CODEN: JASMAN

ISSN: 0001-4966

The Journal of the Acoustical Society of America

Vol. 122, No. 1 July 2007

(Continued)

CONTENTS—*Continued from preceding page*

(Continued)

CONTENTS—*Continued from preceding page*

(Continued)

CONTENTS—*Continued from preceding page*

Vowel formants from the wave equation

Antti Hannukainen, Teemu Lukkari, Jarmo Malinen, and Pertti Palo
Institute of Mathematics, Helsinki University of Technology, P.O. Box 1100, FI-02015 TKK, Espoo, Finland
antti.hannukainen@tkk.fi, teemu.lukkari@tkk.fi, jarmo.malinen@tkk.fi, pertti.palo@tkk.fi

Abstract: This article describes modal analysis of acoustic waves in the human vocal tract while the subject is pronouncing [øː]. The model used is the wave equation in three dimensions, together with physically relevant boundary conditions. The geometry is reconstructed from anatomical MRI data obtained by other researchers. The computations are carried out using the finite element method. The model is validated by comparing the computed modes with measured data.

PACS numbers: 43.70.Bk, 43.20.Ks [AL]
Date Received: March 6, 2007 **Date Accepted:** April 23, 2007

1. Introduction

The purpose of this article is to study vowel production by the wave equation with boundary conditions as specified by Eq. (2). This model constitutes the input part of a (*scattering*) *conservative linear dynamical system* as presented by, e.g., Malinen *et al.* (2006); Malinen and Staffans (2006, 2007). A preliminary version of the present work was presented at the Phonetics Symposium 2006 (Hannukainen *et al.*, 2006).

In the past, the vocal tract (VT) acoustics has been modeled in a number of different ways. Electrical transmission lines have been used for a long time (see, e.g., Dunn, 1950). The celebrated Kelly–Lochbaum model makes use of reflection coefficients obtained from a variable diameter tube (Kelly and Lochbaum, 1962). Such reflection coefficients appear in, e.g., models from geophysics and in interpolation theory (see Foias and Frazho, 1990). We remark that the Kelly–Lochbaum model is closely related to the horn model described by the Webster equation (see Chiba and Kajiyama, 1958; Fant, 1970). All these models have produced very accurate simulation results with a relatively light computational load, and they have applications, e.g., in mobile phones. More advanced two- and three-dimensional descendants of the Kelly–Lochbaum model are the transmission line networks that have been developed by El Masri *et al.* (1996, 1998); Mullen *et al.* (2006). For a recent review and further references, see Palo (2006).

Equation (2) in an anatomically realistic geometry has a more direct basis in physics than any of the approaches discussed in the previous paragraph. This is particularly useful in some applications, for example, in modeling the effects of anatomical abnormalities and maxillofacial surgery on speech (Dedouch *et al.*, 2002a; Nishimoto *et al.*, 2004; Švancara and Horáček, 2006). As solving Eq. (2) analytically is possible only in a radically simplified geometry (see Sondhi, 1986), we solve the problem numerically by the *finite element method* (FEM). This is the approach used by, e.g., Lu *et al.* (1993), Suzuki *et al.* (1993), Kawanishi *et al.* (1996), Niikawa *et al.* (2002), Dedouch *et al.* (2002b), Sasaki *et al.* (2003), and Švancara *et al.* (2004), too. Unfortunately, heavy computations are involved in this method.

We present a modal analysis of an anatomical configuration of [øː] as produced by a native Swedish speaker. We obtain resonance frequencies computationally, which correspond to formants. Unlike the scattering transfer function estimation used by Nishimoto *et al.* (2004) and Sasaki *et al.* (2003), our method does not necessarily require taking into account the radiation impedance at the mouth. Our approach is more closely related to Dedouch *et al.* (2002b) but instead of Neumann boundary condition on the glottis, we use a reflection-free boundary condition slightly above the glottis [see the last lines of Eqs. (2) and (4)]. Using reflection-free

boundary conditions Eq. (3), our Eq. (2) can be coupled to a glottis model in a physically realistic manner. Our results indicate that the computationally obtained formants identify the vowel [øː] correctly in a larger set of measured data.

For numerical computations, a detailed geometric description of the VT is necessary. Nowadays, accurate anatomical data can be obtained using magnetic resonance imaging (MRI). We are indebted to Dr. Olov Engwall (KTH) for kindly providing us with the required data.

2. Acoustic model

Deriving the wave equation for sound pressure starts by assuming that the total pressure $p = p(\mathbf{r},t)$ and the density $\rho = \rho(\mathbf{r},t)$ can be expressed as

$$p(\mathbf{r},t) = p_0 + p'(\mathbf{r}+t), \quad \rho(\mathbf{r},t) = \rho_0 + \rho'(\mathbf{r},t), \tag{1}$$

respectively, where p_0 and ρ_0 are independent of time t and space variable $\mathbf{r}$. For linearization of the equations, it is assumed that $p' = p'(\mathbf{r},t) \ll p_0$ and $\rho' = \rho'(\mathbf{r},t) \ll \rho_0$ are small perturbations at point $\mathbf{r} = (x,y,z) \in \Omega$ at time t. Here $\Omega \subset \mathbb{R}^3$ denotes the interior of the VT with boundary $\partial\Omega = \Gamma_1 \cup \Gamma_2 \cup \Gamma_3$, where Γ_1 is the mouth opening, Γ_2 denotes the walls of the VT, and Γ_3 is a virtual boundary control surface a small distance above the glottis.

By $\mathbf{v} = \mathbf{v}(\mathbf{r},t)$ denote the velocity field of the flow described by p and ρ. A velocity potential $\Phi = \Phi(\mathbf{r},t)$ is any function that satisfies $\mathbf{v} = -\nabla\Phi$. With this notation, our acoustic model is given by

$$\Phi_{tt} = c^2 \Delta\Phi \quad \text{on } \Omega,$$

$$\Phi = 0 \text{ on } \Gamma_1, \quad \frac{\partial\Phi}{\partial\nu} = 0 \text{ on } \Gamma_2, \quad \Phi_t = c\frac{\partial\Phi}{\partial\nu} = 2\sqrt{\frac{c}{\rho_0}}u \text{ on } \Gamma_3, \tag{2}$$

where $u = u(\mathbf{r},t)$ is the incoming power (per unit area) at glottis input, c is the sound velocity in the VT, ν is the exterior unit normal on $\partial\Omega$, and $\partial\Phi/\partial\nu = \nu \cdot \nabla\Phi$. The problem is to compute the velocity potential $\Phi(\mathbf{r},t)$ for a given glottal input function $u(\mathbf{r},t)$.

To derive Eq. (2) from "first principles," one needs to assume that an isentropic thermodynamic equation of state for pressure $p = p(s,\rho)$ holds where s, ρ are the entropy and density, respectively. Then we define the sound speed c by linearizing the equation of state $p' = p(s,\rho_0+\rho') - p(s,\rho_0) \approx c^2\rho'$ where $p_0 = p(s,\rho_0)$ and $c^2 = (\partial p/\partial\rho)(s,\rho_0)$. In this approximation, the entropy s is kept constant since the associated thermodynamic process is assumed to be reversible. In the case of monatomic ideal gas, we have $p/\rho^\gamma = p_0/\rho_0^\gamma$ and $c^2 = \gamma p_0/\rho_0$ where $\gamma = 5/3$ is the adiabatic constant.

Now the wave equation $\Phi_{tt} = c^2\Delta\Phi$ can be derived by a linearization argument involving the continuity equation, Euler equation and linearized equation of state $p' = c^2\rho'$. Having computed Φ, we obtain the perturbation pressure from $p' = \rho_0\Phi_t$. All this can be found, e.g., in Fetter and Walecka (1980, Chap. 9).

Equation (2) is sophisticated enough to capture many relevant properties of wave propagation in three-dimensional geometry (e.g., to detect cross modes). It can also be used as the theoretical starting point in deriving the Webster equation mentioned earlier. However, it does not take into account turbulence, shock formation, or losses due to viscosity, heat conduction, or boundary dissipation on Γ_2.

We also need to take into account the walls and both ends of the VT. The last line in Eq. (2) specifies the required boundary conditions. We regard the mouth as an open end of an acoustic tube, and this is described by the Dirichlet condition $\Phi(\mathbf{r},t) = 0$. More complicated models for the mouth opening or the surrounding acoustic space have been considered by Kawanishi *et al.* (1996) (an impedance model involving Bessel functions), Nishimoto *et al.* (2004) (an impedance model consisting of a small reflecting hemisphere), and Švancara *et al.* (2004) (an exterior model of two concentric spheres with an absorbing outer boundary).

On the walls of the VT, we use the same Neumann condition $(\partial\Phi/\partial\nu)(\mathbf{r},t)=0$ as one would use at the closed end of a resonating tube. These two boundary conditions are discussed by Fetter and Walecka (1980, pp. 306, 307).

At the glottis end, we use a scattering boundary condition that specifies the ingoing sound energy wave. For motivation, we define the ingoing wave $u(\mathbf{r},t)$ and the outgoing wave $y(\mathbf{r},t)$ for $\mathbf{r}\in\Gamma_3$ by

$$u=\sqrt{\frac{\rho_0}{4c}}\left(c\frac{\partial\Phi}{\partial\nu}+\Phi_t\right),\quad y=\sqrt{\frac{\rho_0}{4c}}\left(c\frac{\partial\Phi}{\partial\nu}-\Phi_t\right). \tag{3}$$

The first of these equations coincides with the third boundary condition in Eq. (2). The net power absorbed by the interior domain Ω through the control/observation boundary at time t satisfies

$$\int_{\Gamma_3}|u(\mathbf{r},t)|^2\mathrm{d}\omega(\mathbf{r})-\int_{\Gamma_3}|y(\mathbf{r},t)|^2\mathrm{d}\omega(\mathbf{r})=\int_{\Gamma_3}(-\nu(\mathbf{r}))\cdot\mathbf{j}_e(\mathbf{r},\mathbf{t})\mathrm{d}\omega(\mathbf{r}),$$

where $\mathbf{j}_e=-\rho_0\Phi_t\nabla\Phi=p'\mathbf{v}$ is the energy-flux vector as introduced in Fetter and Walecka (1980, 307 pp.).

Instead of solving Eq. (2), we solve an easier—yet relevant—problem related to Eq. (2). More precisely, we determine the resonance frequencies corresponding to a particular vowel articulation position. By Malinen and Staffans (2006, Theorem 2.3), the resonances of Eq. (2) can be solved by finding the discrete, complex frequencies λ and the corresponding nonzero eigenfunctions $\Phi_\lambda(\mathbf{r})$ such that

$$\lambda^2\Phi_\lambda=c^2\Delta\Phi_\lambda \text{ on } \Omega,$$

$$\Phi_\lambda=0 \text{ on } \Gamma_1,\quad \frac{\partial\Phi_\lambda}{\partial\nu}=0 \text{ on } \Gamma_2,\quad \lambda\Phi_\lambda+c\frac{\partial\Phi_\lambda}{\partial\nu}=0 \quad \text{on } \Gamma_3 \tag{4}$$

are satisfied. The time harmonic extension $\Phi(\mathbf{r},t)=\Phi_\lambda(\mathbf{r})e^{\lambda t}$ of Φ_λ satisfies clearly Eq. (2). Using the connection $p'=\rho_0\Phi_t$, the corresponding perturbation pressure distribution is given by $p'(\mathbf{r},t)=p_\lambda(\mathbf{r})e^{\lambda t}$, where $p_\lambda(\mathbf{r}):=\rho_0\lambda\Phi_\lambda(\mathbf{r})$. Thus Eq. (4) is satisfied with p_λ in place of Φ_λ.

3. Finite element modeling

The variational formulation of Eq. (4) (with p_λ in place of Φ_λ) is

$$\lambda^2\int_\Omega p_\lambda\phi\mathrm{d}\Omega+\lambda_c\int_{\Gamma_3}p_\lambda\phi\mathrm{d}\omega+c^2\int_\Omega\nabla p_\lambda\cdot\nabla\phi\mathrm{d}\Omega=0, \tag{5}$$

where ϕ is an arbitrary test function in Sobolev space $H^1_{\Gamma_1}(\Omega)=\{f\in H^1(\Omega):f(\mathbf{r})=0 \text{ for } \mathbf{r}\in\Gamma_1\}$. The FEM can be used to approximately solve Eq. (5); see, e.g., Johnson (1987) for an elementary treatment. We use piecewise linear shape functions and a tetrahedral mesh of $n=64\,254$ elements which gives sufficiently accurate results. We obtain three $n\times n$ matrices, namely the stiffness matrix $\boldsymbol{K}$, the mass matrix $\boldsymbol{M}$, and $\boldsymbol{P}$ representing the glottis boundary condition in Eq. (4).

When treating Eq. (5) we proceed to solve the following linear algebra problem: Find all complex numbers λ and corresponding nonzero vectors $\mathbf{x}(\lambda)$ such that

$$\lambda^2\boldsymbol{K}\mathbf{x}(\lambda)+\lambda c\boldsymbol{P}\mathbf{x}(\lambda)+c^2\boldsymbol{M}\mathbf{x}(\lambda)=0 \quad\Leftrightarrow\quad \boldsymbol{A}\mathbf{y}(\lambda)=\lambda\boldsymbol{B}\mathbf{y}(\lambda), \tag{6}$$

where $\boldsymbol{A}=\begin{bmatrix}-c\boldsymbol{P} & -c^2\boldsymbol{M}\\ \boldsymbol{I} & 0\end{bmatrix}$, $\boldsymbol{B}=\begin{bmatrix}\boldsymbol{K} & 0\\ 0 & \boldsymbol{I}\end{bmatrix}$, and $\mathbf{y}(\lambda)=\begin{bmatrix}\lambda\mathbf{x}(\lambda)\\ \mathbf{x}(\lambda)\end{bmatrix}$ (Saad, 1992). The numbers λ are good approximations of the λ's appearing in Eq. (4), provided that the number n of elements is high enough. The lowest formants $\mathrm{F1},\mathrm{F2},\ldots$, correspond to the numbers λ in the order of increasing imaginary part.

Table 1. Computed, measured, and scaled formants for [øː] in kilohertz.

	F1	F2	F3	F4
Computed	0.68	1.35	2.71	3.79
Measured	0.50	1.06	2.48	3.24
Scaled	0.56	1.11	2.22	3.10

4. Data

Figure 1 in Hannukainen *et al.* (2006) shows a sliced representation of the VT geometry that we have used as the basis of our analysis. There are 29 slices, each consisting of 51 points, and they define the VT from glottis to mouth. For faster computation, the slices were downsampled by taking into account only every fourth point.

The raw MRI data was collected from a native male speaker of Swedish while he pronounced a prolonged vowel [øː] in supine position. Engwall and Badin (1999) describe the MR imaging procedure and image postprocessing. Corresponding formant measurement data are also available in the same article. The formants were estimated from speech recorded on a different occasion but with the same subject in a similar supine condition.

5. Results and conclusions

The latter form of Eq. (6) was solved in MATLAB environment, and the formants F1 to F4 that we obtained are shown in Table 1. These computed formants are roughly $3\frac{1}{2}$ semitones too high compared to the measured values, and we will discuss the physical background of this discrepancy in the following. The bottom row in Table 1 shows the computed formants multiplied by 0.817, which corresponds to a difference of $3\frac{1}{2}$ semitones.

We also obtained the resonance modes p_λ—see Eq. (4)—corresponding to the formants F1–F4. They are computed as linear combinations of the element basis functions, using the components of $\mathbf{x}(\lambda)$ as weights. We note that the perturbation pressures p_λ are not given here in any physically relevant scale but they have been normalized so that the maximum deviation from the static pressure p_0 is either 1 or −1. Figure 1 shows isobars for the fourth mode. Figure 2 shows the pressure distributions of the modes. Figures 1 and 2 are plotted along a cross-sagittal midline cut (see Fig. 1, Hannukainen *et al.*, 2006). We remark that Fig. 1 supports the hypothesis that a weak cross-mode resonance related to F4 should appear in the oral cavity.

The vowels from Engwall and Badin (1999: Table 4), together with the scaled and computed $[øː]_{s,c}$ from Table 1, are plotted in the (F2, F1) plane in Fig. 3. Clearly, $[øː]_{s,c}$ is closer to measured [øː] than to any other measured vowel, *except* possibly [ɑː]. To further clarify the situation, let us consider the formants F1 to F4 for $[øː]_{s,c}$, [øː], and [ɑː] as vectors: $[øː]_{s,c}=(0.56,1.11,2.22,3.10)$, $[øː]=(0.5,1.06,2.48,3.24)$, and $[ɑː]=(0.56,0.94,2.74,3.24)$. Then the Euclidean distance between $[øː]_{s,c}$ and [øː] is 0.31, but the distance between $[øː]_{s,c}$ and [ɑː] is significantly larger, equaling 0.57. This difference is explained by F3,

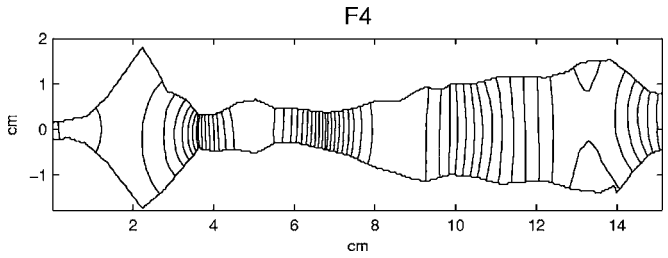


Fig. 1. Isobar corresponding to F4 along a midline cut. The mouth is on the right.

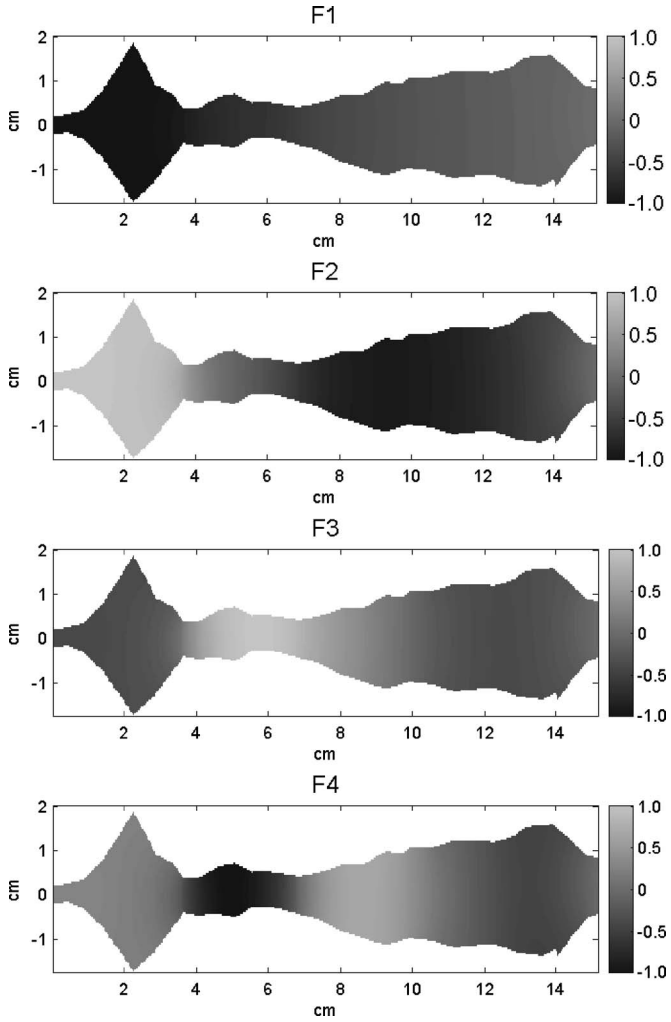


Fig. 2. Pressure distributions for F1–F4 along a midline cut. The mouth is on the right.

since the fourth formants are almost the same. We conclude that the *first two* formants classify the scaled, computed vowel $[\text{ø:}]_{s,c}$ almost correctly. Moreover, if we look at *all four* available formants, even the remaining ambiguity disappears.

