOM RATE: Single/Double: $162.00 USD plus 12.5% tax

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

The Women in Acoustics Committee of the ASA is helping to organize on-site child care for the Portland meeting. Professional nannies will be available to provide child care in a group setting at a rate that depends on the numbers and ages of the children. Priority will be given to members who sign up for the service in advance as coverage will be determined from a pre-conference estimate. Members interested in child care services should contact Dawn Konrad-Martin at dawn.martin@va.gov as early as possible to indicate interest.

Accompanying Persons Program

Spouses and other visitors are welcome at the Portland meeting. The registration fee is $50.00 for preregistration by 27 April and $75.00 at the meeting.

A hospitality room for accompanying persons will be open at the Hilton Portland & Executive Tower from 8:00 a.m. to 10:30 a.m. each day throughout the meeting where information about activities in and around Portland will be provided. Accompanying person tours are being considered dependent on sufficient interest and participation. These might include half day trips to the Columbia River Gorge and Willamette Valley wine tasting. Please check the ASA website at ⟨http://asa.aip.org/meetings.html⟩ for updates about the accompanying persons program.

Portland and its metro area has been nationally recognized as a mecca for visitors seeking both urban adventure and natural beauty. A dynamic arts and culture community, a lively downtown, the proximity of Mt. Hood and the Columbia Gorge, and a dining scene focused on seasonal and locally-grown produce featuring fresh ingredients are keys to this draw. Within minutes of driving from downtown are internationally-lauded Pinot-producing wineries. And, of course, Portland's reputation as a craft-brewing epi-center is well established at this point.

Conference attendee/accompanying person tours are being considered dependent on sufficient interest and participation. These might include half day trips to the Columbia River Gorge and Willamette Valley wine tasting. Please check the ASA website at ⟨http://asa.aip.org/meetings.html⟩ for updates about the accompanying persons program.

Registration Information

The registration desk at the meeting will open on Monday, 18 May, at the Hilton Portland & Executive Tower. Register online at ⟨http://asa.aip.org⟩ or use the form in the printed call for papers. If your registration is not received at the ASA headquarters by 27 April you must register on-site.

Registration fees are as follows:

Category	Preregistration by 27 April	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 2007, 2008, or 2009)	$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 (2009) 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.

A $25 fee will be charged to those who wish to cancel their reigstration after 27 April.

Online registration is available at ⟨http://asa.aip.org⟩.

Cross Language Speech Perception Workshop Registration

Participants may register for the workshop using the registration form in the printed call for papers or online at ⟨http://asa.aip.org⟩. One may register for the workshop only or both the ASA Portland meeting and the workshop in one transaction.

For details about the workshop visit the Workshop web page at ⟨www.asa09crosslangspeech.com⟩.

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.

2009

18–22 May	157th Meeting of the Acoustical Society of America, Portland, OR [Acoustical Society of America, Suite 1NO1, 2 Huntington Quadrangle, Melville, NY 11747-4502; Tel.: 516-576-2360; Fax: 516-576-2377; Email: asa@aip.org; WWW: http://asa.aip.org].
24–28 June	5th International Middle-Ear Mechanics in Research and Otology (MEMRO), Stanford University, Stanford, CA [http://memro2009.stanford.edu].
26–30 Oct	158th Meeting of the Acoustical Society of America, San Antonio, TX [Acoustical Society of America, Suite 1NO1, 2 Huntington Quadrangle, Melville, NY 11747-4502; Tel.: 516-576-2360; Fax: 516-576-2377; Email: asa@aip.org; WWW: http://asa.aip.org].

2010

19–23 April	159th Meeting of the Acoustical Society of America, Baltimore, MD [Acoustical Society of America, Suite 1NO1, 2 Huntington Quadrangle, Melville, NY 11747-4502; Tel.: 516-576-2360; Fax: 516-576-2377; Email: asa@aip.org; WWW: http://asa.aip.org].
15–19 Nov	2nd Iberoamerican Conference on Acoustics (Joint Meeting of the Acoustical Society of America, Mexican Institute of Acoustics, and Iberoamerican Federation on Acoustics), Cancun, Mexico [Acoustical Society of America, Suite 1NO1, 2 Huntington Quadrangle, Melville, NY 11747-4502; Tel.: 516-576-2360; Fax: 516-576-2377; Email: asa@aip.org; WWW: http://asa.aip.org].

ACOUSTICAL NEWS—INTERNATIONAL

Walter G. Mayer
Physics Dept., 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.

March 2009

17–19 **Spring Meeting of the Acoustical Society of Japan**, Tokyo, Japan (www.asj.gr.jp/index-en.html).

23–26 **International Conference on Acoustics (NAG/DAGA 2009)**, Rotterdam, The Netherlands (www.nag-daga.nl).

31–2 **5th International Conference on Bio-Acoustics**, Loughborough, UK (bioacoustics2009.lboro.ac.uk).

April 2009

5–9 **Noise and Vibration: Emerging Methods (NOVEM 2009)**, Oxford, UK (www.isvr.soton.ac.uk/NOVEM2009).

13–17 **2nd International Conference on Shallow Water Acoustics**, Shanghai, China ([soon]: www.apl.washington:edu).

19–24 **International Conference on Acoustics, Speech, and Signal Processing**, Taipei. R.O.C. (icassp09.com).

June 2009

2–5 **XXI Session of the Russian Acoustical Society**, Moscow, Russia (www.akin.ru/main.htm).

17–19 **3rd International Conference on Wind Turbine Noise**, Aalborg, Denmark (www.windturbinenoise2009.org).

21–25 ***13th International Conference "Speech and Computer,"** St. Petersburg, Russia (www.specom.nw.ru).

22-26 **3rd International Conference on Underwater Acoustic Measurements: Technologies and Results**, Nafplion, Peleponnese, Greece (www.uam2009.gr).

July 2009

5–9 **16th International Congress on Sound and Vibration**, Krakow, Poland (www.icsv16.org).

August 2009

12–16 ***7th Triennual Conference of the European Society for Cognitive Science of Music (ESCOM2009)**, Jyväskylä, Finland (www.fyu.fi/hum/laitokset/musikki/en/escom2009).

23–28 **Inter-noise 2009**, Ottawa, Ont., Canada (www.intemoise2009.com).

September 2009

6–10 **InterSpeech 2009**, Brighton, UK (www.interspeech2009.org).

9–11 **9th International Conference on Theoretical and Computational Acoustics**, Dresden, Germany (ictca2009.com).

14–18 **5th Animal Sonar Symposium**, Kyoto, Japan (cse.fra.affrc.go.jp/akamatsu/AimalSonar.html).

15–17 **Autumn Meeting of the Acoustical Society of Japan**, Koriyama, Japan (www.asj.gr.jp/index-en.html).

19–23 **IEEE 2009 Ultrasonics Symposium**, Rome, Italy (e-mail: pappalar@uniroma3.it).

21–23 **10th Western Pacific Acoustics Conference (WESPAC)**, Beijing, China (www.wespax.org).

23–25 **Pacific Rim Underwater Acoustics Conference (PRUAC)**, Xi' an, China (e-mail: lfh@mail.ioa.ac.cn).

23–25 **TECNIACUSTICA2010**, Cádiz, Spain (www.sea-acustica.es).

October 2009

5–7 ***International Conference on Complexity of Nonlinear Waves**, Tallinn, Estonia (www.ioc.ee/cnw09).

26–28 **Euronoise 2009**, Edinburgh, UK (vwvw.euionoise2009.org.uk).

June 2010

9–11 ***14th Conference on Low Frequency Noise and Vibration**, Aalborg, Denmark.

13–16 **INTERNOISE2010**, Lisbon, Portugal, (www.intemoise2010.oig).

August 2010

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

September 2010

26–30 **Interspeech 2010**, Makuhari, Japan (www.interspeech2010.org).

June 2011

27–1 *Forum **Acusticum 2011**, Aalborg, Denmark.

August 2011

27–31 **Interspeech 2011**, Florence, Italy (www.interspeech2011.org).

September 2011

4–7 **International Congress on Ultrasonics**, Gdansk, Poland.

June 2013

2–7 **21stt International Congress on Acoustics (ICA2013)**, Montréal, Canada. (www.ica2013montreal.org).

 0001-4966/2009/125(2)/1255/1/$25.00

BOOK REVIEW

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.

Self-Consistent Methods for Composites, Volume 2—Wave Propagation in Heterogeneous Materials

S. K. Kanaun and V. M. Levin

Springer, The Netherlands, 2008. pp. 294. Price: $189.00. ISBN: 978-1-4020-6967-3.

This book is the second of two volumes on the topic of self-consistent methods in composites. The first volume covered quasistatic problems in elasticity, thermal and electric fields, as well as some interactions between thermal/electric fields and elastic fields in composites. The second volume is devoted to wave propagation problems in composites and other types of heterogeneous materials. The authors have tried to make the two volumes as independent as possible, which leads to some overlaps—such as the two Appendices A and B here, which are essentially the same as Appendices A and E from Volume 1.

The main ideas revolve around two approaches to averaged equations for composites, or random media. They stress the differences between effective field approximations and effective medium approximations. An effective field method (EFM) treats inclusions in heterogeneous media as if isolated in the original matrix material, while the effective field at the inclusion is a sum of external applied field and perturbations due to surrounding inclusions. The effective medium method (EMM) treats each inclusion as isolated in the overall composite medium with the pertinent applied field being the actual applied field. The emphasis of Volume 2 is on wave propagation applications. In addition to the authors' own publications, the references here to well-known works from the physics literature include Bruggeman, Foldy, Lax, Waterman, and Truell, as well as more recent work on coherent potential approximations by Gubernatis, Krumhansl, Stroud, and others. The references to the mechanics literature seem to concentrate on works of Hill, Willis, and co-workers. References to acoustics literature include Kino, Mal, Twersky, and the Varadans.

Subjects covered in the book include: scalar waves, electromagnetic waves, axial shear waves, scattering of long elastic waves, effective wave operators for elastic waves in random media, extended discussion of elastic waves in elastic media spherical inclusions, and elastic waves in polycrystals. Of eight chapters in the book, five are devoted to elastic wave propagation, one to scalar/acoustic waves, one to electromagnetics, and one chapter of Introduction. The overall emphasis of the book is clearly on acoustics and elasticity. The main results of the book are contained in Chapter 7, where detailed comparisons are made between the EFM approach and three different versions of the EMM approach. The final conclusion is that the EFM approach is harder to use, but gives more reliable results, especially when careful comparisons to data are considered—as they are here.

The strangest omissions, at least to the eyes of this reviewer, include Eshelby's work on the mechanics of composites containing isolated ellipsoids, and Keller's work on waves in random media that can be characterized using spatial correlation functions. Taking one particular example, Section 6.1 concerns scattering of elastic waves by random ellipsoidal inclusions, and explicit assumptions are made about the constancy of displacement and strain fields inside the (assumed to be) isolated inclusions for long waves. Invoking Eshelby's famous results at this point in the argument would have been very appropriate and would have increased the overall credibility of the approximations made in this and the following derivations. On the other hand, Keller's work is perhaps omitted because it does not fall neatly into the class of approximations the authors choose to study here. But I think it would have served readers well to clarify the differences between those methods emphasized and others not emphasized to have been somewhat less single-minded in the overall outline of the subjects covered.

The text is appropriate for general background in a graduate level course in wave physics (both acoustics and electromagnetics) in heterogeneous media. This second volume of the series will also be of interest to researchers in the general area of composites, especially for the acoustics/seismology community. The book collects in one place many results of the authors that may be difficult to obtain from the original sources. Although I found disconcerting the seeming disconnection from some of the waves literature that is better known to me presonally, I also found the book particularly useful from the point of view of having a set of independent derivations and proofs of the various results presented. I consider this book to be a positive contribution to our literature overall, and one that I think readers will find useful for both teaching and research purposes.

JAMES G. BERRYMAN
Earth Sciences Division, Geophysics Department,
Lawrence Berkeley National Laboratory,
1 Cyclotron Road, MS 90R1116,
Berkeley, CA 94720

 0001-4966/2009/125(2)/1256/1/$25.00

REVIEWS OF ACOUSTICAL PATENTS

Sean A. Fulop
Dept. of Linguistics, PB92
California State University Fresno
5245 N. Backer Ave., Fresno, California 93740

Lloyd Rice
11222 Flatiron Drive, Lafayette, Colorado 80026

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

Reviewers for this issue:

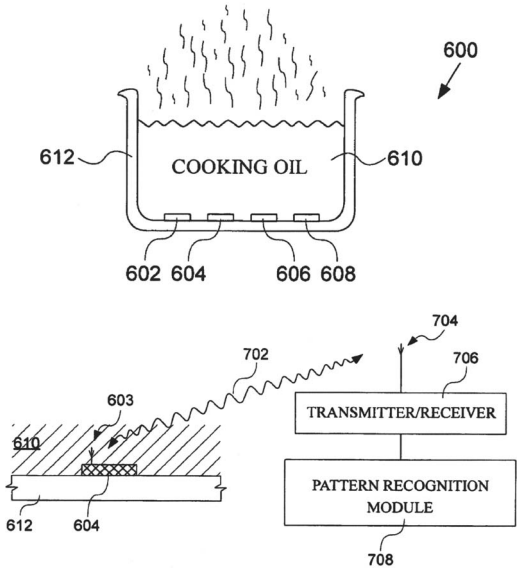
GEORGE L. AUGSPURGER, *Perception, Incorporated, Box 39536, Los Angeles, California 90039*
ANGELO CAMPANELLA, *3201 Ridgewood Drive, Hilliard, Ohio 43026-2453*
JEROME A. HELFFRICH, *Southwest Research Institute, San Antonio, Texas 78228*
DAVID PREVES, *Starkey Laboratories, 6600 Washington Ave. S., Eden Prairie, Minnesota 55344*
CARL J. ROSENBERG, *Acentech Incorporated, 33 Moulton Street, Cambridge, Massachusetts 02138*
NEIL A. SHAW, *Menlo Scientific Acoustics, Inc., Post Office Box 1610, Topanga, California 90290*
ERIC E. UNGAR, *Acentech, Incorporated, 33 Moulton Street, Cambridge, Massachusetts 02138*
ROBERT C. WAAG, *Department of Electrical and Computer Engineering, University of Rochester, Rochester, New York 14627*

7,383,731

43.20.Ye DEEP-FRY OIL QUALITY SENSOR

James Z. T. Liu *et al.*, assignors to Honeywell International Incorporated
10 June 2008 (Class 73/602); filed 2 June 2005

As deep fry cooking oil is heated over a prolonged period of time, the oil **610** can be "converted to free fatty acids and other unhealthy compounds," as well as becoming rancid. The patent describes a system that uses acoustic wave sensors **602**, **604**, **60**, **608** in fry vessel **612** as well as oil condition measurement equipment comprised of antenna **704**, transceiver unit **706**, and pattern recognition module **708**, to determine when it's time for an oil change. The sensors are interrogated and the return signal is compared to a data base in the pattern recognition module. Questions which



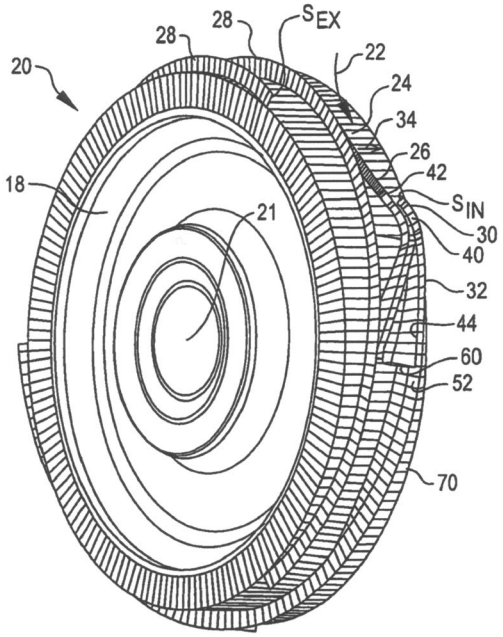
come to mind are: How long do the sensors last in the hot liquid environment? How many batches of fries are needed to populate the pattern recognition module?—NAS

7,334,990

43.25.Ts SUPERSONIC COMPRESSOR

Shawn P. Lawlor *et al.*, assignors to Ramgen Power Systems, Incorporated
26 February 2008 (Class 416/20 R); filed 28 March 2005

An efficient supersonic gas compressor is claimed that requires few parts and has low losses. Rotor **20** is driven at a high RPM (facing edge moving upward) rotating in a close fitting drum (not shown) and has circumferential spiral walls **28** and **70**. Uncompressed gas is to the right. Gas **22** enters narrowed inlet **Sin**. High rotor RPM will cause a shock wave to form at ramp **26–42** behind which compressed gas flows into **Sin** and **30**.



This slower gas flow expands down diffuser **32–44–60** slowing further to move along spiral path **70** to discharge to a reservoir to the left.—AJC

 0001-4966/2009/125(2)/1257/11/$25.00

7,421,899

43.38.Bs RESONANCE METHOD FOR DETERMINING THE SPRING CONSTANT OF SCANNING PROBE MICROSCOPE CANTILEVERS USING MEMS ACTUATORS

Richard K. Workman and Storrs T. Hoen, assignors to Agilent Technologies, Incorporated
9 September 2008 (Class 73/579); filed 17 July 2006

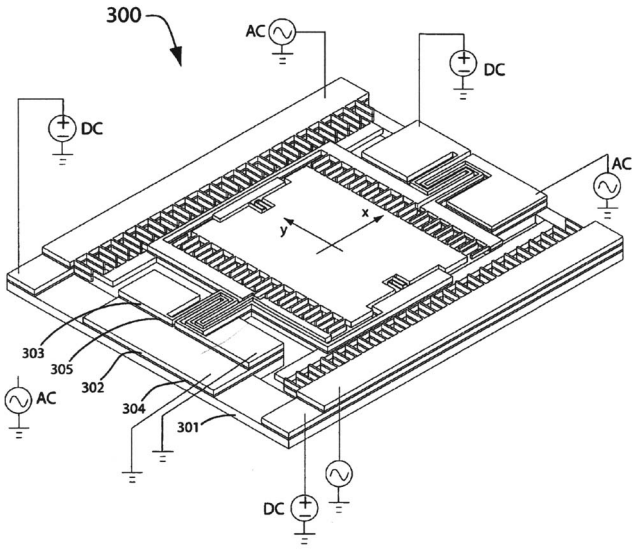
The authors disclose the use of a micro-electronic mechanical system based comb actuator or linear motor to measure the stiffness of an atomic force microscope cantilever. The microscope has to be modified to accept a different mounting arrangement for the cantilevers, but once that is done, the capability exists to shake the cantilever with the tip either touching the surface or not. The authors propose to use this capability to make measurements of the resonant frequency of said cantilever in both states, and use a model that treats the surface as rigid to the tip contact, enabling them to solve for the unknown cantilever stiffness from the resonance frequencies as inputs. This eliminates the nagging problem of determining cantilever stiffness, a recurrent problem with atomic force microscopy.—JAH