As we pointed out earlier, the computed formants F1–F4 differ from the corresponding measured formants by $3\frac{1}{2}$ semitones. Having said that, the *ratios* between the computed formants and the measured formants match each other very well. There is a simple physical explanation why such a discrepancy is to be expected. In Eq. (2), we use the Dirichlet boundary condition on the lip opening. This results in a vibrational node at the opening. In reality, such a node would appear further away outside the mouth. In that sense, the real life VT is effectively longer than the one described by Eq. (2), resulting in lower formants. To get rid of this artifact, we should also model the surrounding acoustic space.

Surrounding acoustic space has been modeled by a lumped impedance for a transmission line (Laine, 1982), by using a "small space" model with impedance termination on the outer shell (Nishimoto *et al.*, 2004), and by using a "large space" model with an absorbing outer boundary (Švancara *et al.*, 2004). The first two of these approaches include a tuning parameter to be determined experimentally so that the measured and computed formants coincide. We

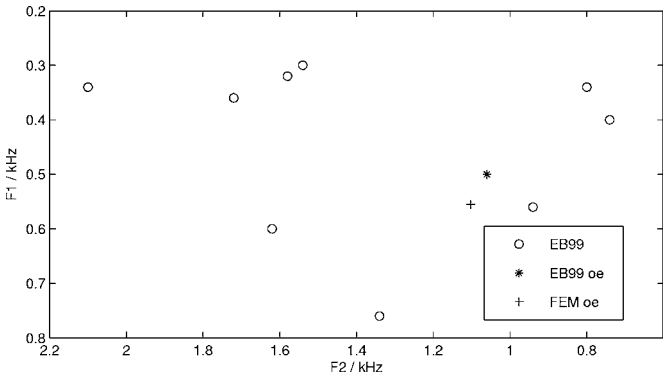


Fig. 3. Vowels in the (F2,F1) plane. FEM oe (+) is the scaled, computed [ø:] EB99 oe (*) is the measured [ø:] and EB99 (o) are other measured vowels. [EB99 denotes Engwall and Badin (1999).].

remark that impedance termination for the wave equation is inherently more difficult than for the transmission line, since the termination must be of boundary control instead just of point control type.

Acknowledgments

We would like to thank Olov Engwall from KTH, Stockholm, for providing the articulation geometry for this study. We would also like to thank an anonymous reviewer for valuable comments. A. H. has been supported by the Academy of Finland.

References and links

Chiba, T., and Kajiyama, M. (**1958**). *The Vowel, Its Nature and Structure*, Phonetic Society of Japan, Tokyo.
Dedouch, K., Horáček, J., Vampola, T., and Černý, L. (**2002a**). "Finite element modelling of a male vocal tract with consideration of cleft palate," in *Proceedings of the Forum Acusticum*, Seville, Sept. 16–20.
Dedouch, K., Horáček, J., Vampola, T., Švec, J., Kršek, P., and Havlík, R. (**2002b**). "Acoustic modal analysis of male vocal tract for Czech vowels," in *Proceedings Interaction and Feedbacks 2002*, Prague, pp. 13–19.
Dunn, H. K. (**1950**). "The calculation of vowel resonances, and an electrical vocal tract," J. Acoust. Soc. Am. **22**, 740–753.
El Masri, S., Pelorson, X., Saguet, P., and Badin, P. (**1996**). "Vocal tract acoustics using the transmission line matrix (TLM) method," in *Proceedings of the Fourth International Conference on Spoken Language Processing*, Philadelphia, Oct. 3–6, pp. 953–956.
El Masri, S., Pelorson, X., Saguet, P., and Badin, P. (**1998**). "Development of the transmission line matrix method in acoustics. Applications to higher modes in the vocal tract and other complex ducts," Int. J. Numer. Model. **11**, 133–151.
Engwall, O., and Badin, P. (**1999**). "Collecting and analysing two- and three-dimensional MRI data for Swedish," *Speech, Music and Hearing, Quarterly Progress and Status Report*, Vol. 40, pp. 11–38.
Fant, G. (**1970**). *Acoustic Theory of Speech Production* (Mouton, The Hague).
Fetter, A., and Walecka, J. (**1980**). *Theoretical Mechanics of Particles and Continua* (McGraw–Hill, New York).
Foias, C., and Frazho, A. E. (**1990**). *The Commutant Lifting Approach to Interpolation Problems*, Operator Theory: Advances and applications, Vol. **44** (Birkhäuser, Basel).
Hannukainen, A., Lukkari, T., Malinen, J., and Palo, P. (**2006**). "Formants and vowel sounds by finite element method," in *The Phonetics Symposium 2006*, Helsinki, Finland, pp. 24–33.
Johnson, C. (**1987**). *Numerical Solution of Partial Differential Equations by the Finite Element Method* (Cambridge University Press, New York).
Kawanishi, Y., Obuchi, M., Dang, J., Nakai, T., and Suzuki, H. (**1996**). "Consideration of the acoustic characteristics of the pyriform fossa in transmission line model, 3D-FEM model and realistic model," Tech. Rep. IEICE **EA96-12**, pp. 1–8.
Kelly, J., and Lochbaum, C. (**1962**). "Speech synthesis," in *Proceedings of the Fourth International Congress on Acoustics*, Paper G42, Copenhagen, Denmark, Sept., pp. 1–4.
Laine, U. K. (**1982**). "Modelling of lip radiation impedance in z-domain," in *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP-82)*, Paris, France, May, Vol. **3**.
Lu, C., Nakai, T., and Suzuki, H. (**1993**). "Finite element simulation of sound transmission in vocal tract," J. Acoust. Soc. Jpn. (E) **92**, 2577–2585.
Malinen, J., and Staffans, O. J. (**2006**). "Conservative boundary control systems," J. Differ. Equations **231**,

290–312.
Malinen, J., and Staffans, O. J. (**2007**). "Impedance passive and conservative boundary control systems," Compl. Anal. Oper. Theory **1**, 279–300.
Malinen, J., Staffans, O. J., and Weiss, G. (**2006**). "When is a linear system conservative?," Q. Appl. Math. **64**, 31–91.
Mullen, J., Howard, D., and Murphy, D. (**2006**). "Waveguide physical modeling of vocal tract acoustics: Flexible formant bandwidth control from increased model dimensionality," IEEE Trans. Speech Audio Process. **14**, 964–971.
Niikawa, T., Ando, T., and Matsumura, M. (**2002**). "Frequency dependence of vocal-tract length," in *Proceedings of the Seventh International Conference on Spoken Language Processing*, Denver, Colorado, Sept. 16–20, pp. 1525–1528.
Nishimoto, H., Akagi, M., Kitamura, T., and Suzuki, N. (**2004**). "Estimation of transfer function of vocal tract extracted from MRI data by FEM," in *The 18th International Congress on Acoustics*, Vol. **II**, Kyoto, Japan, pp. 1473–1476.
Palo, P. (**2006**). "A review of articulatory speech synthesis," Master's thesis, TKK, Helsinki.
Saad, Y. (**1992**). *Numerical Methods for Large Eigenvalue Problems* (Manchester University Press, Manchester).
Sasaki, K., Miki, N., and Miyanaga, Y. (**2003**). "FEM analysis based on 3-D time-varying vocal tract shape," in EUROSPEECH-2003, Geneva, Switzerland, Sept. 1–4, pp. 2357–2360.
Sondhi, M. M. (**1986**). "Resonances of a bent vocal tract," J. Acoust. Soc. Am. **79**, 1113–1116.
Suzuki, H., Nakai, T., Takahashi, N., and Ishida, A. (**1993**). "Simulation of vocal tract with three-dimensional finite element method," Tech. Rep. IEICE **EA93-8**, pp. 17–24.
Švancara, P., and Horáček, J. (**2006**). "Numerical modelling of effect of tonsillectomy on production of Czech vowels," Acta. Acust. Acust. **92**, 681–688.
Švancara, P., Horáček, J., and Pešek, L. (**2004**). "Numerical modelling of production of Czech wovel /a/ based on FE model of the vocal tract," in *Proceedings of International Conference on Voice Physiology and Biomechanics*, Marseille, France, Aug. 18–20.

The time-domain signature of negative acoustic group velocity in microsphere suspensions

Joel Mobley
Department of Physics and Astronomy, National Center for Physical Acoustics, University of Mississippi, University, Mississippi 38677
jmobley@olemiss.edu

Abstract: In the wake of recent reports of superluminal acoustic group velocities in sonic and ultrasonic regions of the acoustic spectrum, this paper describes the time-domain manifestation of such group velocities through simulations of the linear propagation of ultrasonic wave packets in a suspension of elastic microspheres. Conditions under which arbitrarily large and negative group velocities can be observed as the speed of a peak in the envelope of an acoustic pulse are described. Propagation simulations demonstrate the physical signature of negative group velocities, as well as the causal compliance of the superluminal acoustic pulses examined in this work.

PACS numbers: 43.35.Bf, 43.20.Hq, 43.20.Bi [TM]
Date Received: March 16, 2007 **Date Accepted:** April 27, 2007

1. Introduction

In highly dispersive media, the propagation of broadband ultrasonic pulses is marked by the severe reshaping of the pulse envelope. In describing waves undergoing dispersive propagation, up to five different velocities (i.e., phase, group, signal, front, and energy) have been defined, with each characterizing a specific aspect of the evolving pulse.[1] In specific situations, one or more of these velocities may not be clearly defined or physically meaningful. The most fundamental of these quantities is the phase velocity c_p. The phase velocity and attenuation coefficient are the primary components of the complex wave number, which when known for all Fourier components of a signal, can be used to predict the evolution of a pulse in a dispersive linear medium. (Complete knowledge of c_p also implies full knowledge of the complex wave number through Kramers-Kronig relations.) The group velocity c_g can almost always be defined, although its direct physical significance or utility is not always apparent. The group velocity does have some intriguing properties, not the least of which is the fact that it can take on both arbitrarily large and negative values. (According to convention, group velocities such that $c_g \gg c_p$, or $c_g < 0$ are collectively labeled "abnormal".) This paper examines the abnormal regime of the group velocity spectrum in the context of time-localized ultrasonic waves propagating in suspensions of elastic microspheres. The goals of this work are to demonstrate the laboratory signature of negative acoustic group velocity through simulations of wave packet propagation through this two-phase medium, where the solid polymer spheres and the liquid host (saline) have negligible viscous losses and intrinsic attenuation. The transport properties of the microsphere suspension are built upon solutions for acoustic scattering from a sphere in liquid of Faran[2,3] together with the multiple scattering wave number of Waterman and Truell.[4] This suspension model has been shown to accurately describe the attenuation, phase velocity, and backscattering properties of agitated polymer microspheres in saline in the low-MHz (3–30 MHz) ultrasound band[5,6] at solids volume concentrations around 1%. Recently, negative group velocities have been measured in such a suspension at volume fractions of 3%.[7] Simulations of ultrasonic pulse propagation in the suspension demonstrate how negative group velocities are physically manifested in the laboratory and the compatibility of such velocities with causality.

Abnormal group velocities are a general feature of wave propagation in dispersive media. The abnormal regime has been most fully explored with electromagnetic waves, chiefly

in the microwave and optical regions of the spectrum. The idea that electromagnetic group velocities could exceed the speed of light in vacuum, c, was appreciated in the early 20th century.[1] However, it was not clear until many decades later that superluminal group velocities (i.e., $c_g > c$, or $c_g < 0$) do have a physical manifestation as a pulse velocity[8] and that such velocities are consistent with relativistic causality. The term superluminal implies that the time interval marking the appearance of some point-of-reference at two distinct locations is less than the time it would take light to go the same distance in vacuum. This definition naturally includes negative velocities since they describe an "object" that goes a positive distance in a travel time that is less than zero. Since the initial observation for propagation in a linear medium,[9] electromagnetic and electronic pulses with superluminal group velocities have been detected in a variety of systems.[10–13] It has been previously recognized that superluminal acoustic velocities are possible in dispersive media,[14,15] and recently measurements of superluminal velocities of acoustic waves have been reported for sound through an air-filled acoustic filter[16] and ultrasound in an aqueous suspension.[7] It should be noted that the signal velocity (i.e., speed of information) is never superluminal, and in this work the leading edge of the pulse is shown to move at an ordinary acoustic speed.

2. Theory

The dispersive propagation of an acoustic wave packet in a passive linear medium can be described by a transfer function of the form, $H(\omega,\Delta z)=\exp[iK(\omega)\Delta z]$, where $K(\omega)=i\alpha(\omega)+\omega/c_p(\omega)$ is the complex wave number, ω $(=2\pi f)$ is the frequency, and Δz is depth of the observation plane in the medium (e.g., position of a microphone or transducer). The complex wave number describes the dispersive properties of the system and is defined in terms of the phase velocity $c_p(\omega)$ and attenuation coefficient $\alpha(\omega)$.[17] If $\mathrm{Re}\,K(\omega)$ is expanded in a Taylor series, the first-order expansion coefficient defines the group velocity

$$\frac{1}{c_g(\omega)} \equiv \frac{d}{d\omega}\frac{\omega}{c_p(\omega)} = \frac{1}{c_p(\omega)}\left(1 - \frac{\omega}{c_p(\omega)}\frac{dc_p(\omega)}{d\omega}\right). \tag{1}$$

For time-localized signals it has been shown that $c_g(\omega)$ is the velocity of the envelope of a sufficiently narrowband Gaussian-gated continuous wave.[8] Numerical work suggests that this physical manifestation of $c_g(\omega)$ as the speed of a peak holds for many singly peaked, smoothly enveloped (e.g., Hann windowed) narrowband signals. The key condition is that the bandwidth of the signal must be sufficiently narrow that the wave number can be accurately represented by its second-order Taylor expansion over that frequency range.

Derived from Eq. (1), the conditions for the existence of superluminal and negative group velocities in dispersive media are

$$\frac{dc_p}{d\omega} > \left(1 - \frac{c_p}{c}\right)\frac{c_p}{\omega} \quad \text{for } c_g(\omega) > c, \tag{2a}$$

$$\frac{dc_p}{d\omega} > \frac{c_p}{\omega} \quad \text{for } c_g(\omega) < 0. \tag{2b}$$

For acoustic waves in water, the ratio of the phase velocity to the speed of light $c_p(\omega)/c$ is on the order of 5×10^{-6} and so Eqs. (2a) and (2b) are practically equivalent conditions for the time and length scales of interest here. The remainder of the paper is focused on the negative velocity portion of the superluminal regime. Note that in Eq. (1) the relations specifically define the inverse group velocity $s_g(\omega)\equiv 1/c_g(\omega)$, also known as the slowness. This quantity is equivalent to the group delay time per unit depth of propagation and will be referred to from here on simply as the group delay. The delay is a more natural parameter for this subject since it passes from positive to negative smoothly through zero (as opposed to velocity which passes through infinity) and its magnitude varies directly with the effect of interest (as opposed to speed which gets smaller as the delay grows more negative). Note that a delay of zero implies that some reference

point in the acoustic wave form reaches all depths at the same time, and negative delays imply that a reference point in the waveform reaches a more distant observer before an observer closer to the source. In this work, the reference point is the peak of the signal envelope.

The transport model for the microsphere suspensions is built around Faran's solution[2,3] for the scattering amplitude of elastic spheres in an inviscid fluid

$$f(\theta,\omega) = -i\frac{c_w}{\omega}\sum_{n=0}^{\infty}(2n+1)D_n(\omega)P_n(\cos(\theta)), \tag{3}$$

where c_w is the velocity in the fluid. The set of coefficients $\{D_n\}$ depends on the wave-number-radius products for the longitudinal mode in the sphere material, for the shear mode in the sphere material, and for the acoustic mode of the fluid, as well as the densities of the fluid and sphere material. With knowledge of the distribution of scatterer sizes, the acoustic complex wave number for a suspension of these spheres can be determined. Using Eq. (3) and a distribution of sphere radii, the Waterman-Truell multiple-scattering wave number[4] for the coherently propagating wave in the suspension takes the form

$$K(\omega) = \frac{\omega}{c_w} \times \sqrt{\left(1 - i\sigma\sum_q \varepsilon_q \frac{3c_w^2}{2a_q^3}\sum_{n=0}^{\infty}(2n+1)\frac{D_{n,q}(\omega)}{\omega^2}\right)^2 + \left(\sigma\sum_q \varepsilon_q \frac{3c_w^2}{2a_q^3}\sum_{n=0}^{\infty}(-1)^n(2n+1)\frac{D_{n,q}(\omega)}{\omega^2}\right)^2}, \tag{4}$$

where $\sigma\varepsilon_q$ is the volume concentration of spheres of radius a_q. The quantity $\sigma \equiv \Sigma_q \sigma_q$ is the total volume fraction of spheres and $\varepsilon_q \equiv \sigma_q/\sigma$ is the normalized volume fraction for each sphere radius. [Note that the Waterman-Truell result is formulated in terms of far-field scattering amplitudes but it is derived for scatterers of any separation, that is, there is no restriction on the minimum separation of the scatterers. The appearance of the far-field amplitude in the wave number is a result of the derivation and not an assumption. This result has recently been rederived with another approach.[18] Experimentally, the Waterman-Truell result has been shown to describe phase velocity accurately at volume fractions beyond those of interest in this work (e.g., see Ref. 19)].

The microsphere size distribution used here was experimentally determined as part of a previous study.[6] The distribution has a volume-weighted mean radius of 50.5 μm and a full-width at half maximum of 4.8 μm. All further references to a suspension refer to one with the same normalized distribution of sphere radii $\{\varepsilon_q\}$.

Spectral bands that can support negative delays occur where $dc_p(\omega)/d\omega > 0$ (known as regions of anomalous dispersion in optics). Figure 1(a) shows the group delay for three volume fractions. The suspension has three bands for which negative delays are possible. The attenuation coefficient and phase velocity of the suspension are shown in Figs. 1(b) and 1(c). The three bands defined by $dc_p(\omega)/d\omega > 0$ are identified by the pairs of vertical lines in each plot. For the first band, a volume fraction of 10% is sufficient to push the majority of the group delay curve into the negative regime. In Fig. 1(d), the minimum group delay is plotted as a function of volume fraction (occurring between 7.1 and 7.3 MHz).