7,426,066

43.38.Bs MEMS SCANNING MIRROR WITH TUNABLE NATURAL FREQUENCY

Yee-Chung Fu and Ting-Tung Kuo, assignors to Advanced NuMicro Systems, Incorporated
16 September 2008 (Class 359/199); filed 14 November 2005

This patent discloses the ideas involved in a concept for a tilt actuator that is electrostatically driven, whose resonance frequency can be adjusted by using a variable electric bias voltage. While this is not exactly new, the approach here is exceptionally simple and flexible enough to be applied to many other macro-scale problems. The device described here is used for tilting a scanning mirror, where single-frequency operation is quite common



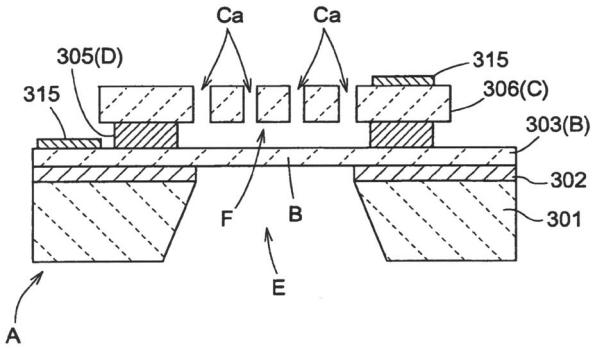
and tuning of the resonance is valuable. This will not be of interest to those wanting to make arrays, as the actuators are a large fraction of the total device area.—JAH

7,386,136

43.38.Gy SOUND DETECTING MECHANISM

Yoshiaki Ohbayashi *et al.*, assignors to Hosiden Corporation
10 June 2008 (Class 381/174); filed in Japan 27 May 2003

A means of manufacturing a single silicon crystal on an insulator condenser microphone is described in clear and informative prose. A silicon oxide or silicon nitride layer is formed on the substrate that functions as a



A: support substrate	301: monocrystal silicon substrate
B: diaphragm	302: silicon oxide film
C: back electrode	303: polycrystal silicon film
D: spacer	305: sacrificial layer
E: acoustic opening	306: polycrystal silicon film
F: void area	315: take-out electrode

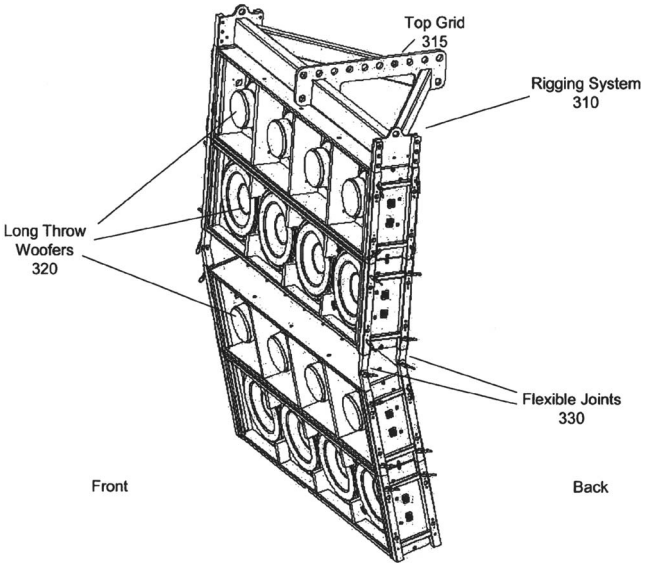
built-in stop layer during one of the fabrication etching steps to control the thickness and stress in the diaphragm.—NAS

7,415,124

43.38.Ja LOW FREQUENCY SURFACE ARRAY

Dragoslav Colich, assignor to HPV Technologies LLC
19 August 2008 (Class 381/335); filed 11 March 2005

About 50 years ago, the "Sweet Sixteen" was a popular project for amateur loudspeaker builders. Sixteen loudspeakers in a 4×4 array were mounted on an open, flat baffle roughly three feet square. In one version, eight speakers faced forward and the others faced backward to reduce distortion. That is the concept patented here, with a few more bells and whistles such as curving the array "…to direct sound at different angles along a



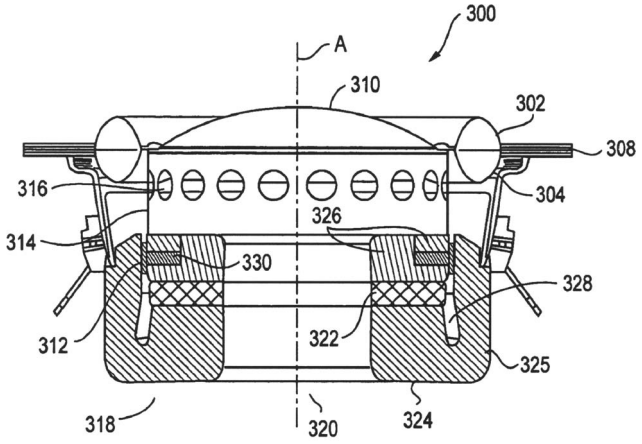
curved surface when a size of the loudspeaker surface array is similar to a size of a wavelength of a sound produced at a predetermined frequency."—GLA

7,418,107

43.38.Ja LOUDSPEAKER

Gilles Milot and Francois Malbos, assignors to Harman Becker Automotive Systems GmbH
26 August 2008 (Class 381/398); filed in the European Patent Office 4 June 2003

The outer suspension of a loudspeaker cone often takes the form of a flexible half-roll. In theory at least, a symmetrical full-roll **302**, **304** is more linear. However, air trapped inside expands and contracts with temperature changes. Also, at long excursions the air is compressed and adds stiffness.



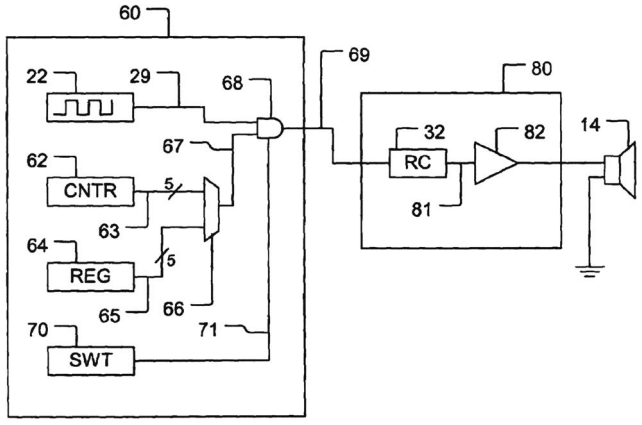
The design described in this patent provides an air leak by making either the upper or lower portion permeable (e.g., using small perforations).—GLA

7,424,120

43.38.Lc METHOD AND APPARATUS FOR VOLUME CONTROL

Douglas Gene Keithley, assignor to Avago Technologies General IP Pte Limited
9 September 2008 (Class 381/107); filed 6 August 2003

Many household appliances emit beeps in response to touchpad entries or to certain operating conditions. Square-wave beep signals are generated by a control module and then reproduced by a separate amplifier/speaker module. Adding a volume control requires additional connections between



the two modules. This patent suggests that pulse width modulation can accomplish the same function without any control wiring.—GLA

7,414,920

43.38.Pf ACOUSTIC GENERATOR FOR DISTANCE SOUNDING WITH MICROPHONE DESIGNED FOR EFFICIENT ECHO DETECTION

Walter Franklin Guion and Steven Lee Herbruck, assignors to WellSonic LC
19 August 2008 (Class 367/144); filed 8 April 2005

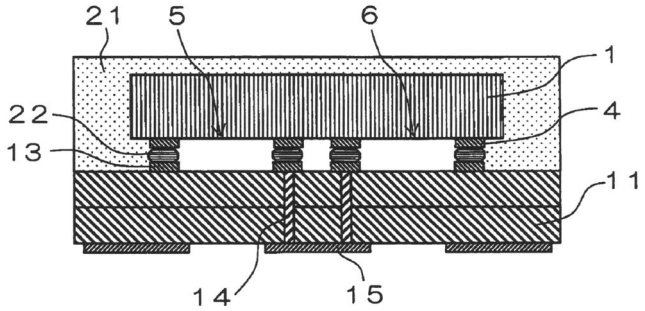
This patent discloses the details of a pulse generator for oil well logging that runs off of a compressed gas supply. The device is capable of producing either positive or negative pressure pulses and can be adjusted for pulse amplitude. It appears to be a real, working device generating frequencies up to 70 Hz at intensities of 160–170 dB SPL. The design is quite ingenious and is shown in detail in the accompanying drawings.—JAH

7,332,986

43.38.Rh SURFACE ACOUSTIC WAVE APPARATUS AND COMMUNICATIONS EQUIPMENT

Wataru Koga *et al.*, assignors to Kyocera Corporation
19 February 2008 (Class 333/193); filed in Japan 28 June 2004

A high power surface acoustic wave (SAW) microwave duplexer is claimed where heat radiating surface **15** is connected by heat and electric



conductors **14** to SAW substrate supports **13-22-4**. **22** is an insulating support resin. **21** is a sealing resin.—AJC

7,339,304

43.38.Rh SURFACE ACOUSTIC WAVE DEVICE

Michio Kadota *et al.*, assignors to Murata Manufacturing Company, Limited
4 March 2008 (Class 310/313 A); filed in Japan 3 October 2003

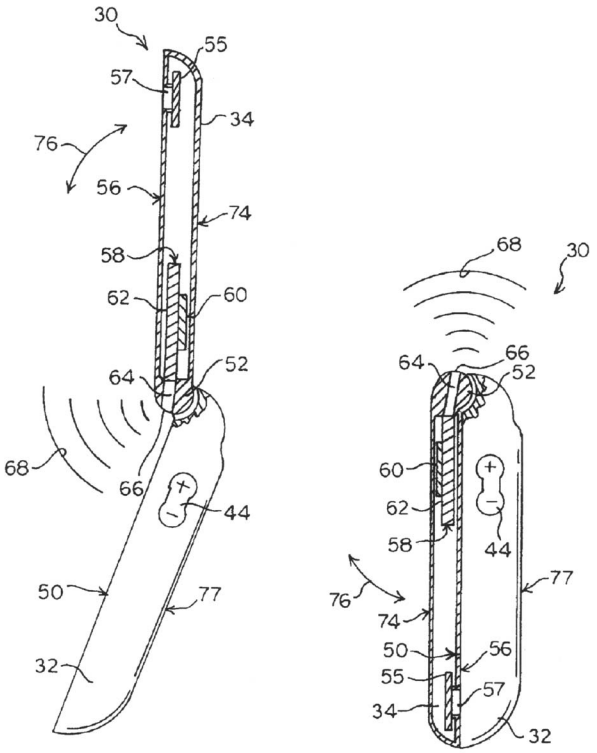
Authors claim and tabulate a variety of Euler angle ranges for the lithium niobate piezoelectric material.—AJC

7,415,290

43.38.Si MOBILE TERMINAL WITH LOUDSPEAKER SOUND REDIRECTION

Mathew J. Murray and William Chris Eaton, assignors to Sony Ericsson Mobile Communications AB
19 August 2008 (Class 455/569.1); filed 14 December 2004

At present, at least three patents per month are devoted to the problem of hearing a cellular phone whether the case is open or closed. Or, as this patent puts it, "Accordingly, there is a need for a mobile handset configuration that provides appropriately adjusted acoustic levels and frequency characteristics at the handset's earpiece via the use of a traditional acoustic receiver unit, provides polyphonic and/or loudspeaker functionality using a separate, remotely mounted loudspeaker, and directs the audio alert and



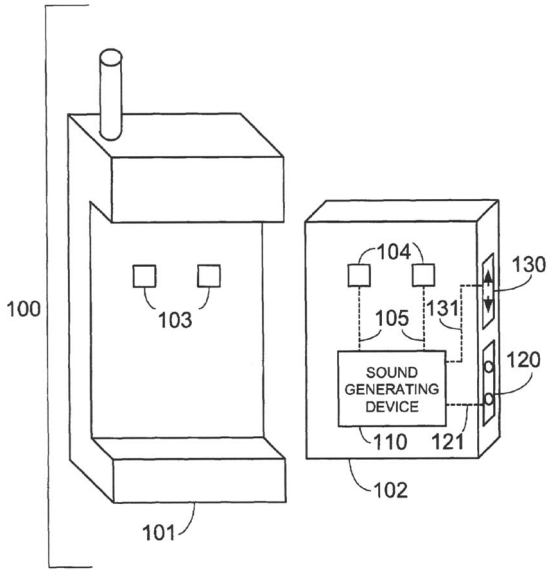
loudspeaker acoustic output generally toward a user in order to provide clear audio whether in the open or closed position." The proposed solution is an end-firing sound exit that automatically guides sound in the desired direction, as shown.—GLA

7,415,291

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

Mark Kirkpatrick, assignor to AT&T Delaware Intellectual Property, Incorporated
19 August 2008 (Class 455/572); filed 28 September 2001

As might be expected from AT&T, more than half of this patent is devoted to its elaborately detailed 21 claims. The basic idea is to incorporate a digital sound library of ring tones as part of a cellular phone battery pack



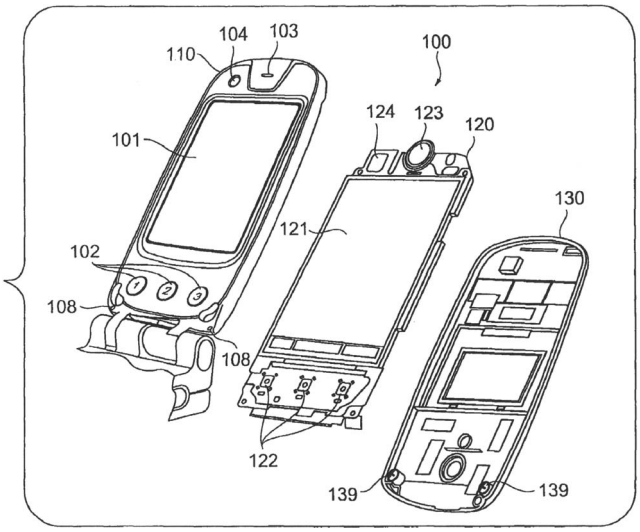
rather than the phone itself. With, say, six battery packs sitting in a charger, the possibilities for self expression become almost endless.—GLA

7,416,048

43.38.Si PORTABLE TERMINAL APPARATUS

Manabu Hongo *et al.*, assignors to Fujitsu Limited
26 August 2008 (Class 181/199); filed in Japan 18 August 2004

This patent describes yet another scheme to make a cellular phone audible whether the case is open or closed. Not only is sound selectively



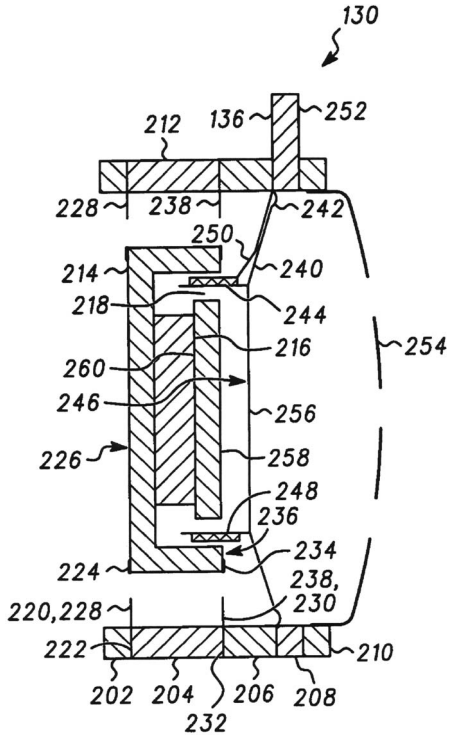
directed to the appropriate opening, but gaskets prevent sound from exiting through inappropriate openings.—GLA

7,421,088

43.38.Si MULTIFUNCTION TRANSDUCER

David B. Cranfill *et al.*, assignors to Motorola, Incorporated
2 September 2008 (Class 381/386); filed 28 August 2003

A dual-purpose transducer for a cellular phone is disclosed. The diaphragm assembly **246** operates normally to reproduce frequencies above 200 Hz or so. However, the magnetic assembly **226** is spring-mounted,



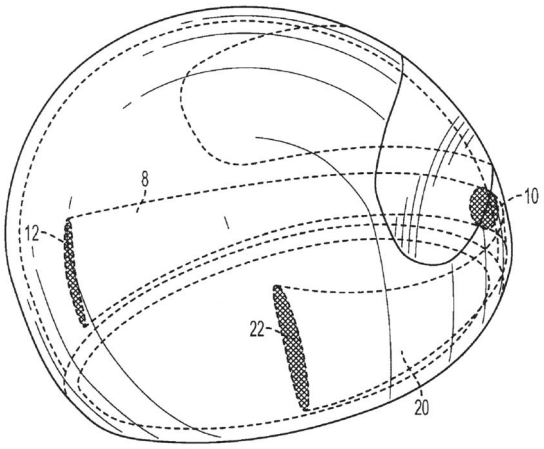
establishing a mechanical resonance around 150 Hz that can also be excited by the voice coil, allowing low frequency vibrations to be felt by the user. The actual resonance frequency is chosen to lie between two musical notes. "The amplitude of excitation of the first resonance by musical notes that might perchance be included in a signal applied to the MFT [multifunction transducer] **130** is reduced, reducing the distortion of sound generated by the MFT **130**, and at least substantially reducing the generation of undesirable mechanical noises." Although the novelty of this idea appears to be patentable, it is almost worthless in practice; under ideal conditions the mechanical resonance as graphed provides only about 6 dB of discrimination. Moreover, musicians tune their instruments to a variety of standards, and pitches may be electronically shifted during mixdown.—GLA

7,421,744

43.38.Si MOTORCYCLE HELMET WITH INTEGRATED ACOUSTIC VOICE AMPLIFIED CHAMBERS

John William Farrell, Manorville, New York
9 September 2008 (Class 2/423); filed 7 March 2007

Wireless communications systems for motorcycle riders are commercially available. These typically have earphones and a microphone embedded in the resilient lining of each user's safety helmet. This patent suggests that a pair of concealed speaking tubes **8**, **20** can pick up the driver's voice