3. Results

For a demonstration of the direct observation of negative group delay, consider the simulated immersion tank experiment illustrated in Fig. 2. The suspension is separated from the water by a thin planar barrier that is acoustically transparent. Two transmitting transducers are placed side-by-side in the water, each facing normal to the water-suspension interface. Each transmitter is axially aligned with a receiving transducer in the suspension. The two receiving transducers are placed at different depths in the suspension relative to the boundary. The transmitters

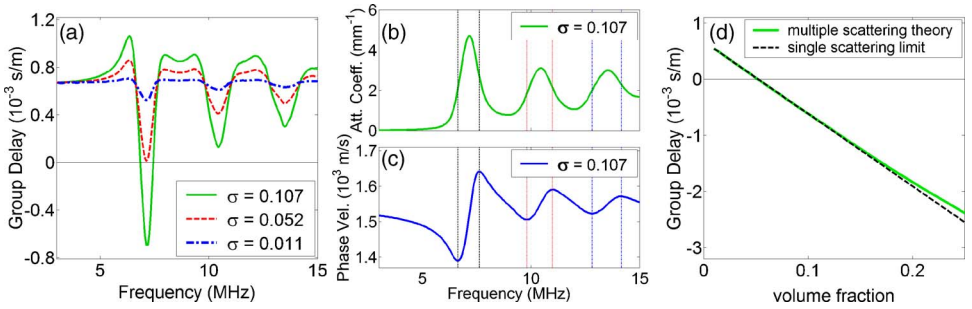


Fig. 1. (Color online) (a) Group delay curves for an experimentally measured system (σ=0.011), and the simulated suspension exhibiting negative delay times for the threshold volume fraction (σ=0.052) and the volume fraction used in the propagation simulations (σ=0.107). The (b) attenuation coefficient [Im $K(\omega)$] and (c) phase velocity [ω/Re $K(\omega)$] for the suspension. The paired vertical lines show the bands where abnormal c_g's are possible. (d) The minimum group delays as functions of σ for the full multiple scattering wave number (7.1–7.3 MHz) and the single scattering limit of the wave number (7.1 MHz).

generate identical pulses, and the only differences in the received pulses are due to the dispersion over the extra depth of travel to the distal (i.e., more distant) receiver. The signal captured by each receiving transducer is displayed by an oscilloscope. The initial signal from each transmitter is a continuous wave gated by a Gaussian window. The Gaussian window is cut off symmetrically at both ends to isolate the signal in time, and this truncated Gaussian curve is downshifted in amplitude so it goes to zero more smoothly. The propagation is simulated using a Fourier domain approach that utilizes the transfer function with the acoustic complex wave number of (4). The magnitudes of the analytic signal for each waveform are calculated and used for the signal envelopes. The volume fraction of spheres is 0.107 and the frequency used is 7.14 MHz, where the greatest negative delay occurs. The group delay curve for this volume fraction is shown in Fig. 1(a). Figure 3 shows the results of the simulations with the two receiving transducers offset by 0.30 mm at various depths in the suspension. In Fig. 3(a) the proximal (i.e., shallower) transducer is just inside the suspension. In Figs. 3(b)–3(f) the depth of the proximal transducer ranges from 0.3 to 1.8 mm.

The propagated signals shown in Fig. 3 consist of two competing elements, the primary signal, which propagates with a negative group delay, and the transient response. At zero depth, the signal is the primary Gaussian-enveloped continuous wave described earlier [see the proximal signal in Fig. 3(a) for its shape]. The frequencies at which negative delays occur for the 0.107 volume fraction are also strongly attenuated; thus the amplitude of the primary signal is sharply reduced with depth as most of its Fourier energy falls within a strong attenuation band. The transient signals are generated at the front and back edges of the pulse due to the nonanalyticity at those points. The component frequencies of the transients are largely those

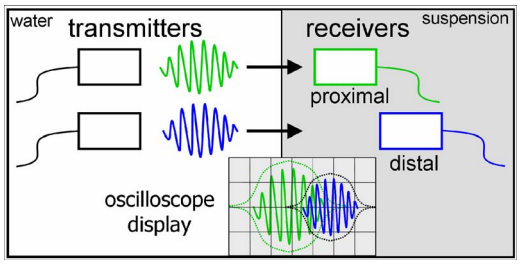


Fig. 2. (Color online) An experimental setup for observing negative group delays in the laboratory. Pulses that are initially identical are propagated between the two pairs of aligned transducers and the captured time-domain waveforms are displayed by an oscilloscope.

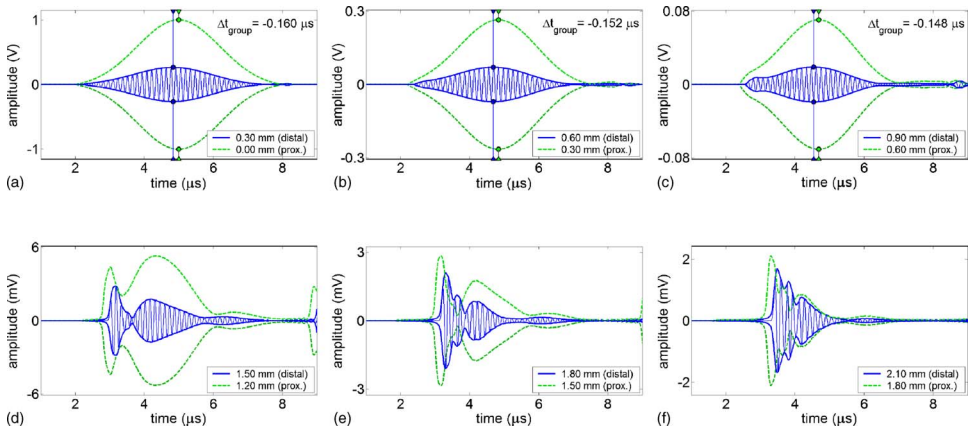


Fig. 3. (Color online) Comparison of waveforms captured at progressive depths in the suspension for σ=0.107. In each panel the receivers are separated by 0.30 mm. For the proximal (shallower) receiver only the signal envelope is shown. In (a)–(c) the positions of the peaks of the envelopes are indicated by vertical lines and the group delay is negative in each. In (d)–(f) the signal is in the transient dominated phase where the Gaussian signals are strongly attenuated and the edge transients become the major feature of the wave forms.

outside the attenuation bands and so are much less attenuated with depth than the primary signal. At shallow depths, the primary signal is the dominant feature in the captured waveform, but as the receivers move deeper the transient responses will begin to dominate as they suffer relatively little attenuation. The primary and the transients also move in different directions in time as a function of depth. With increasing depth, the front edge transient effectively sweeps away the remaining primary signal; at that point and beyond, the waveform delays are positive and the negatively propagating groups are no longer observed. In this manner, causality is maintained.

In order to clearly visualize the laboratory signatures of negative group velocities, several movies of propagation simulations (Mm. 1–5) are provided. All of these simulations refer to the experimental setup described earlier and shown in Fig. 2. For the first set of simulations, both of the transducers remain a fixed distance apart and they are both moved progressively deeper into the suspension with each frame of the movie. The wave forms are those captured by the two receiving transducers. The magnitudes of the analytic signal for each wave form are calculated and used for the signal envelopes. For the proximal receiver (shown in green in the simulations) only the signal envelope is shown so that the wave form of the distal receiver can be clearly visualized. In Mm. 1, the case of a "normal" positive valued group velocity (i.e., $c_g > 0, \sim c_p$) is shown in order to provide a point of comparison. As the transducers move deeper into the suspension, the time ordering of the two signals maintains the usual relationship, with the signal from the deeper transducer always occurring later in time. Mm. 2, however, shows the simulation for a negative group velocity. Now the peak of the signal from the distal transducer occurs earlier in time until it is attenuated away, leaving only the transient signals from the pulse edges. These transient signals keep the "normal" time-ordering with those from the proximal transducer occurring earlier.

Mm. 1. The waveforms from the simulated experiment at f_{center}=6.35 MHz where c_g is "normal" (avi, 3.7 MB).

Mm. 2. The waveforms from the simulated experiment at f_{center}=7.14 MHz where c_g is negative. (avi, 3.1 MB).

A variation on this experiment is to keep one transducer fixed in depth just inside the suspension while continually moving the other receiver deeper into the suspension. The results of these "fixed-reference" simulations for the normal group velocity are shown in Mm. 3, and the negative group velocity case is shown in Mm. 4 and Mm. 5.

Mm. 3. The waveforms from the simulated experiment at $f_{center}=6.35$ MHz where c_g is "normal." (avi, 3.3 MB).

Mm. 4. The waveforms from the simulated experiment at $f_{center}=7.14$ MHz where c_g is negative. (avi, 1.5 MB).

Mm. 5. The waveforms from the simulated experiment at $f_{center}=7.14$ MHz when c_g is negative. Here the reference wave is autoscaled so both signals fill the window. The signals are the same as Mm. 4. (avi, 3.5 MB).

4. Discussion

The compatibility between negative group delays and causality has been widely noted, and is commonly attributed to the "pulse reshaping" effect. The causal consistency can also be verified using Kramers-Kronig relations between group velocity and attenuation[20,21] which stem from the strict ordering of cause and effect. The causal nature of the processes observed in this work is illustrated by the signals produced in the simulations, as in Fig. 3, Mm. 2, 4, and 5. As a receiver is pulled deeper into the suspension, the front edge transient signal moves later in time (positive delay) while the peak of the initially Gaussian primary signal envelope moves forward in time (negative delay). However, the primary envelope never moves through the leading transient to earlier times; rather it is attenuated beyond recognition before its peak can reach the leading edge of the signal. The leading edge transient, which propagates with a "normal" acoustic velocity, reaches all depths first and so its velocity is effectively the signal velocity. If the primary envelope were somehow able to remain intact and pass through the leading transient to earlier times, then there might be something extraordinary at work. However, the envelope effectively rides on top of an ultrasonic tone burst moving at an "ordinary" acoustic velocity, and if parts of the envelope outrun this support structure, they cease to exist.

In summary, the physical signature of negative acoustic group velocity in a suspension of plastic microspheres has been shown through propagation simulations. The negative velocity of the waveform envelope is countered by the positive velocity of the transients at the edges of the pulse. Beyond a sufficient depth, the transients come to dominate the signal as attenuation effectively extinguishes the original pulse. The negative velocity group never overtakes the leading edge transient and causal propagation is maintained.

References and links

[1] L. Brillouin, *Wave Propagation and Group Velocity* (Academic, New York, 1960).
[2] J. J. Faran, "Sound scattering by solid cylinders and spheres," J. Acoust. Soc. Am. **23**, 405–418 (1951).
[3] R. Hickling, "Analysis of echoes from a solid elastic sphere in water," J. Acoust. Soc. Am. **34**, 1482–1592 (1962).
[4] P. C. Waterman and R. Truell, "Multiple scattering of waves," J. Math. Phys. **2**, 512–537 (1961).
[5] C. S. Hall, J. N. Marsh, M. S. Hughes, J. Mobley, K. D. Wallace, J. G. Miller, and G. H. Brandenburger, "Broadband measurements of the attenuation coefficient and backscatter coefficient for suspensions: A potential calibration tool," J. Acoust. Soc. Am. **101**, 1162–1171 (1997).
[6] J. Mobley, K. R. Waters, C. S. Hall, J. N. Marsh, M. S. Hughes, G. H. Brandenburger, and J. G. Miller, "Measurements and predictions of the phase velocity and attenuation coefficient in suspensions of elastic microspheres," J. Acoust. Soc. Am. **106**, 652–659 (1999).
[7] J. Mobley and R. E. Heithaus, "Ultrasonic properties of a suspension of microspheres supporting negative group velocities," Phys. Rev. Lett. (submitted).
[8] C. G. B. Garrett and D. E. McCumber, "Propagation of a Gaussian light pulse through an anomalous dispersion medium," Phys. Rev. A **1**, 305–313 (1970).
[9] S. Chu and S. Wong, "Linear pulse propagation in an absorbing medium," Phys. Rev. Lett. **48**, 738–741 (1982).
[10] R. Y. Chiao and A. M. Steinberg, "Tunneling times and superluminality," Prog. Opt. **37**, 347–406 (1997).
[11] A. Dogariu, A. Kuzmich, and L. J. Wang, "Transparent anomalous dispersion and superluminal light-pulse propagation at a negative group velocity," Phys. Rev. A **63**, 053806 (2001).
[12] M. Mojahedi, K. J. Malloy, G. V. Eleftheriades, J. Woodley, and R. Y. Chiao, "Abnormal wave propagation in passive media," IEEE J. Sel. Top. Quantum Electron. **9**, 30–39 (2003).
[13] E. Recami, M. Zamboni-Rached, K. Z. Nobrega, C. A. Dartora, and H. E. Hernandez, "On the localized superluminal solutions to the Maxwell equations," IEEE J. Sel. Top. Quantum Electron. **9**, 59–73 (2003).
[14] W. M. Robertson, J. Ash, and J. M. McGaugh, "Breaking the sound barrier: Tunneling of acoustic waves through the forbidden transmission region of a one-dimensional acoustic band gap array," Am. J. Phys. **70**, 689–693 (2002).

[15]S. Yang, J. H. Page, Z. Liu, M. L. Cowan, C. T. Chan, and P. Sheng, 'Ultrasound tunneling through 3D phononic crystals," Phys. Rev. Lett. **88**, 104301 (2002).
[16]W. M. Robertson, J. Pappafotis, P. Flannigan, J. Cathey, B. Cathey, and C. Klaus, "Sound beyond the speed of light: Measurement of negative group velocity in an acoustic loop filter," Appl. Phys. Lett. **90**, 014102 (2007).
[17]The wave number convention used here differs from that of Refs. 2 and 6. Both Refs. 2 and 6 use $e^{-i(kz-\omega t)}$ for forward propagation while $e^{-i(\omega t-kz)}$ is the physics convention that is used in this work. The relevance here is that the expansion coefficients $\{D_n\}$ used in this work are the complex conjugates of those in Ref. 6.
[18]Y. C. Angel and C. Aristégui, "Analysis of sound propagation in a fluid through a screen of scatterers," J. Acoust. Soc. Am. **118**, 72–82 (2005).
[19]C. Layman, N. S. Murthy, R. B. Yang, and J. Wu, "The interaction of ultrasound with particulate composites," J. Acoust. Soc. Am. **119**, 1449–1456 (2006).
[20]J. Mobley, K. R. Waters, and J. G. Miller, "Causal determination of acoustic group velocity and frequency derivative of attenuation with finite-bandwidth Kramers-Kronig relations," Phys. Rev. E **72**, 016604 (2005).
[21]J. Mobley, "Finite-bandwidth Kramers-Kronig relations for acoustic group velocity and attenuation derivative applied to encapsulated microbubble suspensions," J. Acoust. Soc. Am. **121**, 1916–1923 (2007).

Testing theories of vowel inherent spectral change[a)]

Geoffrey Stewart Morrison[b)] and Terrance M. Nearey
Department of Linguistics, University of Alberta, Edmonton, Alberta, T6G 2E7, Canada
gsm2@bu.edu, t.nearey@ualberta.ca

Abstract: Three competing accounts of vowel inherent spectral change in English all agree on the importance of initial formant frequencies; however, they disagree about the nature of the perceptually relevant aspects of formant change. The onset+offset hypothesis claims that the final formant values themselves matter. The onset+slope hypothesis claims that only the rate of change counts. The onset+direction hypothesis claims that only the general direction of change in formant frequencies is important. A synthetic-vowel perception experiment was designed to differentiate among the three. Results provide support for the superiority of the onset+offset hypothesis.

PACS numbers: 43.71.−k, 43.71.An, 43.71.Es [JH]
Date Received: February 23, 2007 **Date Accepted:** April 17, 2007

1. Introduction

Traditionally, the English vowel system is often said to comprise true diphthongs, /aɪ, aʊ, ɔɪ/, phonetic diphthongs, /e, o/ (frequently transcribed as /eɪ, oʊ/), and monophthongs, e.g., /i, ɪ, ɛ, æ/. However, for many North American dialects, it has been reported that several nominal monophthongs show significant *vowel inherent spectral change* (VISC), and VISC appears to be important for perception (Andruski and Nearey, 1992; Assmann and Katz, 2005; Assmann *et al.*, 1982; Hillenbrand *et al.*, 2001). Figure 1 provides examples of mean F1-F2 formant trajectories for natural productions of Western Canadian English /e/, /ɪ/, and /ɛ/ in the same context as used in the present study (data from Morrison, 2006).

There are three main accounts of the perceptually relevant aspects of VISC (Gottfried *et al.*, 1993; Nearey and Assmann, 1986). All three hypotheses agree that the initial formant frequencies are perceptually relevant for vowel identification, but disagree on what additional cues are relevant. The *onset+offset* hypothesis states that the relevant perceptual cues are the formant values at the end of the vowel. The *onset+slope* hypothesis states that the relevant cue is the rate of change of formants over time. The *onset+direction* hypothesis states that the only relevant factor is the direction of formant movement in an F1-F2 (or similar) space. To elucidate the difference among the three hypotheses: In the direction hypothesis the rate of change in time (hereafter speed) of formant movement and the formant values achieved at the end of the trajectory are irrelevant. In the slope hypothesis the direction and speed of formant movement are relevant, but the formant values achieved at the end of the trajectory are irrelevant. If the glide portion of one vowel is longer than the glide portion of another vowel, then they could both have the same formant slopes but different final formant values. In the offset hypothesis the formant values achieved at the end of the trajectory are relevant (i.e., direction and magnitude of formant movement are relevant), but the speed of movement toward those values is irrelevant. If the glide portion of one vowel is longer than the glide portion of another vowel, then they could both have the same final formant values but different formant slopes.

[a)]Portions of this letter were presented at the 150th Meeting of the ASA, Minneapolis, October 2005.
[b)]Present address: Department of Cognitive & Neural Systems, Boston University, 677 Beacon Street, Boston, Massachusetts 02215.

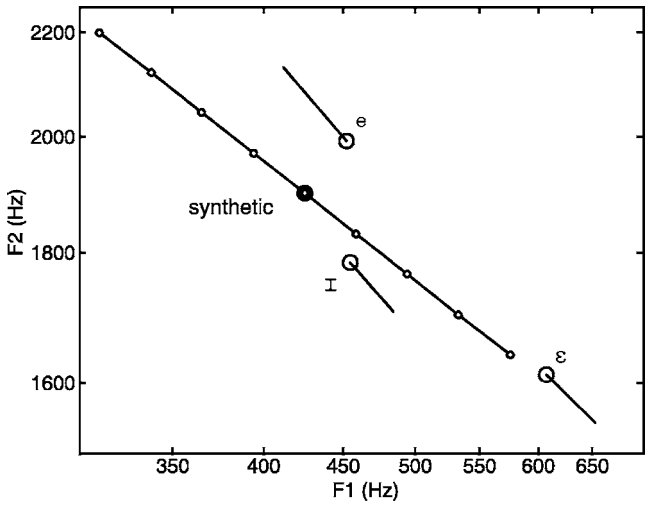


Fig. 1. First and second formant properties of natural and synthetic vowels. The comet plots labeled /e/, /ɪ/, and /ɛ/ represent the mean formant onsets and offsets of these vowels produced in isolated-word /bVpə/ context by seven male speakers of Western Canadian English (ten replications per speaker). Comet heads and tails represent the formant values at 25% and 75% of the duration of the vowels, respectively. (Geometric mean vowel durations for /e/, /ɪ/, and /ɛ/ were 114, 69, and 82 ms, respectively.) For the synthetic stimuli, the large black dot represents the initial formant values and the small white dots represent the nine sets of final formant values: Four magnitudes of diverging VISC, flat (zero) VISC, and four magnitudes of converging VISC.