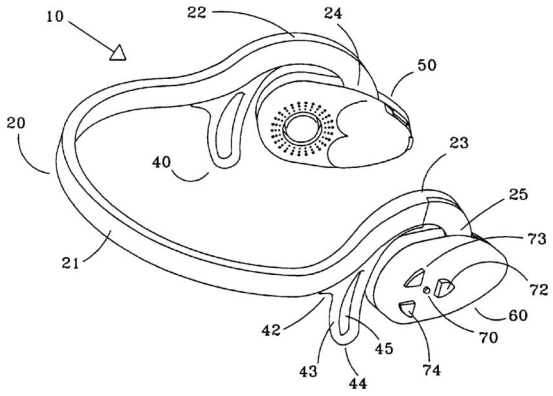
at entry opening **10** and project it from rear-facing exists **12**, **22** to a passenger behind. The idea may work to some degree, and it has the virtue of simplicity, but two-way communication is problematic.—GLA

7,426,282

43.38.Si HEAD SET SPEAKER AND STEREO PLAYING DEVICE

Glen T. Poss, assignor to US Design & Productions
16 September 2008 (Class 381/370); filed 6 September 2000

A combination headset, stereo player, and radio receiver is disclosed in this patent. Yoke **20** is worn behind the head, with flanges **40**, **41** located just behind the user's ears. Pods **50**, **60** are "aerodynamically shaped" and can be adjusted to allow some ambient sound to be heard. Each pod contains a loudspeaker and a battery pack, and one pod contains additional electronic circuitry. Each loudspeaker faces outward, and sound from the front of the cone travels through a coupling chamber to emerge from a ring of concentric perforations. Sound from the rear of the cone is ignored. This arrangement is



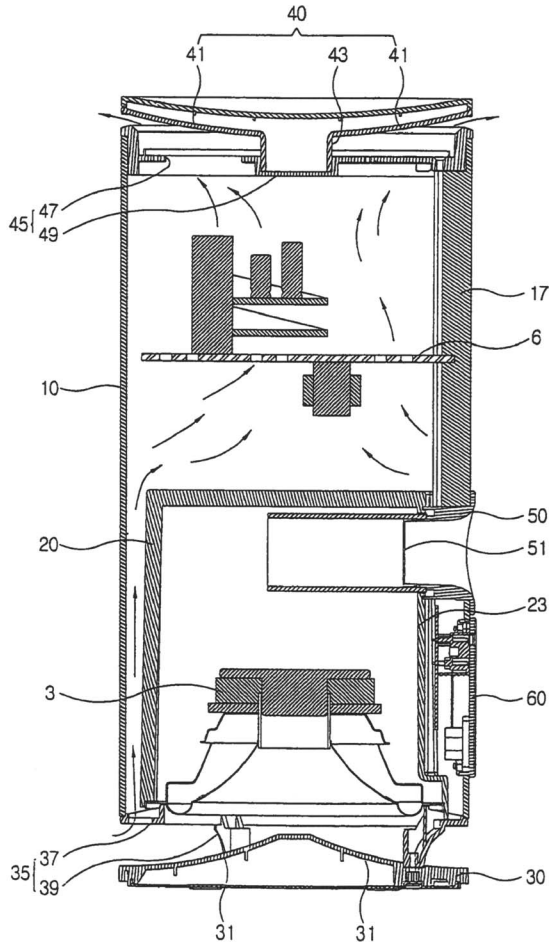
said to exhibit "…improved harmonics and acoustic fidelity." The cup-shaped magnetic assembly protrudes through the housing so that it can contact the user's skin, providing an electrical connection to the radio antenna.—GLA

7,388,963

43.38.Tj SPEAKER APPARATUS

Sang-hyun Han and Woo-nam Byun, assignors to Samsung Electronics Company, Limited
17 June 2008 (Class 381/397); filed in Republic of Korea 1 April 2003

Powered loudspeakers contain amplifiers that can generate heat. Some designs have a heatsink surface mounted on the loudspeaker case, but the interior of the box can become warm as the amplifier is mounted inside the box, and listeners can get burned if they come in contact with the exterior



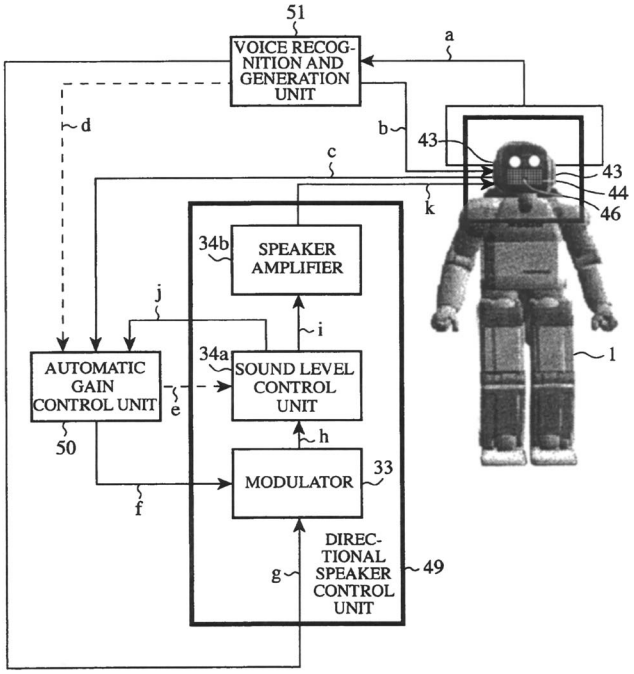
portion of the heatsink (although one could argue that good design may mitigate this). To solve these problems, amplifier assembly **6** is mounted inside casing **10** and is cooled by air entering opening **35** and exiting through design element **40**.—NAS

7,424,118

43.38.Tj MOVING OBJECT EQUIPPED WITH ULTRA-DIRECTIONAL SPEAKER

Kiyofumi Mori *et al.*, assignors to Honda Motor Company, Limited
9 September 2008 (Class 381/77); filed in Japan 10 February 2004

A parametric loudspeaker modulates a highly directional ultrasonic beam to generate audible sound from empty air. This patent explains how an ambulatory robot might be equipped with such a loudspeaker, plus an ultra-



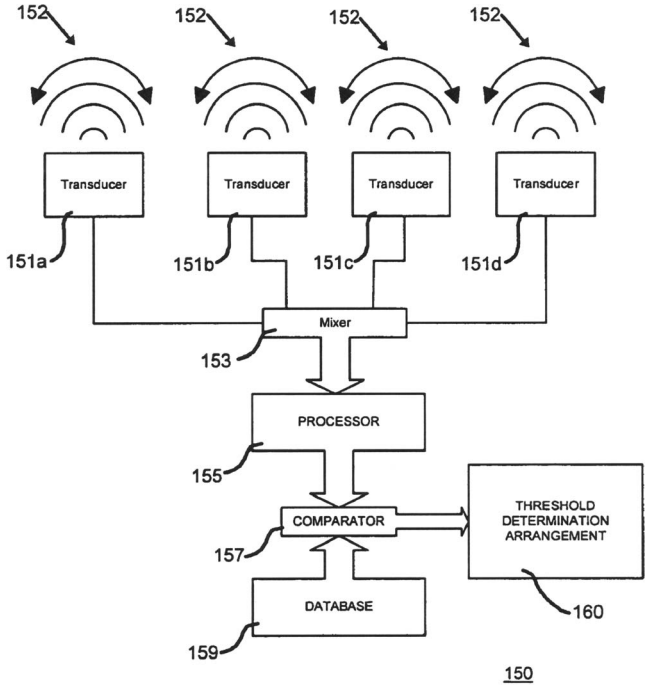
sonic receiver, to project its voice signals to a specific target. By including a range finder function, the distance from the target can be calculated and the sound level adjusted accordingly.—GLA

7,417,536

43.40.At LIVING BEING PRESENCE DETECTION SYSTEM

Sridhar Lakshmanan, Belleville, and Michigan *et al.*
26 August 2008 (Class 340/538); filed 25 April 2006

The authors describe "a system for distinguishing between a first condition corresponding to a living subject being directly in contact with an object of interest, and a second condition corresponding to an absence of contact…" What is being described is a combination of sensor techniques and signal processing techniques that allows the interpretation of signals in the 1–10 Hz range as signals due to the interaction (touching, heartbeat, breathing) of said subject beings with the sensors. It is not a clearly written patent, and the claims are very broad and are not supported in any detail by



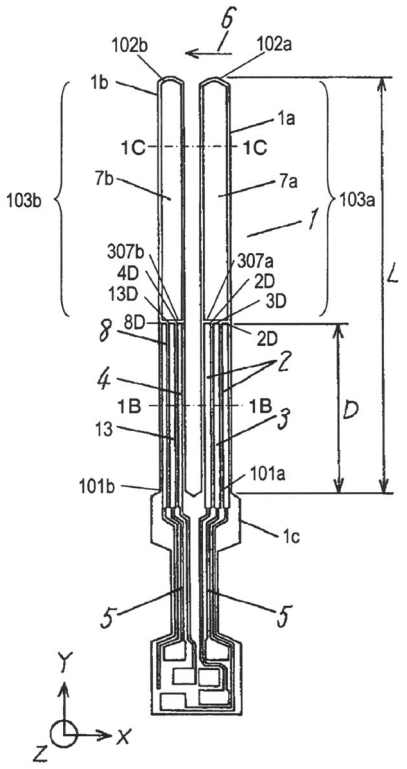
the data given. Nonetheless, there are some ideas discussed in this patent.—JAH

7,337,667

43.40.Cw ANGULAR VELOCITY SENSOR AND ITS DESIGNING METHOD

Satoshi Ohuchi and Hiroyuki Aizawa, assignors to Matsushita Electric Industrial Company, Limited
4 March 2008 (Class 73/504.16); filed in Japan 16 February 2004

A vehicle angular velocity sensor with minimal sensitivity to vehicle environment vibration and etched from a single silicon wafer is claimed. Arm **102a** of length **L** has a piezoelectric driver **2** distributed along extent **D**, which causes arm **102a** vibration **6** in the X-direction. Sensor rotation on



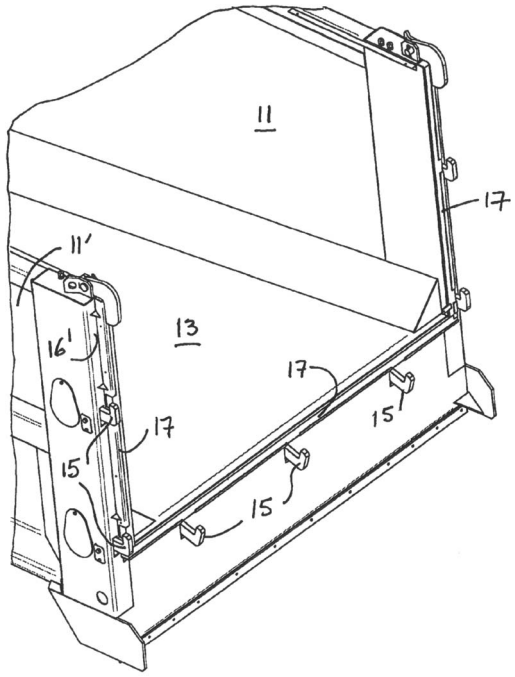
the Y-axis causes a vibrational component in the Z-direction that is coupled to arm **102b** to be detected by piezoelectric detector **13**, also distributed along a segment of length **D**. Immunity from environmental vibration is optimized by **D**/**L** being about 0.4. Criteria for choice of bending modal frequencies of **102a** and **102b** are not clear.—AJC

7,338,106

43.40.Tm SOUND DAMPING AND SEALING SYSTEM FOR TAIL GATE

Dany Poudrier, assignor to Michel Gohier Ltée
4 March 2008 (Class 296/50); filed in Canada 22 September 2005

The patent describes a dump truck tail gate noise and vibration damper for noise emitted when the dump truck driver alternately starts and stops vehicle motion so as to cause the tail gate to swing and impact the truck



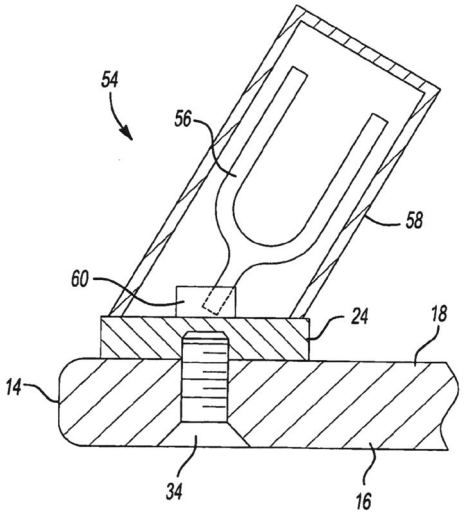
body, jarring the dumping content to complete discharging. Rubber strip **17** is compressed on installation to contact the tail gate and the truck body, thereby damping the sound and vibration of both surfaces.—AJC

7,396,294

43.40.Tm IMPACT FORCE DAMPENING SYSTEM FOR USE WITH A GOLF PUTTER HEAD

Joseph Consiglio, Hilton Head Island, South Carolina
8 July 2008 (Class 473/332); filed 8 December 2005

"The present invention utilizes vibration generating/redirecting components associated with the putter head, offsetting the twisting tendencies of the putter head from acting upon a golf ball contacted offset from a mass centerline associated with the putter head." Nice idea, but exactly how this is accomplished is unclear, as the several embodiments, for example **54**,



have quite different mass spring components and implementations, and the patent does not offer any quantitative analysis.—NAS

7,420,778

43.40.Tm SUSPENSION COMPONENT WITH SEALED DAMPING LAYER

Joseph H. Sassine *et al.*, assignors to Seagate Technology LLC
2 September 2008 (Class 360/244.9); filed 22 December 2003

Damping of small flexible components, such as disc drive head gimbal assemblies, is desirable in order to attenuate resonant vibrations. Friable particles may break loose from damping materials adhered to these components and contaminate sensitive equipment. As described in this patent, a nonfriable sealing layer that is added atop the damping layer to eliminate this contamination may also serve as a constraining layer of a constrained-layer sandwich damping arrangement.—EEU

7,422,190

43.40.Tm DEVICE FOR DAMPING THE VIBRATIONS OF A CABLE AND RELATED DAMPING METHOD

Jean-Pierre Messein and Benoit Lecinq, assignors to Freyssinet International (STUP)
9 September 2008 (Class 248/636); filed in France 3 September 2003

The cable bundles of stayed bridges, which typically are anchored to a tower and to the bridge deck, tend to vibrate as the result of traffic loads and of wind acting on the cables. A cable is surrounded at one of its anchored ends by a tubular collar that is connected relatively rigidly to the bridge structure. Piston-type dampers that act in two or more directions radial to the cable are connected between the cable bundle and the collar.—EEU

7,421,264

43.40.Vn DEVICE AND METHOD FOR REDUCING VIBRATION EFFECTS ON POSITION MEASUREMENT

William Alberth, Jr. and Lawrence Schumbacher, assignors to Motorola, Incorporated
2 September 2008 (Class 455/283); filed 29 October 2002

Electronic devices, such as cell phones, may be equipped with circuitry for global position measurement systems or the like. Such systems are sensitive to small changes or variations in the control signal frequency that may result from mechanical vibrations produced by a vibrator intended to alert the user. The means for overcoming this problem described in this patent consists of detecting when the vibrator is active and not performing position measurements during these time intervals.—EEU

7,421,349

43.40.Yq BEARING FAULT SIGNATURE DETECTION

Jason Stack, assignor to United States of America as represented by the Secretary of the Navy
2 September 2008 (Class 702/35); filed 15 May 2006

Detection of developing faults in a ball bearing is accomplished by measuring the vibrations of the apparatus containing the bearing and searching for signatures unique to bearing faults. This search is done by a detector that relies on signatures predicted by a complex signature fault model. This model is claimed not to be affected by changes in the apparatus' frequency response characteristics (such as may result from changes in coupled loads and mounting tightness) and to operate efficiently in the presence of extraneous vibrations.—EEU

7,424,827

43.40.Yq INSPECTING METHOD OF ELASTIC BODY, INSPECTING APPARATUS THEREOF, AND DIMENSION PREDICTING PROGRAM THEREOF

Tomohiro Yamada *et al.*, assignors to NGK Insulators, Limited
16 September 2008 (Class 73/579); filed in Japan 27 April 2004

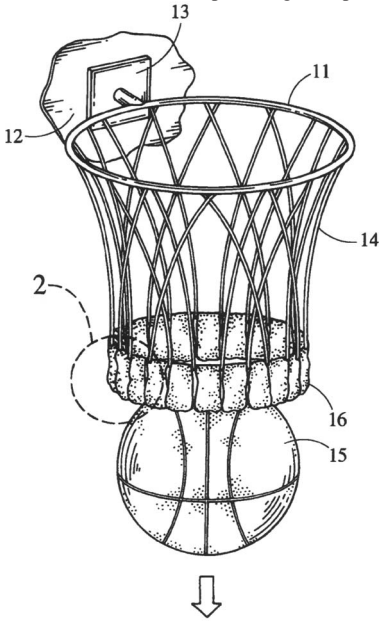
The inventors teach a method for inspection and quality control of actuators (both piezoelectric and electrostrictive) that uses external vibrational excitation in preference to excitation by an applied electric field as is the norm. It is implied in the discussion that this approach offers a more accurate measurement of the dimensions and quality factor of the actuators. What is interesting about their technique is that it uses a combination of mechanical and electrical sensing to develop both space and frequency domain pictures of the vibrations of the device under test. Unfortunately, the text is very difficult to understand and a lot of details pertaining to the sensing methods are left undisclosed. It appears from the sample data given that this method is actually in use at their company.—JAH

7,390,274

43.50.Ed BASKETBALL NET

William L. Bradford, Detroit, Michigan
24 June 2008 (Class 473/485); filed 12 September 2005

Device **2** is attached to basketball net **14** to enhance and augment the sound produced when basketball **15** drops through hoop **11**.—NAS

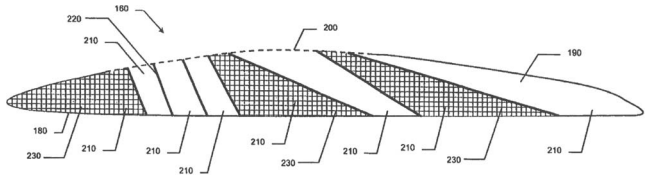


7,334,998

43.50.Gf LOW-NOISE FAN EXIT GUIDE VANES

Michael G. Jones *et al.*, assignors to The United States of America as represented by the Administrator of the National Aeronautics and Space Administration
26 February 2008 (Class 416/227 R); filed 6 December 2004