Nearey and Assmann (1986) and Gottfried *et al.* (1993) used pattern recognition models to test alternate parametrizations consistent with the three VISC hypotheses. In both studies, although there was a slight advantage for the offset model, all three parametrizations performed well in terms of correct identification (and, in the earlier study, correlation with listeners' response patterns).[1] In contrast to the former studies, the stimuli in the present experiment are explicitly designed to differentiate among the three hypotheses. Specifically, the hypotheses are directly compared in a perceptual experiment using a synthetic /e/-/ɪ/-/ɛ/ continuum which has a fixed onset, but which varies in offset, duration, and slope.

2. Method

2.1 Stimuli

Stimuli were synthesized in a male-speaker range using an implementation of the Klatt cascade formant synthesizer (Klatt and Klatt, 1990). The stimuli were /bVpə/, consisting of synthetic /bVp¯/ plus a natural recording of the final /pə/ syllable from the burst onward.[2] Care was taken (adjusting glottal slope and breathiness parameters) to match the voice quality of the synthetic speech to the natural portion of the stimuli. To limit the stimulus space, pilot studies were conducted to find a single initial set of F1-F2 values from which it was possible to obtain /e/, /ɪ/, and /ɛ/ percepts by changing only the final formant values and duration. The initial F1 and F2 values were 425 and 1900 Hz, respectively (6.052 and 7.550 log Hz), see Fig. 1.

Three types of vowels were synthesized: Formant trajectories were either *straight* (in log hertz) from the beginning to the end of the vowel [Figs. 2(a), 2(c), and 2(e)]; or they were *elbowed* with an initial steady state for the first 25% of the vowel duration, followed by a glide [Figs. 2(b), 2(d), and 2(f)]. Elbowed diphthongs thus had the same initial and final formant values as straight diphthongs, but the slope of the glide portion was 33% steeper. The third vowel type was *flat*, having no formant movement during the vowel.

Each stimulus had one of three directions of movement. The flat stimuli had zero movement. The other two directions were opposite in the log F1-F2 space (see Fig. 1), stimuli either had converging VISC [F1 rose and F2 fell, see Figs. 2(c)–2(f)], or diverging VISC [F1 fell

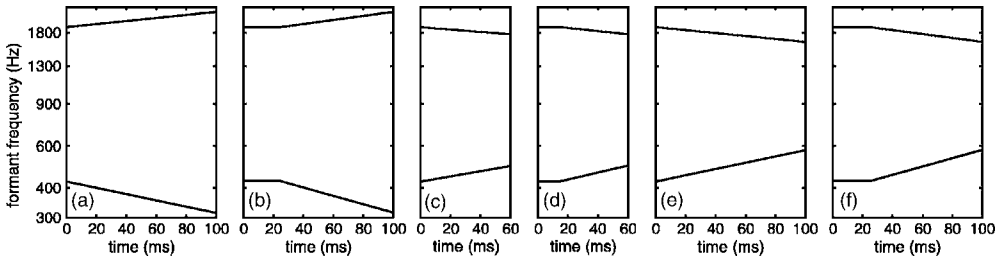


Fig. 2. Examples of formant trajectories for synthetic stimuli (excluding consonant transitions). The examples are straight [(a), (c), (e)] and elbowed [(b), (d), (f)] versions of the stimuli which, in the perception experiment, received the greatest number of /e/ (a), (b), /ɪ/ (c), (d), and /ɛ/ (e), (f) responses. The percentages of responses (pooled across listeners) for each category were: (a) 85 /e/, 3 /ɪ/, 12 /ɛ/. (b) 79 /e/, 4 /ɪ/, 17 /ɛ/. (c) 21 /e/, 60 /ɪ/, 20 /ɛ/. (d) 22 /e/, 53 /ɪ/, 26 /ɛ/. (e) 23 /e/, 35 /ɪ/, 41 /ɛ/. (f) 25 /e/, 31 /ɪ/, 44 /ɛ/. Audio files Mm. 1 through 6 correspond to (a)–(f).

and F2 rose, see Figs. 2(a) and 2(b)].[3] Nonflat stimuli had one of four magnitudes of VISC (multiples of ± 0.0756 log Hz for F1 and ∓ 0.0367 log Hz for F2). Because natural productions of the three vowels examined here differ in duration as well as spectral properties, stimuli also ranged over three duration values: 60, 80, and 100 ms (excluding consonant transitions). This resulted in a total of 51 stimuli: Eight straight, eight elbowed, and one flat, all multiplied by three durations.

Mm. 1. Audio file of stimulus shown in Fig. 2(a) (8.03kB). This file is of type "wav."

Mm. 2. Audio file of stimulus shown in Fig. 2(b) (8.03kB). This file is of type "wav."

Mm. 3. Audio file of stimulus shown in Fig. 2(c) (7.25kB). This file is of type "wav."

Mm. 4. Audio file of stimulus shown in Fig. 2(d) (7.25kB). This file is of type "wav."

Mm. 5. Audio file of stimulus shown in Fig. 2(e) (8.03kB). This file is of type "wav."

Mm. 6. Audio file of stimulus shown in Fig. 2(f) (8.03kB). This file is of type "wav."

2.2 Listeners

Listeners were 23 undergraduate-student volunteers. They were monolingual English speakers who had grown up in western Canada, and reported no hearing deficits.

2.3 Procedures

Listeners were tested one at a time in a sound booth. The stimuli were played at a comfortable volume (approximately 65 dB SPL) via a Roland Edirol UA-30 sound card and a calibrated Mackie HR824 studio monitor. In each trial, the listeners heard a stimulus, then saw three buttons on a computer screen labeled *baypa, bippa*, and *beppa* representing /e/, /ɪ/, and /ɛ/, respectively (prior to the experiment, listeners were trained on the orthography to vowel-category relationship). They indicated their response via a mouse click. A new stimulus was automatically presented 500 ms after a response to the previous stimulus was provided. The whole set of stimuli were presented eight times in randomized blocks, resulting in a total of 408 trials per listener.

3. Results and discussion

Direct statistical tests of the three rival hypotheses are not possible because such tests require a strict nesting of terms, and the hypotheses are not related to each other in this manner. An indirect strategy analogous to partial correlation analysis was therefore adopted. The procedure is as follows: Models of two rival hypotheses (e.g., A and B) are fitted to the data, then a third

larger model (e.g., C) including all the terms of the first two models is fitted. Thus, both smaller models are properly nested within the third. If the larger model fits significantly better than A but not significantly better than B, then it is reasonable to infer support for hypothesis B. Roughly speaking, B includes essentially all the relevant information in the larger hypothesis, while adding the extra terms from A adds little explanatory information.

3.1 Offset versus slope

In order to compare the adequacy of the offset versus the slope hypotheses, *Vector smoothed generalized additive logistic regression models* (Yee and Wild, 1996) were fitted to the perceptual response data (the raw counts of responses for each category given to each stimulus), and compared for goodness-of-fit. Three models were fitted to each of the listeners' data sets. In all models, duration (three values) was entered in milliseconds and fitted via a (saturated) quadratic polynomial. Duration parameters were included to control for the effect of duration which is not of interest in the present letter. Other parameters entered into the models related directly to formant movement. These were fitted via smoothing splines with two effective degrees of freedom.[4] Since all three models include initial formant specifications, the differences among them can be summarized by their characterization of spectral change:

(A) *Offset model:* $\Delta F1$, the change of F1 in log Hz from the beginning to the end of the vowel.

(B) *Slope model:* $\Delta F1/\Delta t$, the slope of F1 in log Hz/s over the glide portion of the vowel (from the beginning or elbow of the vowel, for straight and elbowed stimuli, respectively, to the end of the vowel).

(C=A+B) *Combined model:* A model containing both $\Delta F1$ and $\Delta F1/\Delta t$. This allows for both the offset and the slope to have effects on perception.

Parametrizations A and B are equivalent to those of Nearey and Assmann (1986). Since F2 in the synthetic stimuli in the present study was perfectly correlated with F1, it was only necessary to enter one set of formant values into the models.

The model-fitting procedure minimizes the deviance statistic (G^2). Table 1 gives results of ΔG^2 likelihood ratio tests comparing the differences in the goodness-of-fit between the larger model (model C) and each of the two smaller models (models A and B). Models fitted to data from 10 of the 23 listeners had a significant (nominal $p < 0.05$) improvement in goodness-of-fit when the offset parameters were added to a model already containing slope parameters. In contrast, when slope parameters were added to a model already containing offset parameters, there was a significant improvement in goodness-of-fit for data from only three listeners, and these three data sets had also had a significant improvement of fit when the offset parameters were added to the slope model. The results therefore provide greater support for the onset +offset hypothesis than for the onset+slope hypothesis.

3.2 Offset versus direction

If the direction hypothesis is correct, then there should be no difference in listeners' responses to stimuli which have the same VISC direction in the F1-F2 plane but different final formant values. The direction versus offset hypotheses were tested in the same manner as the slope versus offset hypotheses. The formant-movement parameters in the models were:

(A) *Offset model:* $\Delta F1$ (the same as noted earlier).

(D) *Direction model:* The three directions of formant movement in the stimuli, diverging, flat, and converging, were entered as three discrete interval levels, −1, 0, and +1.[5]

(E=A+D) *Combined model:* A model containing both ΔF1 and direction parameters.

Results of the ΔG^2 likelihood ratio tests are given in Table 2. Models fitted to data from 16 of the 23 listeners had a significant (nominal $p < 0.05$) improvement in goodness-of-fit when the offset parameters were added to a model already containing direction parameters. In contrast, when the direction parameters were added to a model already containing offset param-

Table 1. Likelihood ratio tests for combined model (C) vs offset model (A) and slope model (B). ΔG^2 is decrease in deviance statistic (lack of fit measure) in the larger vs smaller model. Δedf is the change in effective degrees of freedom (see footnote 4). Reported p values are nominal significance levels for individual tests referred to χ^2 distributions with Δedf degrees of freedom.

	Models compared					
	C vs B (added terms $\Delta F1$)			C vs A (added terms $\Delta F1/\Delta t$)		
Listener	ΔG^2	Δedf	p	ΔG^2	Δedf	p
1	13.53	3.42	0.005[a]	1.95	3.47	0.666
2	3.77	3.93	0.427	5.38	3.97	0.247
3	11.17	3.94	0.024[a]	4.49	3.81	0.318
4	13.09	4.03	0.011[a]	12.38	3.83	0.013[a]
5	3.39	3.49	0.414	2.56	3.51	0.556
6	5.35	3.68	0.217	3.91	3.72	0.376
7	4.68	3.90	0.308	6.46	3.77	0.148
8	8.62	3.78	0.062	2.98	3.82	0.533
9	3.01	3.71	0.510	3.00	3.75	0.517
10	4.98	4.04	0.295	3.89	3.98	0.418
11	13.05	3.57	0.008[a]	11.54	3.75	0.017[a]
12	30.00	4.00	0.000[a]	12.33	3.68	0.012[a]
13	8.17	3.99	0.085	5.51	3.87	0.225
14	13.86	3.63	0.006[a]	2.98	3.73	0.518
15	4.73	3.91	0.305	1.52	3.90	0.812
16	11.76	3.95	0.019[a]	2.44	3.90	0.640
17	4.14	3.93	0.377	2.97	3.86	0.540
18	9.47	3.94	0.048[a]	2.17	3.76	0.670
19	14.10	3.99	0.007[a]	3.40	3.91	0.478
20	4.45	3.77	0.318	4.08	3.70	0.352
21	7.49	3.91	0.107	1.25	3.80	0.850
22	12.86	3.76	0.010[a]	1.61	3.50	0.740
23	7.68	3.66	0.085	5.62	3.77	0.206

[a]Significant at a nominal α level of 0.05.

eters, there was a significant improvement in goodness-of-fit for data from only one listener. The results therefore provide much greater support for the offset hypothesis than for the direction hypothesis.

3.3 Remaining issues

There was considerable interlistener variation in response patterns. Also, a number of listeners gave a preponderance of /e/ responses, even for stimuli with converging VISC. Note that interlistener variation and bias do not invalidate the results of the analyses. All that is required for the reasonable application of the modeling procedure is that the response probabilities change noticeably under at least some of the stimulus manipulations at issue. For each listener there was a significant improvement in goodness-of-fit when a model including only duration information was compared with the offset model (model A), this indicates that the requirement was met with respect to formant movement.

All hypotheses agree on the importance of initial formant values, and the onset point in the experiment may have given a strong cue for /e/ perception. Examination of Fig. 1 indicates that the ratio of initial F1 and F2 values were closest to the mean of natural /e/ productions [the

Table 2. Likelihood ratio tests for combined model (E) vs offset model (A) and direction model (D). (See caption of Table 1 for explanation of abbreviations.)

	Models compared					
	E vs D (added terms $\Delta F1$)			E vs A (added terms *direction coding*)		
Listener	ΔG^2	Δedf	p	ΔG^2	Δedf	p
1	26.97	3.57	0.000[a]	6.64	3.98	0.155
2	6.16	3.93	0.181	5.14	3.99	0.273
3	13.45	3.73	0.007[a]	3.93	3.83	0.390
4	11.93	3.85	0.016[a]	4.89	4.01	0.300
5	2.44	3.48	0.573	5.05	4.00	0.282
6	13.07	3.68	0.008[a]	3.19	4.01	0.529
7	5.38	3.91	0.240	9.44	3.99	0.051
8	6.47	3.73	0.144	0.69	4.00	0.952
9	33.78	3.66	0.000[a]	4.66	3.94	0.316
10	13.73	4.03	0.008[a]	7.10	3.99	0.130
11	35.36	3.70	0.000[a]	3.16	4.00	0.531
12	16.98	3.71	0.001[a]	3.05	4.03	0.554
13	3.38	3.95	0.488	6.28	3.99	0.178
14	18.99	3.69	0.001[a]	1.58	4.03	0.815
15	13.66	3.85	0.007[a]	0.57	4.00	0.966
16	14.30	3.86	0.006[a]	5.37	4.00	0.251
17	2.27	3.91	0.672	1.93	4.00	0.750
18	19.77	3.82	0.000[a]	5.90	4.00	0.207
19	12.19	3.93	0.015[a]	5.18	3.99	0.268
20	1.15	3.80	0.868	11.67	4.02	0.020[a]
21	16.33	3.83	0.002[a]	2.66	3.99	0.616
22	26.23	3.64	0.000[a]	6.25	3.81	0.165
23	13.83	3.78	0.007[a]	5.51	4.03	0.242

[a]Significant at a nominal α level of 0.05.

Morrison (2006) data were collected after the present study was conducted]. It is also possible that no single onset point will allow for clear percepts of all three vowels by all listeners, and additional experiments using multiple onset points may be warranted.

Furthermore, it is conceivable that a perceptual mechanism registering rate of spectral change may have operational limits (e.g., temporal smearing) that produce differences in "effective slope" that are less than the nominal 33% differences between corresponding straight and elbowed stimuli. Thus, the distinction between slope and offset hypotheses should probably also be assessed in contexts with longer mean vowel durations (e.g., isolated CVCs), where absolute physical differences between straight and elbowed stimuli would be larger.

Finally, there is perhaps a weak indication that more complex characterizations of formant trajectories should be investigated (see footnote 1 for some possible parametrizations). The combined models C and E contain information not available in their constituents (A+B and A+D, respectively). Three of 23 listeners for model C and one for model E showed significant improvement over both the relevant simpler models. Since at the nominal 0.05 α level, one might expect to find only about two false positives (Type I errors), it would seem judicious to examine more complex trajectory models in a larger and presumably more powerful experiment.

4. Conclusion

The weight of evidence presented here indicates that the onset+offset hypothesis is superior to the onset+slope and onset+direction hypotheses.

Acknowledgments

This research was supported by the Social Sciences and Humanities Research Council of Canada. Thanks to Peter F. Assmann, Michael Kiefte, the Associate Editor, and two anonymous reviewers for comments on earlier drafts of this letter.

References and links

[1]More complex curve-fitting parametrizations of VISC have been tested (Zahorian and Jagharghi, 1993); however, Hillenbrand *et al.* (2001) failed to find an improvement over the onset+offset model for models using polynomials and discrete-cosine transforms. But see the discussion in Sec. 3.3.

[2]This was one of a series of several experiments designed to examine English and Spanish listeners' perception of VISC. The /bVpə/ pattern results in possible but nonexistent words in both languages. In some experiments the /bVpə/ words were embedded in natural-speech carrier sentences in the same voice as the natural-speech /pə/ portion of the stimuli.

[3]The stimulus identification rates reported in the caption of Fig. 2 are based on Western Canadian English listeners' responses, and the reader may categorize these stimuli differently. Also, the reader may have difficulty distinguishing the straight and elbowed versions of each pair of example stimuli (although this was not the case for the younger of the two authors). This might be expected if slope is not a perceptually relevant aspect of VISC.

[4]The fitting procedure implements smoothing splines via a maximum penalized likelihood method. The analyst can request a target degrees-of-freedom value, and the software will seek a smoothness penalty that comes close to that value. See Hastie and Tibshirani (1990); Hastie *et al.* (2001, Chap. 9); and Yee and Wild (1996) for details. Two requested degrees of freedom in the representation of a stimulus property results in (approximately) an added four degrees of freedom in the fitted models because the smoothing splines are estimated independently for each of two nonredundant response contrast terms for the three-way vowel distinction. Similar results were obtained with linear stimulus terms; however, for most listeners there was a significant improvement in goodness-of-fit when offset and slope models were fitted using smoothing splines with two effective degrees of freedom. This modest increase in the complexity of the models maintained all the essential differences among the three main hypotheses examined, and provided some added flexibility to allow for factors such as possible floor and ceiling effects in the stimulus variables in question.

[5]Direction in the stimuli is restricted to one dimension (see Fig. 1), and the three values −1, 0, and +1 exhaust all possible one-dimensional directions. A discrete-level coding system was therefore adopted, with direction dummy coded using two parameters (i.e., [1 0], [0 0], and [0 1] for −1, 0, and +1, respectively). The smallest nonzero $\Delta F2$ magnitudes in the stimuli (68Hz) just exceeded the largest mean threshold for F2 movement detection (66Hz) reported by Kewley-Port and Goodman (2005). The stimuli in the present study also had F1 movement (minimum 31Hz), and the combined effect of F1 and F2 movement was therefore expected to make the difference between zero-magnitude and the smallest non-zero-magnitude stimuli detectable (results of Monte Carlo permutation tests were consistent with this expectation).

Andruski, J. E., and Nearey, T. M. (**1992**). "On the sufficiency of compound target specification of isolated vowels in /bVb/ syllables," J. Acoust. Soc. Am. **91**, 390–410.

Assmann, P. F., and Katz, W. F. (**2005**). "Synthesis fidelity and time-varying spectral change in vowels," J. Acoust. Soc. Am. **117**, 886–895.

Assmann, P. F., Nearey, T. M., and Hogan, J. T. (**1982**). "Vowel identification: Orthographic, perceptual, and acoustic aspects," J. Acoust. Soc. Am. **71**, 975–989.

Gottfried, M., Miller, J. D., and Meyer, D. J. (**1993**). "Three approaches to the classification of American English diphthongs," J. Phonetics **21**, 205–229.

Hastie, T., and Tibshirani, R. (**1990**). *Generalized Additive Models*, Chapman and Hall, London.

Hastie, T., Tibshirani, R., and Friedman, J. (**2001**). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction* (Springer, New York).