A turbofan engine noise absorbing air bypass duct guide vane is claimed. Guide vane **160** has porous surface **200** that communicates sound into resonant cavities **210** tuned to various noise frequencies to absorb that



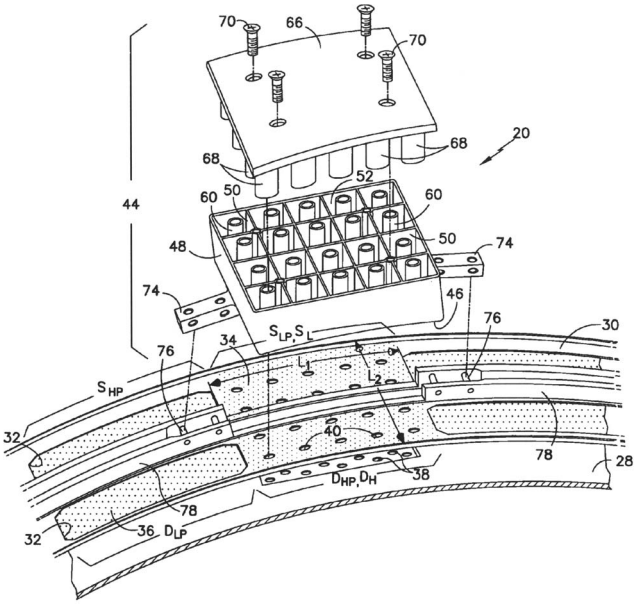
sound energy. Chambers **F** and **G** are filled with sound absorbing material.—AJC

7,337,875

43.50.Gf HIGH ADMITTANCE ACOUSTIC LINER

William Proscia *et al.*, assignors to United Technologies Corporation
4 March 2008 (Class 181/214); filed 28 June 2004

A tuned gas turbine aircraft engine exhaust noise attenuator lining **44**



is claimed. Exhaust noise enters openings **40** to pass into Helmholtz resonators **60–68** to be absorbed.—AJC

7,412,801

43.55.Ev SOUND ABSORBING PANEL

Panagiotis Papakonstantinou, GR-13671 Acharnai Attikis, Greece
19 August 2008 (Class 52/144); filed in Greece 21 October 2004

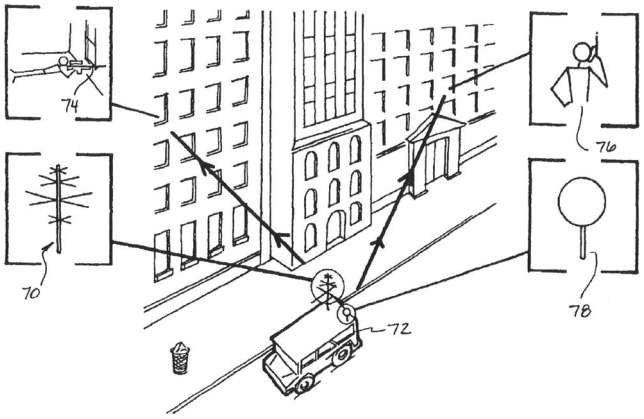
The patent is for a strong, durable, self-contained panel with a metal backing into which a sound absorptive material is inserted. In this manner, the panel can be interchanged, replaced, repaired, joined with other panels, and moved independently from the structural sound isolation barrier to which it is attached. The metal backing is folded at the edges to form a rigid perimeter flange for the absorptive insert.—CJR

7,423,934

43.60.Gk SYSTEM FOR DETECTING, TRACKING, AND RECONSTRUCTING SIGNALS IN SPECTRALLY COMPETITIVE ENVIRONMENTS

Charles A. Uzes, Athens, Georgia
9 September 2008 (Class 367/135); filed 9 November 2006

This patent is a continuation of earlier U.S. Patent 7,123,548, filed in 2005. The technique described can be used with electromagnetic or acoustic waves. "The present invention relates to a system for three dimensional multiple signal tracking, and reconstruction for use in connection with search and rescue, surveillance, storm and severe weather alerting, animal



and bird migration, subsurface mapping, anti-terrorism, conventional warfare, etc." Information from one or more receivers is digitized and then converted to "…a signal vector providing a mathematical model of the physical wave field." The conversion process may use a library of appropriate signal vectors for comparison.—GLA

7,415,125

43.66.Ts APPARATUS AND METHOD FOR CREATING ACOUSTIC ENERGY IN A RECEIVER ASSEMBLY WITH IMPROVED DIAPHRAGMS-LINKAGE ARRANGEMENT

Daniel M. Warren *et al.*, assignors to Knowles Electronics, LLC
19 August 2008 (Class 381/418); filed 10 May 2004

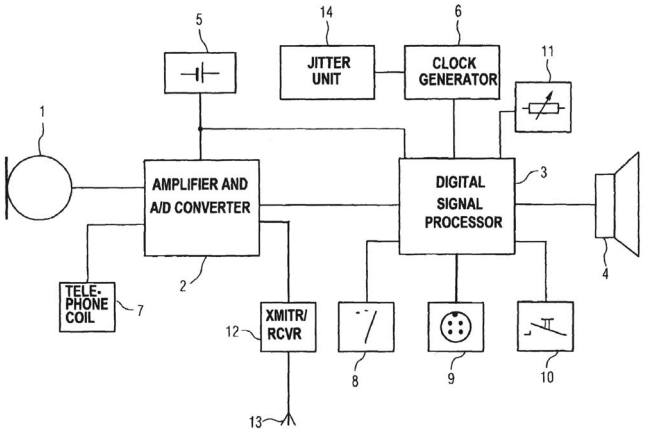
Transducer design methodology is described for internally damping the vibrations and magnetic signals that are produced by a hearing aid receiver. Bellowslike members with accordionlike structures expand and contract in response to flexing of linkage assemblies coupled to two diaphragm assemblies and to a two-legged armature.—DAP

7,421,085

43.66.Ts HEARING AID DEVICE OR HEARING DEVICE SYSTEM WITH A CLOCK GENERATOR

Kunibert Husung and Torsten Niederdränk, assignors to Siemens Audiologische Technik GmbH
2 September 2008 (Class 381/312); filed in Germany 30 September 2002

To prevent internally generated electromagnetic interference from interfering with wireless signal transmission in a hearing device, the system clock frequency is modulated slightly around an average frequency with a



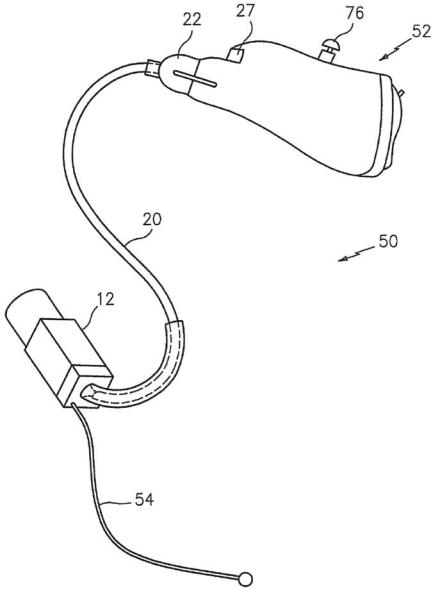
high frequency sinusoid or noise to produce jitter that varies the digitally interfering signals.—DAP

7,421,086

43.66.Ts HEARING AID SYSTEM

Natan Bauman *et al.*, assignors to Vivatone Hearing Systems, LLC
2 September 2008 (Class 381/328); filed 13 January 2006

A behind-the-ear hearing aid or other hearing device with external microphone leaves the ear open by utilizing a speaker housed in a casing having a 0.23 in. maximum lateral dimension. The design is said to generate



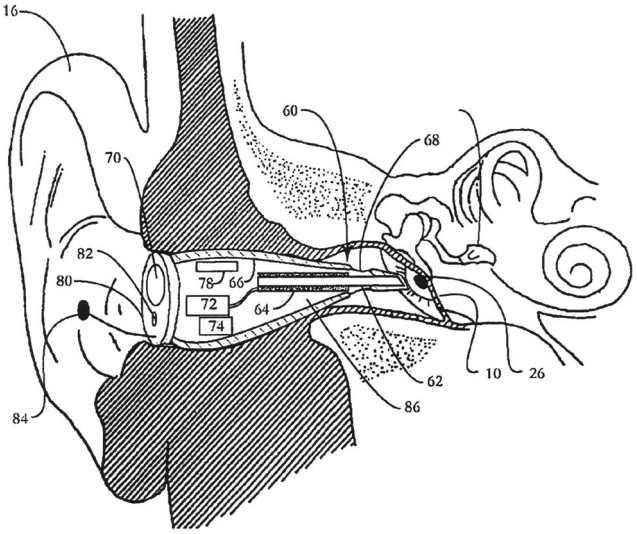
a maximum of 8 dB of insertion loss and a maximum of 8 dB occlusion effect over "human audible" frequencies.—DAP

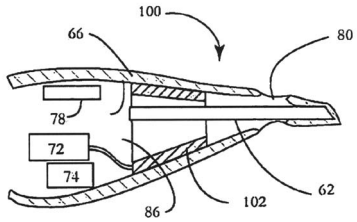
7,421,087

43.66.Ts TRANSDUCER FOR ELECTROMAGNETIC HEARING DEVICES

Rodney C. Perkins *et al.*, assignors to EarLens Corporation
2 September 2008 (Class 381/331); filed 28 July 2004