Hillenbrand, J. M., Clark, M. J., and Nearey, T. N. (**2001**). "Effect of consonant environment on vowel formant patterns," J. Acoust. Soc. Am. **109**, 748–763.

Kewley-Port, D., and Goodman, S. G. (**2005**). "Thresholds for second formant transitions in front vowels," J. Acoust. Soc. Am. **118**, 3252–3560.

Klatt, D. H., and Klatt, L. C. (**1990**). "Analysis, synthesis, and perception of voice quality variations among female and male talkers," J. Acoust. Soc. Am. **87**, 820–857.

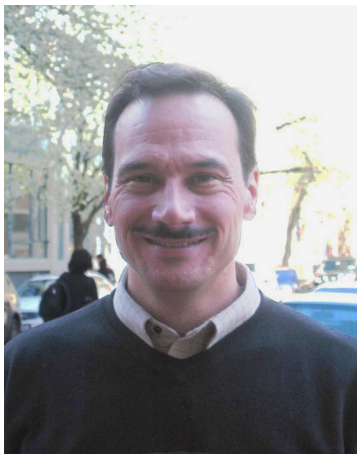
Morrison, G. S. (**2006**). "L1 & L2 production and perception of English and Spanish vowels: A statistical modelling approach," PhD dissertation, University of Alberta, Edmonton, Alberta, Canada.
Nearey, T. M., and Assmann, P. F. (**1986**). "Modeling the role of vowel inherent spectral change in vowel identification," J. Acoust. Soc. Am. **80**, 1297–1308.
Yee, T. W., and Wild, C. J. (**1996**). "Vector generalized additive models," J. R. Stat. Soc. Ser. B (Methodol.) **58**, 481–493.
Zahorian, S., and Jagharghi, A. (**1993**). "Spectral-shape features versus formants as acoustic correlates for vowels," J. Acoust. Soc. Am. **94**, 1966–1982.

ACOUSTICAL NEWS—USA

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

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

New Fellows of the Acoustical Society of America



Paul E. Barbone —For contributions to structural acoustics and acoustical imaging.

Kai Ming Li—For contributions to the study of sound propagation in high-rise cities.

Claire Prada—For contributions to time-reversal acoustic remote sensing.

Karim Sabra—For contributions to the use of ambient noise for imaging.

 0001-4966/2007/122(1)/1/3/$23.00

Accountants' Review Report

To Board of Directors
Acoustical Society Foundation, Inc.

We have reviewed the accompanying statement of financial position of Acoustical Society Foundation, Inc. (a nonprofit organization) as of June 30, 2006 and the related statements of activities and net assets and cash flows for the year then ended, in accordance with Statements on Standards for Accounting and Review Services issued by the American Institute of Certified Public Accountants. All information included in these financial statements is the representation of the management of Acoustical Society Foundation, Inc.

A review consists principally of inquiries of Organization personnel and analytical procedures applied to financial data. It is substantially less in scope than an audit in accordance with auditing standards generally accepted in the United States of America, the objective of which is the expression of an opinion regarding the financial statements taken as a whole. Accordingly, we do not express such an opinion.

Based on our review, we are not aware of any material modifications that should be made to the accompanying financial statements in order for them to be in conformity with generally accepted accounting principles.

Robinson & Gordon, CPA's, P.C.
Rochester, New York
October 23, 2006

ACOUSTICAL SOCIETY FOUNDATION, INC.
STATEMENT OF FINANCIAL POSITION
JUNE 30, 2006

ASSETS

CURRENT ASSETS	
Cash, including pooled income fund cash of $8,611	$ 31,234
Investments	755,699
Pooled income fund securities	167,667
TOTAL ASSETS	$954,600

LIABILITY AND NET ASSETS

LIABILITY	
Deferred revenue	$ 38,166
NET ASSETS	
Unrestricted-operating	388,022
Board Designated:	
Stetson Scholarship Fund for Speech	79,583
Beranek Fund	69,795
Various small funds	20,466
	169,844
Temporarily Restricted—Medwin Prize in Acoustical Oceanography	40,350
Permanently Restricted:	
Pooled Income Fund	138,113
Robert W. Young Memorial Fund	79,705
Rossing Prize for Acoustics Education	100,400
	318,218
TOTAL NET ASSETS	916,434
TOTAL LIABILITY AND NET ASSETS	$954,600

ACOUSTICAL SOCIETY FOUNDATION, INC.
STATEMENT OF CASH FLOWS
FOR THE YEAR ENDED JUNE 30, 2006

CASH FLOWS—OPERATING ACTIVITIES	
Increase in net assets	$51,255
Adjustments to reconcile increase in net assets to net cash provided from operating activities:	
Contributed securities—Investments	(1,100)
Net realized gain on sale of investments	(1,248)
Net realized gain on sale of pooled income fund assets	(578)
Unrealized gain on investment	(12,672)
Unrealized gain on pooled income fund assets	(7,072)
Change in value of split interest agreements	(10,302)
Increase in deferred revenue (liability to life beneficiaries)	6,451
Uncollected pledges receivable	7,025
Collected pledges receivable	500
NET CASH PROVIDED FROM OPERATING ACTIVITIES	32,259
CASH FLOWS—INVESTING ACTIVITIES	
Purchase of securities:	
Investments	(36,359)
Pooled income fund assets	(22,027)
Proceeds from sale of securities:	
Investments	26,385
Pooled income fund assets	18,292
NET CASH USED FOR INVESTING ACTIVITIES	(13,709)
NET INCREASE IN CASH	18,550
Cash at beginning of year	12,684
CASH AT END OF YEAR	$31,234

USA Meetings Calendar

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

2007

24–26 July	Revolutionary Aircraft for Quiet Communities, Hampton, VA [A NASA workshop hosted by the National Institute of Aerospace and co-sponsored by the Joint Planning and Development Office and the Council of European Aerospace Societies; Web: www.nianet.net].
5–8 Oct.	123rd Audio Engineering Society Convention, New York, NY [Audio Engineering Society, 60 E. 42 St., Rm. 2520, New York, NY 10165-2520; Tel.: 212-661-8528; Fax: 212-682-0477; Web: www.aes.org]
22–24 Oct.	NOISE-CON 2007, Reno, NV [Institute of Noise Control Engineering, INCE Business Office, 210 Marston Hall, Ames, IA 50011-2153; Tel.: (515) 294-6142; Fax: (515) 294-3528; Email: ibo@inceusa.org]
27 Nov.–2 Dec.	154th Meeting of the Acoustical Society of America, New Orleans, Louisiana (note Tuesday through Saturday) [Acoustical Society of America, Suite 1NO1, 2 Huntington Quadrangle, Melville, NY 11747-4502; Tel.: 516-576-2360; Fax: 516-576-2377; Email: asa@aip.org; Web: http://asa.aip.org].

2008

29 June–4 July	Acoustics08, Joint Meeting of the Acoustical Society of America (ASA), European Acoustical Association (EAA), and the Acoustical Society of France (SFA), Paris, France [Acoustical Society of America, Suite 1NO1, 2 Huntington Quadrangle, Melville, NY 11747-4502; Tel.: 516-576-2360; Fax: 516-576-2377; E-mail: asa@aip.org; Web: http://asa.aip.org/meetings.html].
27–30 July	NOISE-CON 2008, Dearborn, MI [Institute of Noise Control Engineering, INCE Business Office, 210 Marston Hall, Ames, IA 50011-2153; Tel.: (515) 294-6142; Fax: (515) 294-3528; E-mail: ibo@inceusa.org]
28 Jul.–1 Aug.	9th International Congress on Noise as a Public Health Problem Quintennial meeting of ICBEN, the International Commission on Biological Effects of Noise). Foxwoods Resort, Mashantucket, CT [Jerry V. Tobias, ICBEN 9, Post Office Box 1609, Groton CT 06340-1609; Tel.: 860-572-0680; Web: www.icben.org; Email icben2008@att.net]

ACOUSTICAL STANDARDS NEWS

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

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

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

Take Note!

In April 2007 Accredited Standards Committee S3, Bioacoustics, approved the following erratum to ANSI S3.22-2003 *American National Standard Specification of Hearing Aid Characteristics*:

Erratum to ANSI S3.22-2003 American National Standard Specification of Hearing Aid Characteristics

In clause **6.11, Harmonic distortion**, two equations are shown with a "v" where a square root sign "√" should appear.

The equations should read:

a) $\%THD = 100\sqrt{([p_2^2 + p_3^2 + p_4^2 + \ldots.]/p_1^2)}$

b) $\%THD = 100\sqrt{([p_2^2 + p_3^2 + p_4^2 + \ldots.]/[p_1^2 + p_2^2 + p_3^2 + p_4^2 \ldots.])}$

If you own a copy of this standard, you can download a free copy of this erratum from the ASA Store at http://asastore.aip.org/.

Standards Meetings Calendar – National

At the 154th ASA Meeting in New Orleans, Louisiana, 27 November–1 December 2007 (Note: Tuesday through Saturday meeting), the ASA Committee on Standards (ASACOS) and ASACOS Steering Committees will meet as below:

•Tuesday, 27 November 2007

ASACOS Steering Committee

•Wednesday, 28 November 2007

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

Accredited Standards Committee on Acoustics, S1

(P. Battenberg, Chair; R.J. Peppin, Vice Chair)

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

S1 WORKING GROUPS

S1/Advisory–Advisory Planning Committee to S1 (P. Battenberg)

S1/WG1–Standard Microphones and their Calibration (V. Nedzelnitsky)

S1/WG4–Measurement of Sound Pressure Levels in Air (M. Nobile, Chair; E. Dunens, Vice Chair)

S1/WG5–Band Filter Sets (A. H. Marsh)

S1/WG17–Sound Level Meters and Integrating Sound Level Meters (B. M. Brooks)

S1/WG19–Insertion Loss of Windscreens (A. J. Campanella)

S1/WG20–Ground Impedance (K. Attenborough, Chair; J. Sabatier, Vice Chair)

S1/WG22–Bubble Detection and Cavitation Monitoring (Vacant)

S1/WG25–Specification for Acoustical Calibrators (P. Battenberg)

S1/WG26–High Frequency Calibration of the Pressure Sensitivity of Microphones (A. J. Zuckerwar)

S1/WG27–Acoustical Terminology (J. S. Vipperman)

S1/WG28–Passive Acoustic Monitoring for Marine Mammal Mitigation for Seismic Surveys (A. M. Thode)

S1 STANDARDS ON ACOUSTICS

ANSI S1.1-1994 (R 2004) American National Standard Acoustical Terminology

ANSI S1.4-1983 (R 2006) American National Standard Specification for Sound Level Meters. This Standard includes ANSI S1.4A-1985 (R 2001) Amendment to ANSI S1.4-1983

ANSI S1.6-1984 (R 2006) American National Standard Preferred Frequencies, Frequency Levels, and Band Numbers for Acoustical Measurements

ANSI S1.8-1989 (R 2006) American National Standard Reference Quantities for Acoustical Levels

ANSI S1.9-1996 (R 2006) American National Standard Instruments for the Measurement of Sound Intensity

ANSI S1.11- 2004 American National Standard Specification for Octave-Band and Fractional-Octave-Band Analog and Digital Filters

ANSI S1.13-2005 American National Standard Measurement of Sound Pressure Levels in Air

ANSI S1.14-1998 (R 2003) American National Standard Recommendations for Specifying and Testing the Susceptibility of Acoustical Instruments to Radiated Radio-Frequency Electromagnetic Fields, 25 MHz to 1 GHz

ANSI S1.15-1997/Part 1 (R 2006) American National Standard Measurement Microphones, Part 1: Specifications for Laboratory Standard Microphones

ANSI S1.15-2005/Part 2 American National Standard Measurement Microphones, Part 2: Primary Method for Pressure Calibration of Laboratory Standard Microphones by the Reciprocity Technique

ANSI S1.16-2000 (R 2005) American National Standard Method for Measuring the Performance of Noise Discriminating and Noise Canceling Microphones

ANSI S1.17-2004/Part 1 American National Standard Microphone Windscreens, Part 1: Measurements and Specification of Insertion Loss in Still or Slightly Moving Air

ANSI S1.18-1999 (R 2004) American National Standard Template Method for Ground Impedance

ANSI S1.20-1988 (R 2003) American National Standard Procedures for Calibration of Underwater Electroacoustic Transducers

ANSI S1.22-1992 (R 2007) American National Standard Scales and Sizes for Frequency Characteristics and Polar Diagrams in Acoustics

ANSI S1.24 TR-2002 ANSI Technical Report Bubble Detection and Cavitation Monitoring

ANSI S1.25-1991 (R 2007) American National Standard Specification for Personal Noise Dosimeters

ANSI S1.26-1995 (R 2004) American National Standard Method for Calculation of the Absorption of Sound by the Atmosphere

ANSI S1.40-2006 American National Standard Specifications and Verification Procedures for Sound Calibrators. (Revision of ANSI S1.40-1984)

ANSI S1.42-2001 (R 2006) American National Standard Design Response of Weighting Networks for Acoustical Measurements

ANSI S1.43-1997 (R 2007) American National Standard Specifications for Integrating-Averaging Sound Level Meters

Accredited Standards Committee on Mechanical Vibration and Shock, S2

(R.L. Eshleman, Chair; A.T. Herfat, Vice Chair)

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

S2 WORKING GROUPS

S2/WG1–S2 Advisory Planning Committee (R. L. Eshleman, Chair; A. T. Herfat, Vice Chair)

S2/WG2–Terminology and Nomenclature in the Field of Mechanical Vibration and Shock and Condition Monitoring and Diagnostics of Machines (D. J. Evans)

S2/WG3–Signal Processing Methods (T. S. Edwards)

S2/WG4–Characterization of the Dynamic Mechanical Properties of Viscoelastic Polymers (W. M. Madigosky, Chair, J. Niemiec, Vice Chair)

S2/WG5–Use and Calibration of Vibration and Shock Measuring Instruments (D. J. Evans, Chair; B. E. Douglas, Vice Chair)

S2/WG6–Vibration and Shock Actuators (G. Booth)

S2/WG7–Acquisition of Mechanical Vibration and Shock Measurement Data (B. E. Douglas)

S2/WG8–Analysis Methods of Structural Dynamics (B. E. Douglas)

S2/WG9–Training and Accreditation (R. L. Eshleman, Chair; D. Corelli, Vice Chair)

S2/WG10–Measurement and Evaluation of Machinery for Acceptance and Condition (R. L. Eshleman, Chair; H. C. Pusey, Vice Chair)

S2/WG10/Panel 1– Balancing (R. L. Eshleman)

S2/WG10/Panel 2– Operational Monitoring and Condition Evaluation (R. Bankert)

S2/WG10/Panel 3–Machinery Testing (R. L. Eshleman)

S2/WG10/Panel 4–Prognosis (R. L. Eshleman)

S2/WG10/Panel 5–Data Processing, Communication, and Presentation (K. Bever)

S2/WG11–Measurement and Evaluation of Mechanical Vibration of Vehicles (A. F. Kilcullen)

S2/WG12–Measurement and Evaluation of Structures and Structural Systems for Assessment and Condition Monitoring (B. E. Douglas)

S2/WG13–Shock Test Requirements for Shelf-mounted and Other Commercial Electronics Systems (J. Baker)

S2/WG39 (S3)–Human Exposure to Mechanical Vibration and Shock– Parallel to ISO/TC 108/SC 4 (D. D. Reynolds, Chair; R. Dong, Vice Chair)

S2 STANDARDS ON MECHANICAL VIBRATION AND SHOCK

ANSI S2.1-2000 / ISO 2041:1990 American National Standard Vibration and Shock—Vocabulary (Nationally Adopted International Standard)

ANSI S2.2-1959 (R 2006) American National Standard Methods for the Calibration of Shock and Vibration Pickups

ANSI S2.4-1976 (R 2004) American National Standard Method for Specifying the Characteristics of Auxiliary Analog Equipment for Shock and Vibration Measurements

ANSI S2.7-1982 (R 2004) American National Standard Balancing Terminology

ANSI S2.8-1972 (R 2006) American National Standard Guide for Describing the Characteristics of Resilient Mountings

ANSI S2.9-1976 (R 2006) American National Standard Nomenclature for Specifying Damping Properties of Materials

ANSI S2.16-1997 (R 2006) American National Standard Vibratory Noise Measurements and Acceptance Criteria of Shipboard Equipment

ANSI S2.17-1980 (R 2004) American National Standard Techniques of Machinery Vibration Measurement

ANSI S2.19-1999 (R 2004) American National Standard Mechanical Vibration—Balance Quality Requirements of Rigid Rotors, Part 1: Determination of Permissible Residual Unbalance, Including Marine Applications

ANSI S2.20-1983 (R 2006) American National Standard Estimating Air Blast Characteristics for Single Point Explosions in Air, with a Guide to Evaluation of Atmospheric Propagation and Effects

ANSI S2.21-1998 (R 2002) American National Standard Method for Preparation of a Standard Material for Dynamic Mechanical Measurements

ANSI S2.22-1998 (R 2002) American National Standard Resonance Method for Measuring the Dynamic Mechanical Properties of Viscoelastic Materials

ANSI S2.23-1998 (R 2002) American National Standard Single Cantilever Beam Method for Measuring the Dynamic Mechanical Properties of Viscoelastic Materials

ANSI S2.24-2001 (R 2006) American National Standard Graphical Presentation of the Complex Modulus of Viscoelastic Materials

ANSI S2.25-2004 American National Standard Guide for the Measurement, Reporting, and Evaluation of Hull and Superstructure Vibration in Ships

ANSI S2.26-2001 (R 2006) American National Standard Vibration Testing Requirements and Acceptance Criteria for Shipboard Equipment

ANSI S2.27-2002 American National Standard Guidelines for the Measurement and Evaluation of Vibration of Ship Propulsion Machinery

ANSI S2.28-2003 American National Standard Guide for the Measurement and Evaluation of Vibration of Shipboard Machinery

ANSI S2.29-2003 American National Standard Guide for the Measurement and Evaluation of Vibration of Machine Shafts on Shipboard Machinery

ANSI S2.31-1979 (R 2004) American National Standard Methods for the Experimental Determination of Mechanical Mobility, Part 1: Basic Definitions and Transducers

ANSI S2.32-1982 (R 2004) American National Standard Methods for the Experimental Determination of Mechanical Mobility, Part 2: Measurements Using Single-Point Translational Excitation

ANSI S2.34-1984 (R 2005) American National Standard Guide to the Experimental Determination of Rotational Mobility Properties and the Complete Mobility Matrix

ANSI S2.42-1982 (R 2004) American National Standard Procedures for Balancing Flexible Rotors

ANSI S2.43-1984 (R 2005) American National Standard Criteria for Evaluating Flexible Rotor Balance

ANSI S2.46-1989 (R 2005) American National Standard Characteristics to be Specified for Seismic Transducers

ANSI S2.48-1993 (R 2006) American National Standard Servo-Hydraulic Test Equipment for Generating Vibration–Methods of Describing Characteristics

ANSI S2.60-1987 (R 2005) American National Standard Balancing Machines—Enclosures and Other Safety Measures

ANSI S2.61-1989 (R 2005) American National Standard Guide to the Mechanical Mounting of Accelerometers

ANSI S2.70-2006 American National Standard Guide for the Measurement and Evaluation of Human Exposure to Vibration Transmitted to the Hand, a revision of ANSI S3.34-1986

ANSI S2.71-1983 (R 2006) American National Standard Guide to the Evaluation of Human Exposure to Vibration in Buildings, formerly ANSI S3.29-1983

ANSI S2.72-2002/Part 1/ISO 2631-1:1997 (R2007) Mechanical vibration and shock—Evaluation of human exposure to whole-body vibration—Part 1: General requirements (Nationally Adopted International Standard), formerly ANSI S3.18-2002 / ISO 2631-1:1997

ANSI S2.72-2003/Part 4/ISO 2631-4:2001 (R2007) Mechanical vibration and shock—Evaluation of human exposure to whole body vibration—Part 4: Guidelines for the evaluation of the effects of vibration and rotational motion on passenger and crew comfort in fixed-guideway transport systems (Nationally Adopted International Standard), formerly ANSI S3.18-2003/Part 4/ISO 2631-4:2001

ANSI S2.73-2002/ISO 10819:1996 (R2007) Mechanical vibration and shock—Hand-arm vibration—Method for the measurement and evaluation of the vibration transmissibility of gloves at the palm of the hand (Nationally Adopted International Standard), formerly ANSI S3.40-2002/ISO 10819:1996

Accredited Standards Committee on Bioacoustics, S3

(C. A. Champlin, Chair; D. A. Preves, Vice Chair)

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

S3 WORKING GROUPS

S3/Advisory– Advisory Planning Committee to S3 (C. Champlin, Chair; D. A. Preves, Vice Chair)

S3/WG35–Audiometers (R. L. Grason)

S3/WG36– Speech Intelligibility (R. S. Schlauch)

S3/WG37– Coupler Calibration of Earphones (B. Kruger)

S3/WG39–Human Exposure to Mechanical Vibration and Shock (parallel to ISO/TC 108/SC 4) (D. D. Reynolds, Chair; R. Dong, Vice Chair)

S3/WG43–Method for Calibration of Bone Conduction Vibrators (J. D. Durrant)

S3/WG48– Hearing Aids (D. A. Preves)

S3/WG51–Auditory Magnitudes (R. P. Hellman)

S3/WG56–Criteria for Background Noise for Audiometric Testing (J. Franks)

S3/WG59–Measurement of Speech Levels (M. C. Killion and L. A. Wilber, Co-Chairs)

S3/WG60– Measurement of Acoustic Impedance and Admittance of the Ear (Vacant)

S3/WG62–Impulse Noise with Respect to Hearing Hazard (J. H. Patterson, Chair; R. Hamernik, Vice Chair)

S3/WG67–Manikins (M. D. Burkhard)

S3/WG72–Measurement of Auditory Evoked Potentials (R. F. Burkard)

S3/WG76–Computerized Audiometry (A. J. Miltich)

S3/WG79–Methods for Calculation of the Speech Intelligibility Index (C. V. Pavlovic)

S3/WG81–Hearing Assistance Technologies (L. Thibodeau and L. A. Wilber, Co-Chairs)

S3/WG82–Basic Vestibular Function Test Battery (C. Wall, III)

S3/WG83–Sound Field Audiometry (T. R. Letowski)

S3/WG84–Otoacoustic Emission (G. R. Long)

S3/WG88–Standard Audible Emergency Evacuation and Other Signals (Vacant)

S3/WG89–Spatial Audiometry in Real and Virtual Environments (J. Besing)

S3/WG90–Animal Bioacoustics (A. E. Bowles)

S3/WG91–Text-to-Speech Synthesis Systems (C. Bickley and A. K. Syrdal, Co-Chairs)

S3/WG92–Effects of Sound on Fish and Turtles (R. R. Fay and A. N. Popper, Co-Chairs)

S3 Liaison Group

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

S3 STANDARDS ON BIOACOUSTICS

ANSI S3.1-1999 (R 2003) American National Standard Maximum Permissible Ambient Noise Levels for Audiometric Test Rooms

ANSI S3.2-1989 (R 1999) American National Standard Method for Measuring the Intelligibility of Speech over Communication Systems

ANSI S3.4-2005 American National Standard Procedure for the Computation of Loudness of Steady Sound

ANSI S3.5-1997 (R 2002) American National Standard Methods for Calculation of the Speech Intelligibility Index

ANSI S3.6-2004 American National Standard Specification for Audiometers *(Revision of ANSI S3.6-1996)*

ANSI S3.7-1995 (R 2003) American National Standard Method for Coupler Calibration of Earphones

ANSI S3.13-1987 (R 2002) American National Standard Mechanical Coupler for Measurement of Bone Vibrators

ANSI S3.20-1995 (R 2003) American National Standard Bioacoustical Terminology

ANSI S3.21-2004 American National Standard Methods for Manual Pure-Tone Threshold Audiometry *(Revision of ANSI S3.21-1978)*

ANSI S3.22-2003 American National Standard Specification of Hearing Aid Characteristics *(Revision of ANSI S3.22-1996)*

ANSI S3.25-1989 (R 2003) American National Standard for an Occluded Ear Simulator

ANSI S3.35-2004 American National Standard Method of Measurement of Performance Characteristics of Hearing Aids under Simulated Real-Ear Working Conditions

ANSI S3.36-1985 (R 2006) American National Standard Specification for a Manikin for Simulated *in situ* Airborne Acoustic Measurements

ANSI S3.37-1987 (R 2002) American National Standard Preferred Earhook Nozzle Thread for Postauricular Hearing Aids

ANSI S3.39-1987 (R 2002) American National Standard Specifications for Instruments to Measure Aural Acoustic Impedance and Admittance (Aural Acoustic Immittance)

ANSI S3.41-1990 (R 2001) American National Standard Audible Emergency Evacuation Signal

ANSI S3.42-1992 (R 2002) American National Standard Testing Hearing Aids with a Broad-Band Noise Signal

ANSI S3.44-1996 (R 2006) American National Standard Determination of Occupational Noise Exposure and Estimation of Noise-Induced Hearing Impairment

ANSI S3.45-1999 American National Standard Procedures for Testing Basic Vestibular Function

ANSI S3.46-1997 (R 2002) American National Standard Methods of Measurement of Real-Ear Performance Characteristics of Hearing Aids

Accredited Standards Committee on Noise, S12

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

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

S12 Working Groups

S12/Advisory–Advisory Planning Committee to S12 (R. D. Hellweg)

S12/WG3–Measurement of Noise from Information Technology and Telecommunications Equipment (K. X. C. Man)

S12/WG11–Hearing Protector Attenuation and Performance (E. H. Berger)

S12/WG13–Method for the Selection of Hearing Protectors that Optimize the Ability to Communicate (D. Byrne)

S12/WG14–Measurement of the Noise Attenuation of Active and/or Passive Level Dependent Hearing Protective Devices (W. J. Murphy, Chair)

S12/WG15–Measurement and Evaluation of Outdoor Community Noise (P. D. Schomer)

S12/WG18–Criteria for Room Noise (R. J. Peppin)

S12/WG23–Determination of Sound Power (J. Schmitt and B. M. Brooks, Co-Chairs)

S12/WG31–Predicting Sound Pressure Levels Outdoors (R. J. Peppin, Chair; L. Pater, Vice Chair)

S12/WG32–Revision of ANSI S12.7–1986 Methods for Measurement of Impulse Noise (A. H. Marsh)

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

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

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

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

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

S12/WG45–Measurement of Occupational Noise Exposure from Telephone Equipment (K. A. Woo, Chair; L. A. Wilber, Vice Chair)

S12/WG46–Acoustical Performance Criteria for Relocatable Classrooms (T. Hardiman, P. D. Schomer, Co-Chairs)

S12/WG47–Underwater Noise Measurements of Ships (M. Bahtiarian)

S12/WG48–Railroad Horn Sound Emission Testing (J. Erdreich, Chair; J. J. Earshen, Vice Chair)

S12/WG49–Noise from Hand-operated Power Tools, Excluding Pneumatic Tools (B. M. Brooks)

S12 LIAISON GROUPS

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

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

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

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

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

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

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

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

S12 STANDARDS ON NOISE

ANSI S12.1-1983 (R 2006) American National Standard Guidelines for the Preparation of Standard Procedures to Determine the Noise Emission from Sources

ANSI S12.2-1995 (R 1999) American National Standard Criteria for Evaluating Room Noise

ANSI S12.3-1985 (R 2006) American National Standard Statistical Methods for Determining and Verifying Stated Noise Emission Values of Machinery and Equipment

ANSI S12.5-2006/ISO 6926:1999 American National Standard Acoustics—Requirements for the Performance and Calibration of Reference Sound Sources Used for the Determination of Sound Power Levels (Nationally Adopted International Standard)

ANSI S12.6-1997 (R 2002) American National Standard Methods for Measuring the Real-Ear Attenuation of Hearing Protectors *(Revision of ANSI S12.6-1984)*

ANSI S12.7-1986 (R 2006) American National Standard Methods for Measurements of Impulse Noise

ANSI S12.8-1998 (R 2003) American National Standard Methods for Determining the Insertion Loss of Outdoor Noise Barriers

ANSI S12.9-1988/Part 1 (R 2003) American National Standard Quantities and Procedures for Description and Measurement of Environmental Sound, Part 1

ANSI S12.9-1992/Part 2 (R 2003) American National Standard Quantities and Procedures for Description and Measurement of Environmental Sound, Part 2: Measurement of Long-Term, Wide-Area Sound

ANSI S12.9-1993/Part 3 (R 2003) American National Standard Quantities and Procedures for Description and Measurement of Environmental Sound, Part 3: Short-Term Measurements with an Observer Present

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

ANSI S12.9-1998/Part 5 (R 2003) American National Standard Quantities and Procedures for Description and Measurement of Environmental Sound, Part 5: Sound Level Descriptors for Determination of Compatible Land Use

ANSI S12.9-2000/Part 6 (R 2005) American National Standard Quantities and Procedures for Description and Measurement of Environmental Sound, Part 6: Methods for Estimation of Awakenings Associated with Aircraft Noise Events Heard in Homes

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

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

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

ANSI S12.12-1992 (R 2002) American National Standard Engineering Method for the Determination of Sound Power Levels of Noise Sources Using Sound Intensity

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

ANSI S12.14-1992 (R 2002) American National Standard Methods for the Field Measurement of the Sound Output of Audible Public Warning Devices Installed at Fixed Locations Outdoors

ANSI S12.15-1992 (R 2002) American National Standard For Acoustics—Portable Electric Power Tools, Stationary and Fixed Electric Power Tools, and Gardening Appliances—Measurement of Sound Emitted

ANSI S12.16-1992 (R 2002) American National Standard Guidelines for the Specification of Noise of New Machinery

ANSI S12.17-1996 (R 2006) American National Standard Impulse Sound Propagation for Environmental Noise Assessment

ANSI S12.18-1994 (R 2004) American National Standard Procedures for Outdoor Measurement of Sound Pressure Level

ANSI S12.19-1996 (R 2006) American National Standard Measurement of Occupational Noise Exposure

ANSI S12.23-1989 (R 2006) American National Standard Method for the Designation of Sound Power Emitted by Machinery and Equipment

ANSI S12.30-1990 (R 2002) American National Standard Guidelines for the Use of Sound Power Standards and for the Preparation of Noise Test Codes

ANSI S12.42-1995 (R 2004) American National Standard Microphone-in-Real-Ear and Acoustic Test Fixture Methods for the Measurement of Insertion Loss of Circumaural Hearing Protection Devices

ANSI S12.43-1997 (R 2002) American National Standard Methods for Measurement of Sound Emitted by Machinery and Equipment at Workstations and Other Specified Positions

ANSI S12.44-1997 (R 2002) American National Standard Methods for Calculation of Sound Emitted by Machinery and Equipment at Workstations and Other Specified Positions from Sound Power Level

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

ANSI S12.51-2002 / ISO 3741:1999 American National Standard Acoustics—Determination of sound power levels of noise sources using sound pressure—Precision method for reverberation rooms. This Standard includes Technical Corrigendum 1-2001 (Nationally Adopted International Standard). *This standard replaces ANSI S12.31-1990 and ANSI S12.32-1990.*

ANSI S12.53/Part 1-1999 (R 2004)/ISO 3743-1:1994 American National Standard Acoustics—Determination of sound power levels of noise sources—Engineering methods for small, movable sources in reverberant fields—Part 1: Comparison method for hard-walled test rooms (Nationally Adopted International Standard). *This standard, along with ANSI S12.53/Part 2-1999, replaces ANSI S12.33-1990.*

ANSI S12.53/Part 2-1999 (R 2004)/ISO 3743-2:1994 American National Standard Acoustics—Determination of sound power levels of noise sources using sound pressure—Engineering methods for small, movable sources in reverberant fields—Part 2: Methods for special reverberation test rooms (Nationally Adopted International Standard). *This standard, along with ANSI S12.53/Part 1-1999 replaces ANSI S12.33-1990.*

ANSI S12.54-1999 (R 2004)/ISO 3744:1994 American National Standard Acoustics—Determination of sound power levels of noise sources using sound pressure—Engineering method in an essentially free field over a reflecting plane (Nationally Adopted International Standard). *This standard replaces ANSI S12.34-1988.*

ANSI S12.55-2006/ISO 3745:2003 American National Standard Acoustics—Determination of sound power levels of noise sources using sound pressure—Precision methods for anechoic and hemi-anechoic rooms (Nationally Adopted International Standard). *This standard replaces ANSI S12.35-1990.*

ANSI S12.56-1999 (R 2004)/ISO 3746:1995 American National Standard Acoustics—Determination of sound power levels of noise sources using sound pressure—Survey method using an enveloping measurement surface over a reflecting plane (Nationally Adopted International Standard). *This standard replaces ANSI S12.36-1990.*

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

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

ANSI S12.65-2006 American National Standard for Rating Noise with Respect to Speech Interference (R*evision of ANSI S3.14-1977)*

ASA Committee on Standards (ASACOS)

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

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

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

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

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

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

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

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

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

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

STANDARDS NEWS FROM THE UNITED STATES

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

American National Standards Call for Comment on Proposals Listed

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

ASA (ASC S2) (Acoustical Society of America)

REAFFIRMATIONS

BSR S2.21-1998 (R200x), Method for Preparation of a Standard Material for Dynamic Mechanical Measurements (Reaffirmation of ANSI S2.21-1998 (R2002)) (11 June 2007)

Applies to the preparation of a standard material for calibration of instruments for measuring the dynamic mechanical properties of viscoelastic materials. The purpose of this Standard is to assist users of dynamic mechanical test equipment in preparing the standard material from its basic components. The standard material is used for the calibration of new instruments in comparison with other instruments and in checking the operation of the same instrument at different times.

BSR S2.22-1998 (R200x), Resonance Method for Measuring the Dynamic Mechanical Properties of Viscoelastic Materials (Reaffirmation of ANSI S2.22-1998 (R2002)) (11 June 2007)

Defines a procedure for measurement and analysis of the dynamic properties of viscoelastic materials using a resonance method. The Standard applies to materials used in sound and vibration damping systems operating at frequencies from a fraction of a hertz to about 20 kHz.

BSR S2.23-1998 (R200x), Single Cantilever Beam Method for Measuring the Dynamic Mechanical Properties of Viscoelastic Materials (Reaffirmation of ANSI S2.23-1998 (R2002)) (11 June 2007)

Defines a method for measuring the dynamic mechanical properties of viscoelastic materials using a cantilever beam technique. The dynamic mechanical properties are expressed in terms of the frequency dependence of Young's modulus and loss factor at a given reference temperature. The Standard provides information for constructing such equipment and analyzing the results obtained.

BSR S2.27-2002 (R200x), Guidelines for the Measurement and Evaluation of Vibration of Ship Propulsion Machinery (Reaffirmation of ANSI S2.27-2002) (11 June 2007)

Contains guidelines for the measurement and evaluation of vibration of ship propulsion systems including limits for acceptability. It is applicable to all ocean-going ships and inland vessels. Test conditions, instrumentation, data analysis and evaluation, and reporting requirements are described.

BSR S2.72-2002-Part 1/ISO 2631-1-1997 (R200x), Mechanical vibration and shock—Evaluation of human exposure to whole-body vibration—Part 1: General requirements (Reaffirmation and redesignation of ANSI S3.18-1979 (R1999)) (23 April 2007)

Defines methods for the measurement of periodic, random, and transient whole-body vibration. Indicates the principal factors that combine to determine the degree to which a vibration exposure will be acceptable. Informative annexes indicate current opinion and provide guidance on the possible effects of vibration on health, comfort, perception, and motion sickness. The frequency range considered is: 0.5 to 80 Hz for health, comfort, and perception; and 0.1 to 0.5 Hz for motion sickness.

BSR S2.72-2003/Part 4/ISO 2631-4-2001 (R200x), Mechanical vibration and shock—Evaluation of human exposure to whole-body vibration—Part 4: Guidelines for the evaluation of the effects of vibration and rotational motion on passenger and crew comfort in fixed-guideway transport systems (Reaffirmation and redesignation of ANSI S3.18-2003, Part 4/ISO 2631-4-2001) (23 April 2007)

Aids in the design and evaluation of fixed-guideway passenger systems with regard to the impact of vibration and repetitive motions on passenger comfort.

BSR S2.73-2002/ISO 10819:1996 (R200x), Mechanical vibration and shock—Hand-arm vibration—Method for the measurement and evaluation of the vibration transmissibility of gloves at the palm of the hand (Reaffirmation and redesignation of ANSI S3.40-2002/ISO 10819:1996) (23 April 2007)

Specifies method for laboratory measurement, data analysis and reporting of vibration transmissibility of gloves in terms of vibration transmission from handle to palm of hand in frequency range 31.5–1250 Hz. Intended to define screening test for vibration transmission through gloves. Many factors influence transmission of vibration through gloves, so the transmissibility value according to this standard is not sufficient to assess health risk due to vibration.

ASA (ASC S3) (Acoustical Society of America)

NEW STANDARDS

BSR S3.47-200x, Specification of Hearing Assistance Devices/Systems (new standard) (23 April 2007)

Provides air conduction methods for evaluation of hearing assistance device/systems (HADS) that are packaged for individual use. Among the test methods described are family of output curves, output sound pressure curve for 90-dB sound pressure level input, frequency range, total harmonic distortion, noise level with no input, static and dynamic AGC characteristics, and volume control linearity.

REVISIONS

BSR S3.4-200x, Procedure for the Computation of Loudness of Steady Sounds (revision of ANSI S3.4-2005) (14 May 2007)

Specifies a procedure for calculating the loudness of steady sounds as perceived by a typical group of listeners with normal hearing, based on the spectra of the sounds. The possible sounds include simple and complex tones (both harmonic and inharmonic) and bands of noise.

REAFFIRMATIONS

BSR S3.5-1997 (R200x), Methods for Calculation of the Speech Intelligibility Index (Reaffirmation of ANSI S3.5-1997 (R2002)) (7 May 2007)

Defines a method for computing a physical measure that is highly correlated with the intelligibility of speech as evaluated by speech perception tests given a group of talkers and listeners.

BSR S3.13-1987 (R200x), Mechanical Coupler for Measurement of Bone Vibrators (Reaffirmation of ANSI S3.13-1987 (R2002)) (28 May 2007)

Specifies requirements for mechanical couplers used for calibrating bone-conduction audiometers and making measurements on bone vibrators and bone-conduction hearing aids. Specific design features are given for the mechanical coupler when driven by a vibrator with a prescribed plane circular tip area and applied with a specific static force. An appendix provides an example of a specific construction of a mechanical coupler.