An in-the-ear-canal transmitter sends inductive signals to a transducer that is releasably attached to the tympanic membrane or to either the malleus, incus, or stapes in the middle ear. Variations in the magnetic field emitted by the transmitter produce vibrations in the middle ear via the passive transducer (typically a magnet) that are transmitted to the cochlea. The





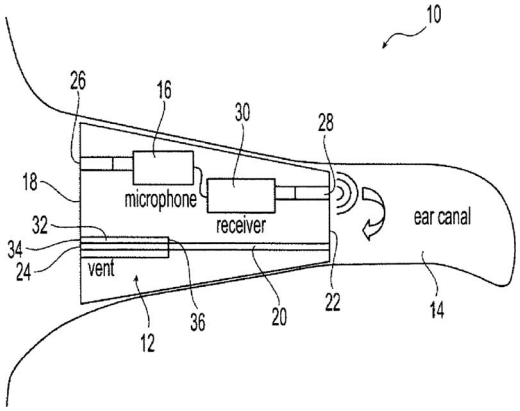
transmitter includes a coil would on a core that is positioned at a predetermined distance and orientation relative to the implanted transducer.—DAP

7,424,122

43.66.Ts HEARING INSTRUMENT VENT

James G. Ryan, assignor to Sound Design Technologies, Limited
9 September 2008 (Class 381/322); filed 5 April 2004

To reduce acoustic feedback, a hearing aid vent is surrounded by a series of cells which are designed as quarter-wave resonators so they propagate sound out of the ear canal at a specific acoustic feedback-prone frequency out of phase with the energy propagating from the vent. The result is that the total radiated sound propagating out of the ear canal is reduced, thus reducing the tendency for acoustic feedback oscillation.—DAP



7,424,123

43.66.Ts CANAL HEARING DEVICE WITH TUBULAR INSERT

Adnan Shennib and Richard C. Urso, assignors to Insound Medical, Incorporated
9 September 2008 (Class 381/328); filed 24 February 2004

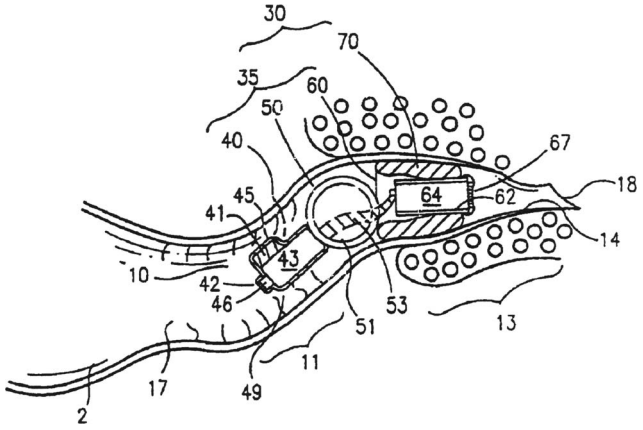
A two-piece mass-producable, standard hearing aid that fits deep into the ear canal uses a dual acoustic seal system whose goals are to improve high frequency response, prevent acoustic feedback, and minimize occlusion. In the preferred embodiment, a first module is positioned in the cartilaginous portion of the ear canal. A second replaceable, disposable module consists of a tubular insert with flexible sound conduction tube that provides a primary concentric acoustic seal with pressure vent in the bony region of the ear canal and a secondary concentric seal with larger occlusion-relief vent in the cartilaginous region.—DAP

7,424,124

43.66.Ts SEMI-PERMANENT CANAL HEARING DEVICE

Adnan Shennib *et al.*, assignors to Insound Medical, Incorporated
9 September 2008 (Class 381/328); filed 26 April 2005

A three-part hearing device includes a soft sealing retainer that seats deeply in the bony portion of the ear canal and surrounds a moisture-proof, encapsulated receiver assembly. The device also has a nonoccluding encapsulated microphone assembly that sits in the cartilaginous part of the ear canal and a separate battery assembly. The device may contain a reed-switch and magnet for remote power switching or control.—DAP

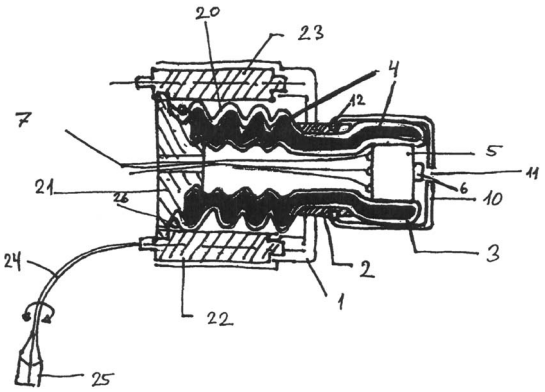


7,425,196

43.66.Ts BALLOON ENCAPSULATED DIRECT DRIVE

Martin Bondo Jorgensen and Karsten Videbaek, assignors to Sonion Roskilde A/S
16 September 2008 (Class 600/25); filed 22 December 2003

A receiver module of a hearing aid designed to fit into the ear canal includes a receiver housing with a surrounding expansible structure that is encircled by an elastic encapsulation material which forms a waterproof seal for the receiver. An electrically activated miniature pump integrated into the



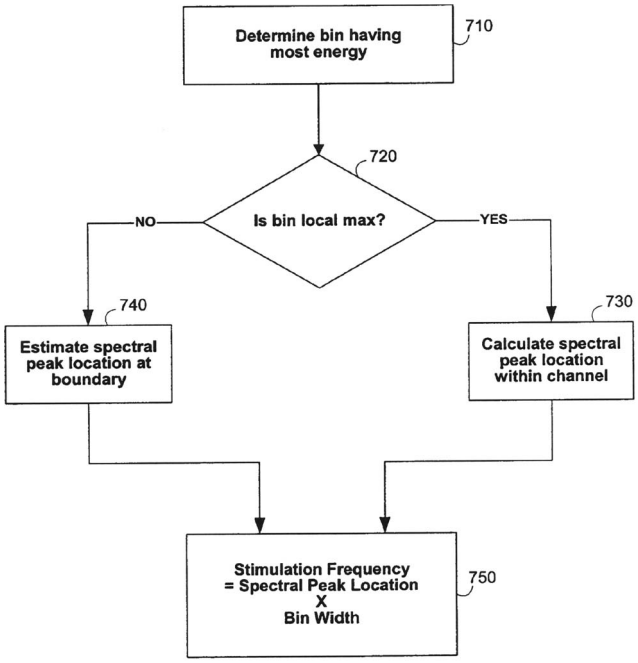
hearing aid controls the pressure in the expansible structure as a function of acoustic gain in response to signals from the hearing aid signal processor.—DAP

7,426,414

43.66.Ts SOUND PROCESSING AND STIMULATION SYSTEMS AND METHODS FOR USE WITH COCHLEAR IMPLANT DEVICES

Leonid M. Litvak *et al.*, assignors to Advanced Bionics, LLC
16 September 2008 (Class 607/56); filed 14 March 2005

Simultaneous stimulation of implanted electrodes including temporal information utilizes stimulation of virtual electrodes positioned in the cochlea at a location corresponding to a frequency at which a spectral peak is



located within an assigned channel. The modulation depth of the waveform decreases with increasing selected frequency and is based on a rate at which each of the channels is updated.—DAP

7,424,098

43.72.Gy SELECTABLE AUDIO AND MIXED BACKGROUND SOUND FOR VOICE MESSAGING SYSTEM

Renee M. Kovales *et al.*, assignors to International Business Machines Corporation
9 September 2008 (Class 379/76); filed 31 July 2003

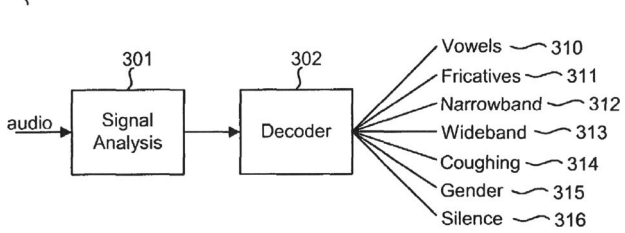
Contextual information, including music or background sounds related to emotional cues of the caller, are added as embedded audio files to enhance voice mail messages.—DAP

7,424,427

43.72.Gy SYSTEMS AND METHODS FOR CLASSIFYING AUDIO INTO BROAD PHONEME CLASSES

Daben Liu and Francis G. Kubala, assignors to Verizon Corporate Services Group Incorporated
9 September 2008 (Class 704/256.1); filed 16 October 2003

Sound is classified as a vowel or a fricative with first or second phoneme-based hidden Markov models, respectively. A non-phoneme-based



model is also used to classify based on bandwidth, silence, the speaker gender, and other nonspeech sounds such as coughing.—DAP

7,415,093

43.80.Vj METHOD AND APPARATUS OF CT CARDIAC DIAGNOSTIC IMAGING USING MOTION A PRIORI INFORMATION FROM 3D ULTRASOUND AND ECG GATING

John Eric Tkaczyk *et al.*, assignors to General Electric Company
19 August 2008 (Class 378/8); filed 30 October 2006

Ultrasound and electrocardiogram data *t* are acquired from the heart in real-time. The data are used prospectively to gate acquisition of x-ray CT (computer tomography) data. An image is reconstructed from the CT data.—RCW