BSR S3.37-1987 (R200x), Preferred Earhook Nozzle Thread for Postauricular Hearing Aids (Reaffirmation of ANSI S3.37-1987 (R2002)) (7 May 2007)

Describes a preferred thread for earhook nozzles on postauricular hearing aids. The need for such a standard arises from the wide variety of earhooks that hearing aid dispensers are required to keep in inventory to utilize different postauricular hearing aids from several manufacturers. This standard applies only to those postauricular hearing aids that utilize screw-on threads.

BSR S3.39-1987 (R200x), Specifications for Instruments to Measure Aural Acoustic Impedance and Admittance (Aural Acoustic Immittance) (Reaffirmation of ANSI S3.39-1987 (R2002)) (7 May 2007)

Provides specifications for instruments designed to measure acoustic impedance, acoustic admittance, or both quantities, within the human external ear canal. Terms that apply to these instruments and to related measurements are defined. Four types of instruments are classified. Characteristics, specifications, and recommended calibration procedures are then provided.

BSR S3.42-1992 (R200x), Testing Hearing Aids with a Broad-Band Noise Signal (Reaffirmation of ANSI S3.42-1992 (R2002)) (7 May 2007)

Defines a test method with which to characterize the steady-state frequency response and input/output characteristics of hearing aids as the input level varies. This method is particularly useful for those hearing aids that have automatic gain control or other types of adaptive circuitry.

BSR S3.46-1997 (R200x), Methods of Measurement of Real-Ear Performance Characteristics of Hearing Aids (Reaffirmation of ANSI S3.46-1997 (R2002)) (28 May 2007)

Provides definitions for terms used in the measurement of real-ear performance characteristics of hearing aids, provides procedural and reporting guidelines, and identifies essential characteristics to be reported by the manufacturer of equipment used for this purpose. Acceptable tolerances

for the control and measurement of sound pressure levels are indicated. Where possible, sources of error have been identified and suggestions provided for their management.

ASA (ASC S12) (Acoustical Society of America)

NEW STANDARDS

BSR S12.67-200x, Pre-Installation Airborne Sound Measurements and Acceptance Criteria of Shipboard Equipment (new standard) (14 May 2007)

Describes instrumentation and procedures for the pre-installation measurement and analysis of airborne noise generated by shipboard equipment. Maximum noise level criteria are presented for several types of equipment. This standard is based on MIL STD 740-1, "Airborne Sound Measurements and Acceptance Criteria of Shipboard Equipment" and MIL-STD-1474D, Requirement 5, "Shipboard Equipment Noise".

REVISIONS

BSR S12.9-Part 5-200x, Quantities and Procedures for Description and Measurement of Environmental Sound—Part 5: Sound Level Descriptors for Determination of Compatible Land Use (revision of ANSI S12.9-Part 5-1998 (R2003)) (23 April 2007)

Provides guidance on the compatibility of various human uses of land with the acoustical environment, using the yearly average total day-night adjusted sound exposure or the yearly average adjusted day-night average sound level to characterize the acoustical environment. An informative annex provides guidance to local authorities for designation of land uses compatible with existing or predicted yearly average total day-night adjusted sound exposure or yearly average adjusted day-night average sound level.

ASTM (ASTM International)

REAFFIRMATIONS

BSR/ASTM F2154-2001 (R200x), Specification for Sound-Absorbing Board, Fibrous Glass, Perforated Fibrous Glass Cloth Faced (Reaffirmation of ANSI/ASTM F2154-2001) (21 May 2007)

IEEE (Institute of Electrical and Electronics Engineers)

NEW STANDARDS

BSR/IEEE C57.127-200x, Guide for the Detection and Location of Acoustic Emissions from Partial Discharges in Oil-Immersed Power Transformers and Reactors (new standard) (3 July 2007)

Applies to the detection and location of acoustic emissions from partial discharges and other sources in oil-immersed power transformers and reactors. Both electrical sources (partial discharge) and mechanical sources (such as loose clamping, bolts, or insulation parts) generate these emissions. There are descriptions of acoustic instrumentation, test procedures, and interpretation of results.

Project Initiation Notification System (PINS)

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

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

ASA (ASC S12) (Acoustical Society of America)

BSR S12.6-200x, Methods for Measuring the Real-Ear Attenuation of Hearing Protectors (revision of ANSI S12.6-1997 (R2002))

Specifies laboratory-based procedures for measuring, analyzing, and reporting the noise-reducing capabilities of hearing protection devices. The methods consist of psychophysical tests conducted on human subjects to determine real-ear attenuation at threshold. Two methods are provided, differing in their subject selection, training, hearing protector fitting procedures, and experimenter involvement, but corresponding in all electroacoustic and psychophysical aspects. Project need: To modify certain aspects of the fitting protocols for better consistency. Stakeholders: Purchasers, manufacturers, administrators, and regulators of hearing protection.

SCTE (Society of Cable Telecommunications Engineers)

BSR/SCTE 62-200x, Measurement Procedure for Noise Figure (revision of ANSI/SCTE 62-2002)

This procedure defines a method of measurement for Noise Figure of active Cable Telecommunications equipment. It is intended for measurement of 75-ohm devices having type "F" or 5/8-24 KS connectors, and for the measurement of true broadband noise as opposed to narrowband disturbances. Project need: To define the method of measurement. Stakeholders: Cable Telecommunications Industry.

Final actions on American National Standards

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

ASA (ASC S1) (Acoustical Society of America)

REAFFIRMATIONS

ANSI S1.22-1992 (R2007), Scales and Sizes for Frequency Characteristics and Polar Diagrams in Acoustics (Reaffirmation of ANSI S1.22-1992 (R2002)) (16 March 2007)

ANSI S1.25-1991 (R2007), Specification for Personal Noise Dosimeters (Reaffirmation of ANSI S1.25-1991 (R2002)) (16 March 2007)

ANSI S1.43-1997 (R2007), Specifications for Integrating-Averaging Sound Level Meters (Reaffirmation of ANSI S1.43-1997 (R2002)) (16 March 2007)

ASA (ASC S2) (Acoustical Society of America)

REVISIONS

ANSI S2.8-2007, Technical Information Used for Resilient Mounting Applications (revision of ANSI S2.8-1972 (R2006)) (22 March 2007)

ASSE (ASC A10) (American Society of Safety Engineers)

NEW STANDARDS

ANSI A10.46-2007, Hearing Loss Prevention in Construction and Demolition Workers (new standard) (5 March 2007)

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

REVISIONS

ANSI C63.19-2007, Methods of Measurement of Compatibility between

Wireless Communication Devices and Hearing Aids (revision of ANSI C63.19-2006) (28 March 2007)

TIA (Telecommunications Industry Association)

ADDENDA

ANSI/TIA 470.110-C-1-2007, Telecommunications—Telephone Terminal Equipment—Handset Acoustics Performance Requirements—Addendum 1 (addenda to ANSI/TIA 470-110-C-2004) (11 April 2007)

STANDARDS NEWS FROM ABROAD

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

International Organization for Standardization (ISO)

Newly Published ISO and IEC Standards

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

ISO Standards

ACOUSTICS (TC 43)

ISO 1996-2:2007, Acoustics—Description, measurement and assessment of environmental noise—Part 2: Determination of environmental noise levels

IEC Standards

ROTATING MACHINERY (TC 2)

IEC 60034-9 Amd.1 Ed. 4.0 b:2007, Amendment 1—Rotating electrical machines—Part 9: Noise limits

IEC 60034-14 Ed. 3.1 b:2007, Rotating electrical machines—Part 14: Mechanical vibration of certain machines with shaft heights 56 mm and higher—Measurement, evaluation and limits of vibration severity

ELECTROACOUSTICS (TC 29)

IEC 60645-3 Ed. 2.0 b:2007, Electroacoustics—Audiometric equipment—Part 3: Test signals of short duration

ISO Draft Standard

ACOUSTICS (TC 43)

ISO/DIS 13472-2, Acoustics—Measurement of sound absorption properties of road surfaces in situ—Part 2: Spot method for reflective surfaces (27 May 2007)

MECHANICAL VIBRATION AND SHOCK (TC 108)

ISO/DIS 20283-2, Mechanical vibration—Measurement of vibration on ships—Part 2: Structural vibration (6 June 2007)

BOOK REVIEWS

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

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

Audio Signal Processing and Coding

Andreas Spanias, Ted Painter, and Venkatraman Atti

Wiley-Interscience, John Wiley & Sons, Hoboken, New Jersey, 2007. 464 pp. Price $95.00 (hardcover), ISBN: 978-0-471-79147-8

This text is designed "to provide an in-depth treatment of audio compression algorithms and standards." The key idea of the text is to present audio coding/compression algorithms "to obtain compact digital representations of high-fidelity (wideband) audio signals for the purpose of efficient transmission or storage." That is, the objective is "to represent the signal with the minimum number of bits while achieving *transparent* signal reproduction." The authors of this text are able to achieve their objectives in writing such a text in a commendable manner; however, the background required to comprehend much of their treatment is higher than that of an "audiophile"—more likely students in digital signal processing (DSP) and multimedia would not have trouble appreciating such a text.

Unfortunately, the Introduction *does* require some knowledge of signal processing jargon to appreciate the discussion. Such terms as data rates, sampling, resolution, bits, pulse-coded modulation, and frequency (kHz) are not really defined until the detailed chapters of the book are encountered. This approach encompasses the entire introduction implying a higher entry level of the readership with some fundamental knowledge of DSP. With this in mind, the subsequent chapter on digital signal processing is adequate for audio processing purposes providing a reasonable discussion of the necessary topics for comprehension of the more detailed subjects to follow.

The discussion of introductory quantization and coding is nicely done incorporating both scalar and vector quantization techniques. Rudimentary codes are introduced with enough example problems to demonstrate each approach and make comparisons. The next chapter briefly discusses linear predictive coding (LPC), which is heavily used in speech systems finding high success in telephony. Here the discussion briefly outlines the basic concepts of analysis-synthesis LPC techniques and demonstrates how they are applied to typical speech systems.

It is rather surprising that the heart of these techniques relies on modulation schemes such as pulse-coded modulation (PCM) or differential PCM, etc., that are merely referenced throughout the text and never really defined mathematically, again implying more *a priori* knowledge of communication theory than a typical audiophile or beginner might actually have.

The chapter on psychoacoustic principles is well done. The authors introduce the concept of psychoacoustics along with the idea of perceptual processing. Here the idea is to characterize the human audio system along with perception of sounds with the goal of removing irrelevant information to achieve a highly compressed audio signal. Masking of tones, noise, etc., along with the concept of perceptual entropy is introduced as a metric that establishes the absolute limit of signal compression. This chapter carefully defines the concepts and principles behind such phenomena as masking and leads the reader very nicely through an example of applying psychoacoustic principles to processing an audio signal along with all of the detailed processing along the way. The authors cleverly review the current practice and introduce many practical and useful formulas from the literature.

A chapter on time-frequency processing basics follows in which the authors introduce properties of the audio signal amenable for block analysis. Here they show how a variety of bandpass filter banks can be designed and applied to identify perceptual irrelevancies and aide in extracting a set of block parameters amenable to quantization. The authors carefully develop a variety of bandpass filter bank designs and discuss transform techniques that can be used in the development of an audio encoder. They discuss both discrete Fourier (DFT) and discrete cosine transform (DCT) techniques and offer examples to illustrate their main design points. Time-frequency analysis for signal processing and coding is emphasized and designers are led through the various techniques in an attempt to develop "perfect reconstruction" of the audio signal. This is a very informative chapter.

Next follows the discussion of transform coders which utilize the DFT, DCT, and other transforms for the time-frequency analysis section of the audio coder. Here a variety of coders are briefly discussed ranging from the optimum frequency domain coding, to the perceptual transform coder (monaural and stereo), to hybrid coders and adaptive spectral coders. In all cases, the coders are described along with detailed references cited for the interested reader. Subband coders are then discussed in Chapter 8 incorporating the discrete wavelet transform (DWT). Details of the wavelet transform are deferred to the references, but enough information is provided to comprehend the coding schemes discussed. The DWT coders are discussed including a variety of structures along with hybrid coders exploiting both the wavelet and harmonic (frequency domain) signal representations. Chapter 9 concludes the detailed coder discussion with the important sinusoidal coder that is capable of achieving good quality for low rate coding. It is a parametric technique that depends on the superposition of sinusoidal models and associated time-varying frequencies, amplitudes, and phases. This model is used as a parametric tracking model for speech and music applications. It is generalized to a frequency modulation counterpart that is currently being investigated as an alternative. Illustrative examples are included to demonstrate the usefulness of these parametric harmonic representations.

The next chapter, 10, is concerned with audio coding standards and techniques. It encompasses the current state-of-the-art in audio ranging from the musical instrument digital interface (MIDI), to surround sound, to the Motion Pictures Experts Group (MPEG) audio systems with descriptions and applications. Encoders/decoders are discussed along with the particular algorithms employed as well as specific coding tools available for application (e.g. MPEG-7, data description Language (DDL)).

Lossless coding and watermarking are the subject of the next chapter. Here the authors present the current state-of-the art in both subjects. Lossless coding occurs when real-time storage/processing exist in storage-rich environments. First, the principles of lossless coding are discussed followed by a brief review of the major algorithmic approaches eliminating statistical dependencies using prediction algorithmic techniques (LPC, polynomial methods, etc.). Watermarking, used for copyright protection and identification, is also briefly discussed along with a survey and tables with references for the reader interested in more details.

The final chapter includes a discussion on measures of quality, both subjective and quantitative for perceptual audio coding. Although brief, a couple of examples are presented to establish the basic ideas. Future research directions are also discussed.

In summary, this text is more like an encyclopedia or detailed survey on audio signal processing and coding with a very rich list of references and enough of the details for any audio engineer. As an introductory text, it appears to require more detailed knowledge (as discussed earlier) of digital signal processing, communications, and coding techniques than that of an audiophile. On the other hand, this text is a "must buy" for any practicing audio engineer, since it represents an encyclopedia of standards, algorithms, and approaches to current audio coding problems.

JAMES V. CANDY
University of California
Lawrence Livermore National Laboratory
and UC Santa Barbara Livermore, CA 94551

 0001-4966/2007/122(1)/15/1/$23.00

Perceptual Audio Evaluation—Theory, Method and Application

Søren Bech and Nick Zacharov

John Wiley & Sons, Chichester, 2006.
462 pp. $135.00 (hardcover). ISBN: 0470869232

This book is a practitioner's handbook on the design and analysis of experiments involving human listeners in the area of sound quality evaluation. The scope is on audio and telecommunication applications and associated perceptual evaluation procedures. As the authors state in the preface, investigating the perception of sound quality requires knowledge in many scientific areas: audio recording and production, electronic engineering, signal processing, acoustics, experimental psychology, statistics, and perhaps in several others. The book provides insight into these fields from an application-oriented point of view, and it covers the entire process of an audio-evaluation study, from its initial conceptualization to the final reporting of the results.

The book is divided into two major parts: The first part entitled "Experimental considerations" starts out with two rather introductory chapters. In Chap. 2, the authors give a brief overview of fundamental epistemological principles (rationalism versus empiricism). They address the problem of formulating a testable hypothesis and give an example of how this might be approached in the context of audio evaluation. Chapter 3 goes more into the practical details encountered when conducting an experiment. It sketches the individual steps involved in an audio-evaluation study and summarizes them in a flow chart that serves as an overview of the entire book. Special emphasis is put on the fact that for an industrial application a perceptual study might not even be necessary; this might be the case when sufficient results from the literature are available and/or a suitable perceptual model of the situation already exists.

Chapter 4 is entitled "Quantification of impression," but it not only covers methods for quantifying the strength of sensations or of perceptual attributes, but also methods for their identification or elicitation in the first place. The authors start out by describing in detail prevalent *direct* attribute-elicitation methods. These methods are called "direct" because participants are asked to directly put a label to their sensations; consequently, these methods are "based on the assumption that there is a close relationship between a given sensation and the verbal descriptors used by the subject to describe the sensation" (p. 44). In contrast, indirect elicitation methods abandon this assumption; they are based on rather simple judgments, and attempt to infer the relevant perceptual attributes or features by subsequent analysis. The following sections are dedicated to the scaling proper, and the distinction between direct and indirect methods is kept. Among the direct methods, the authors describe category scaling and Stevens's magnitude estimation techniques; among the indirect methods, Fechner's classical JND scaling and Thurstone's and Luce's paired-comparison scaling are treated. The authors present some guidelines that might help the reader to choose the scaling method most appropriate to his or her application.

Chapter 5, "Experimental variables," covers major independent variables in an audio evaluation, such as the characteristics of the signal, the reproduction system, the listening room, and the listener. The main focus of the chapter, however, is on listener selection and training. An attempt is made to categorize listeners according to their expertise or experience; to this effect, a categorization scheme adopted from the food sciences is presented. Procedures are described to select subjects for a permanent listening panel, to train their desired hearing abilities, and to monitor their performance.

The final chapter of the first part deals with data analysis *per se*. It describes important aspects of the experimental design that have an influence on the choice of analysis method, such as the types of variables encountered in an experiment, the assignment of subjects to experimental conditions, and the required sample size. Subsequently, the authors treat methods of statistical analysis, starting from descriptive statistics and afterwards elaborating on procedures of inferential statistics, including tests for one, two, or more samples, and focusing on ANOVA techniques for both between and within-subjects design. At the end, the authors present a statistical checklist; while originally meant for reviewers of medical papers, the points on this list will also be worth considering when conducting perceptual studies.

The second part, "Technical considerations," is comprised of three chapters. Chapter 7 focuses on listening rooms and other spaces designed for conducting audio evaluations, and it introduces relevant standards (from IEC, ITU, and EBU) that give recommendations for the acoustical requirements of such rooms. The authors discuss the positioning of the listener and the loudspeaker(s) in monophonic, stereophonic, and multichannel reproduction; in addition, the controversial issue of the positioning of one or several subwoofers is considered. Subsequently, recommendations for experiments that include the presentation of accompanying visual stimuli, such as pictures and movies, are given. Finally, requirements for electronic equipment, such as amplifiers, digital-to-analog converters, and sound cards are presented.

The following chapter is dedicated to the issue of calibration in both the physical and the perceptual domains. Calibration methods that work on the intensity of signals, such as root-mean-square (RMS) calibration and metrics based on the weighted sound pressure level are presented. Subsequently, prevalent models (by Stevens, Zwicker, Moore) to estimate the loudness of signals are discussed, and some recommendations regarding reference reproduction levels are given. The remainder of this chapter covers topics related to the calibration of loudspeakers and headphones.

In Chap. 9, "Test planning, administration and reporting," the authors deal with administrative and logistic issues when conducting an audio-evaluation study. They discuss ethical guidelines for experiments involving human subjects (adopted from the American Psychological Association) and regulations regarding maximum exposure levels in such experiments. They give advice on how to write instructions and present some examples. Prevalent software packages for stimulus presentation and response collection are reviewed. The chapter closes with a short guide on how to summarize the outcomes of a study in a research report.

To sum up, this book thoroughly introduces the discipline of perceptual audio evaluation, and it does so using a hands-on approach. Many techniques and procedures are illustrated in flow charts for the reader to obtain a quick overview. Examples are provided in each chapter, in order to render the content easily accessible. The annotated collection of key standards relating to perceptual evaluation in one of the appendices is also worth mentioning. Although the primary audience seems to be research and development engineers in the audio and telecommunication industries, the book will provide a great source of information to anybody faced with the challenge of conducting audio-evaluation experiments.

FLORIAN WICKELMAIER
Department of Psychology
University of Tuebingen
Friedrichstrasse 21, 72072 Tuebingen, Germany

 0001-4966/2007/122(1)/16/1/$23.00

REVIEWS OF ACOUSTICAL PATENTS

Lloyd Rice
11222 Flatiron Drive, Lafayette, Colorado 80026

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

Reviewers for this issue:

GEORGE L. AUGSPURGER, *Perception, Incorporated, Box 39536, Los Angeles, California 90039*
ANGELO CAMPANELLA, *3201 Ridgewood Drive, Hilliard, Ohio 43026-2453*
ALIREZA DIBAZAR, *Department of BioMed Engineering, University of Southern California, Los Angeles, California 90089*
DIMITRI DONSKOY, *Stevens Institute of Technology, Castle Point on the Hudson, Hoboken, New Jersey 07030*
JOHN M. EARGLE, *JME Consulting Corporation, 7034 Macapa Drive, Los Angeles, California 90068*
GEOFFREY EDELMANN, *Naval Research Laboratory, Code 7145, 4555 Overlook Ave. SW, Washington,DC 20375*
SEAN A. FULOP, *California State University, Fresno, 5245 N. Backer Avenue M/S PB92, Fresno, California 93740-8001*
JEROME A. HELFFRICH, *Southwest Research Institute, San Antonio, Texas 78228*
DAVID PREVES, *Starkey Laboratories, 6600 Washington Ave. S., Eden Prarie, Minnesota 55344*
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*

7,164,621

43.30.Tg UNDERWATER SOUNDING APPARATUS

Akira Okunishi *et al.*, assignors to Furuno Electric Company, Limited
16 January 2007 (Class 367/131); filed in Japan 28 November 2003

This patent describes a noise-canceling device designed to remove propeller/shaft noise from an acoustic ocean sounding device (e.g., side scan and subbottom profilers). The patent asserts that the self-generated noise component is measured and removed via time delay and gain. The bulk of the document relates to a cylindrical array dedicated to measuring the propeller noise by wrapping around the shaft itself. Even if the noise field could be accurately measured, why remove noise of a frequency well below any modern bottom profiler?—GFE

7,164,618

43.30.Vh DUAL UNIT EIDETIC TOPOGRAPHER

Anthony Matthews, assignor to The United States of America as represented by the Secretary of the Navy
16 January 2007 (Class 367/11); filed 24 March 2005

This patent describes a method to produce a 3D image of an underwater target of interest (e.g., a mine or wreck). The method described requires two underwater vehicles, one of which remains stationary with a vertical array and the second circles the target while transmitting. This highly coordinated feat utilizes both beamforming and synthetic aperture processing.—GFE

7,164,622

43.30.Wi ACOUSTIC PROPAGATION DELAY MEASUREMENTS USING TRANSMISSION OF KNOWN BROAD BANDWIDTH WAVEFORMS

Barry B. Baker and Richard D. Haskell, both of Gales Ferry, Connecticut
16 January 2007 (Class 367/134); filed 10 September 2004

One of the problems in deploying a long, flexible sensor array, such as an underwater towing hydrophone array, is uncertainty in the position of an individual receiver. The patent describes a method of reducing this uncertainty by utilizing one or more transmitters broadcasting broadband signals. Measuring a time delay of the transmitted signals, a relative position of the receivers can be determined. The primary goal of the patent is to minimize the risk that a third party will detect the operation of the transmitter. This goal is achieved by using pseudorandom wideband waveforms with minimal signal-to-noise ratio at all frequencies. The broadcast signal energy should preferably be below the acoustic emission from the towing platform. Knowing the broadcast waveform and using matched filtering, the lag-space localization can be achieved with reduced detectability by a third party.—DMD

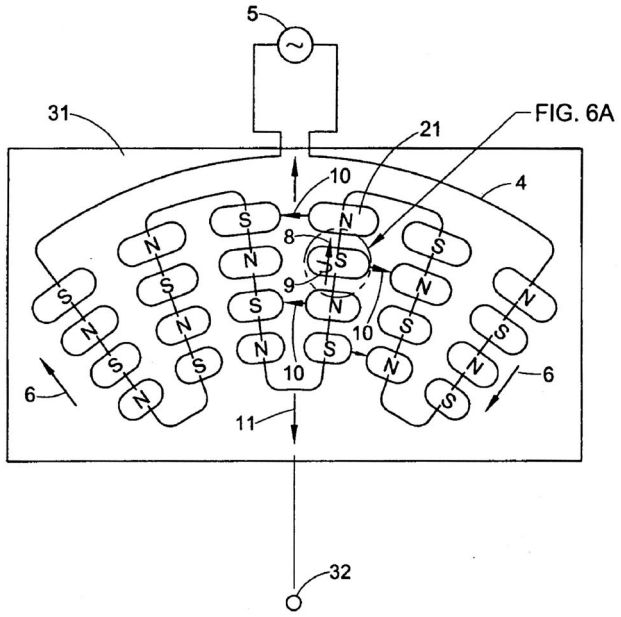
7,165,453

43.35.Zc FLEXIBLE ELECTROMAGNETIC ACOUSTIC TRANSDUCER SENSOR

John Flora *et al.*, assignors to Electric Power Research Institute
23 January 2007 (Class 73/643); filed 20 July 2005

The author claims a curved nondestructive testing (NDT) electromagnetic acoustic transducer (EMAT) that can be cut or bent into a curved shape to conform to the metal object being inspected for cracks or other faults. Magnetic pieces **N-S** are arranged into an array on a flexible insulating substrate **31**. These may be polarized by a dc field **8**. A rf current source **5** sends current **6** by magnetic pole faces **21** at the metal inspection surface. Via Lorenz force **9** in the metal surface, pressure waves **10** are generated that will propagate as shear or Rayleigh waves across the metal surface, i.e., this is a Surface Acoustic Wave (SAW) system. The EMAT is a reciprocal transducer; the current source **5** is replaced with a rf receiver to detect

 0001-4966/2007/122(1)/17/18/$23.00



reflected SAWs. The NDT function proceeds as SAWs will propagate well through good material surfaces, but will be scattered by flaws. A variety of EMAT materials and conformal shapes is claimed.—AJC

7,165,455

43.35.Zc SURFACE ACOUSTIC WAVE SENSOR METHODS AND SYSTEMS

Steven J. Magee *et al.*, assignors to Honeywell International Incorporated
23 January 2007 (Class 73/650); filed 18 December 2004

A torque and temperature sensor **100** is claimed where surface acoustic wave (SAW) devices **104, 106** are fastened to a thin section of a carrier

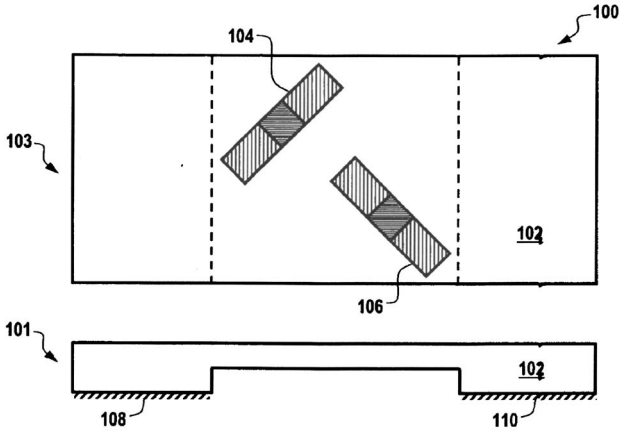


plate whose ends **102** are adhered to stressed material **108, 110** such as a propeller shaft. Material strains due to torque and temperature will change the SAW propagation time. The frequency of an oscillator with the SAW in the feedback path is accordingly changed. The difference in frequency between the two SAW device oscillators is a measure of torque, while the sum is a measure of temperature. More complex SAW arrays are also claimed.—AJC

7,168,322

43.35.Zc METHOD FOR ULTRASONIC CONTROL OF WELD JOINTS

Olivier Bardoux *et al.*, assignors to L'Air Liquide Société Anonyme à Directoire et Conseil de Surveillance pour l'Etude et l'Exploitation des Precédés Georges Claude
30 January 2007 (Class 73/588); filed in France 17 June 2002

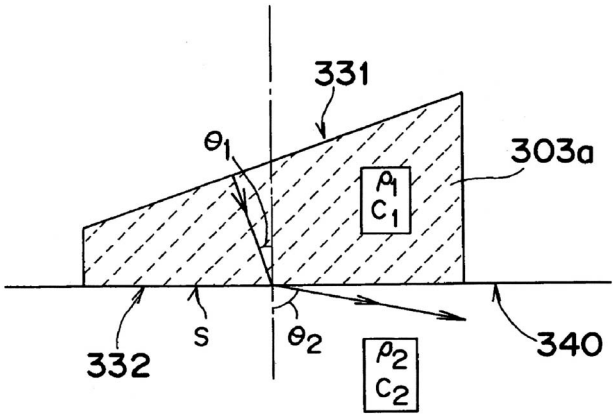
This patent describes a means for inspecting welds in metal plates with thicknesses between 5 and 100 mm without changing transducer arrangements. It covers this wide range of thicknesses by combining the time-of-flight diffraction (TOFD) technique with the creeping wave (CW) technique. A typical transducer arrangement that straddles the weld to be inspected and that can be moved along the length of the weld consists of at least two sending transducers on one side of the weld and at least one receiving transducer on the other side. One or more sending transducers inject ultrasound normal to the plate surface for TDOF measurements, whereas one or more other sending transducers generate longitudinal waves at an angle for CW measurements. The two sets of sending transducers are activated successively, the first to locate flaws at relatively great depths, the second to locate flaws at lesser depths.—EEU

7,162,930

43.38.Ar ULTRASONIC SENSOR

Masahiko Hashimoto *et al.*, assignors to Matsushita Electric Industrial Company, Limited
16 January 2007 (Class 73/861.25); filed 28 April 2003

This is a transducer that projects ultrasound into air and other gases. An ultrasound source is placed on surface **331**. Sound waves proceed through matching medium **303a**, such as silica gels, having lower density



and sound velocity as compared to the source transducer. At surface **S** the waves are refracted to a shallow angle as they emerge into the surrounding gas medium. Projection of sound at shallow angles is possible.—AJC

7,168,318

43.38.Ar ISOLATED PLANAR MESOGYROSCOPE

A. Dorian Challoner and Kirill V. Shcheglov, assignors to California Institute of Technology
30 January 2007 (Class 73/504.13); filed 12 April 2005

This patent discusses the factors involved in creating a gyroscope using frequency shifts on standing vibrational waves in a structure. This effect was first exploited in the 1980s with the invention of the hemispherical resonator gyro, and this patent takes that idea one step further to a planar geometry, explaining the trade-offs involved in going from a hemispherical resonator to a disk resonator. The patent is well-written and full of interest-

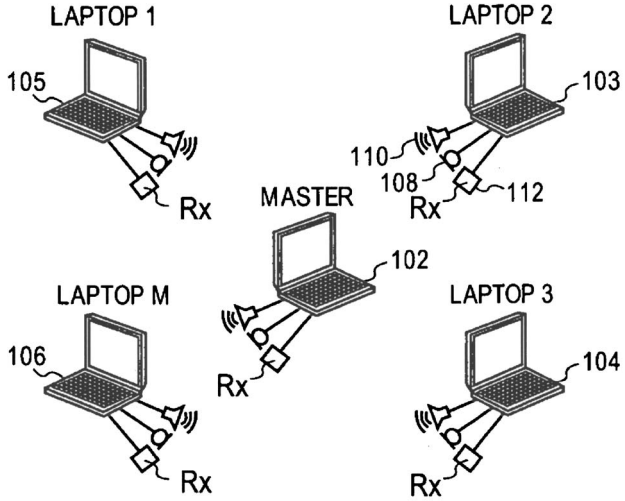
ing detail. Unfortunately, it also relies heavily on citations of previous work to explain many of the decisions made, and so for full comprehension when reading the patent, one needs to have access to much of the prior work.—JAH

7,035,757

43.38.Hz THREE-DIMENSIONAL POSITION CALIBRATION OF AUDIO SENSORS AND ACTUATORS ON A DISTRIBUTED COMPUTING PLATFORM

Vikas C. Raykar *et al.*, assignors to Intel Corporation
25 April 2006 (Class 702/150); filed 9 May 2003

A sound recording setup is described in which multiple small computing devices, such as laptops, are used at the individual collection points, rather than the more typical multiple-microphone arrangement in which all microphone signals are collected by a single multichannel recording device. The various acoustic collection devices would communicate with each other



by the use of wireless broadcasting technology. The primary purpose of the system seems to be the determination of the location of one or more acoustic sources. The system would differ from a typical microphone array in that the pickup points are described as distributed in the acoustic space.—DLR

7,162,043

43.38.Hz MICROPHONE ARRAY SOUND SOURCE LOCATION SYSTEM WITH IMAGING OVERLAY

Takeshi Sugiyama *et al.*, assignors to Chubu Electric Power Company, Incorporated
9 January 2007 (Class 381/92); filed in Japan 2 October 2000

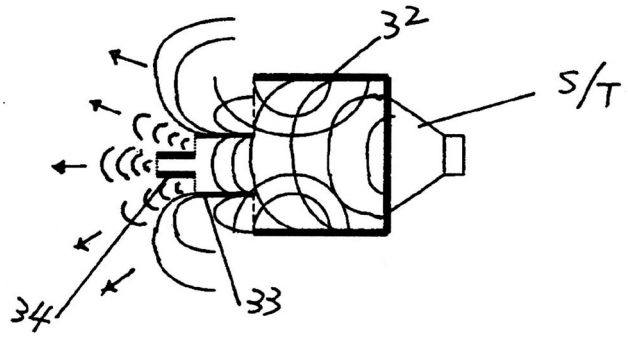
The patent describes a first-order, three-dimensional array of five microphones whose outputs enable the user to ''zero in'' on a distant sound source via beamforming. Actually, this can be done with only four microphones, and it is difficult to see where the fifth microphone fits in.—JME

7,130,438

43.38.Ja ACOUSTIC ENCLOSURE FOR SINGLE AUDIO TRANSDUCER

Fong-Jei Lin, Saratoga, California
31 October 2006 (Class 381/337); filed 22 December 2003

A fanciful interpretation of waves, diffraction, and sound propagation, among other things, yields an acoustic housing **32-33-34** that, when placed



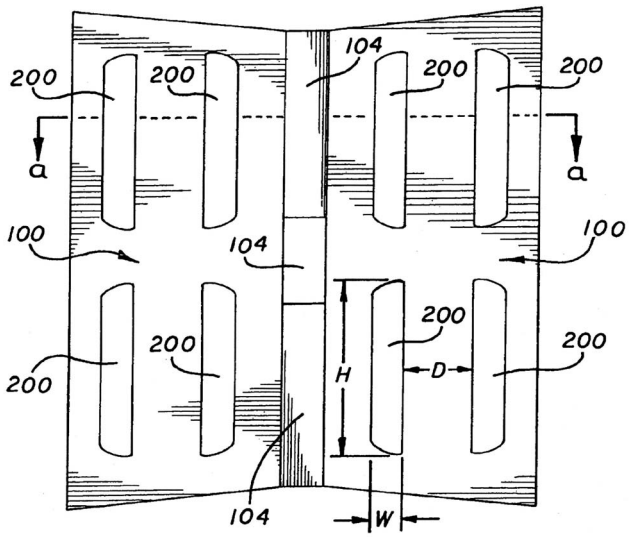
on a conventional cone-type electroacoustic transducer **S/T**, enables the low, mid, and high portions of the frequency spectrum to emanate from apertures in **32-33-34**, respectively. No measurement or other data are presented. This may be in an alternate universe, but not in a parallel universe.—NAS

7,134,523

43.38.Ja SYSTEM FOR INTEGRATING MID-RANGE AND HIGH-FREQUENCY ACOUSTIC SOURCES IN MULTI-WAY LOUDSPEAKERS

Mark Engebretson, assignor to Harman International Industries, Incorporated
14 November 2006 (Class 181/144); filed 22 November 2002

A means is disclosed of integrating both mid- and high-frequency drivers into a combined baffle **100** (called a radiation boundary integrator or RBI in the patent). Midfrequency drivers are placed behind the pairs of vertical slots while the high-frequency drivers are placed behind the central slots. The side-slot pairs can have foam placed in them that is said to be transondent to midfrequency radiation, but not so to high-frequency radiation. The geometry of the baffle behind the side-slot pairs **200** can be contoured so that a phase plug of sorts is formed. The foam is said to form an



acoustical barrier for the high frequencies radiating outward across the RBI front surface, and also to serve as a volume displacement device for the midfrequency drivers due to the contoured back of the slot pairs. The invention appears to relate to a line array system marketed by a subsidiary of the assignee.—NAS

7,136,498

43.38.Ja LOUDSPEAKER HAVING A DUAL CHAMBER ACOUSTICAL ENCLOSURE WITH TWO EXTERNAL VENTS AND ONE INTERNAL VENT

Wayne M. Schott, assignor to Koninklijke Philips Electronics N.V.
14 November 2006 (Class 381/351); filed 16 December 1999

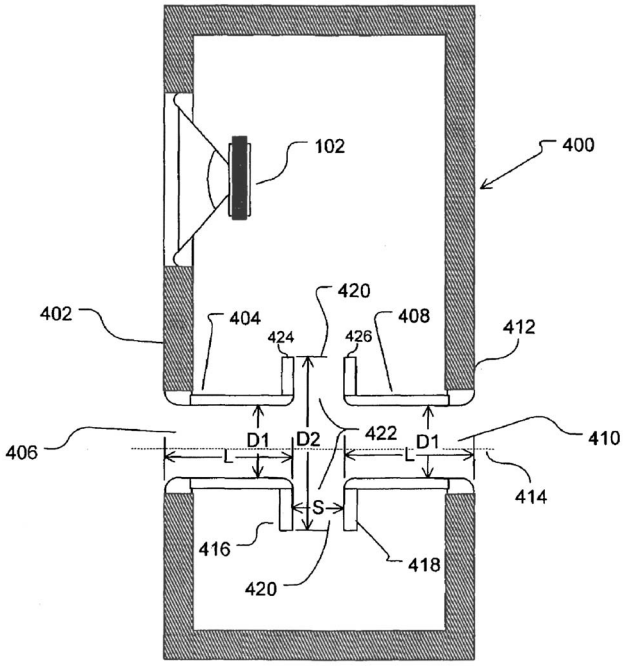
This clearly written patent discloses a method of designing triple-vented, dual-chamber enclosures for bass- and extended-frequency reproduction which can use a loudspeaker with about one-half the magnetic efficiency (and thus lower cost) as needed in an enclosure per United States Patent 4,549,631, by Bose, and has greater output over a broader operating band compared to the enclosures per United States Patent 4,875,546, by Krnan, and Japanese published application number 4-301998, by Tamura. This is done by selecting the ratio of the volumes and the ratios of the acoustic masses of the vents. The patent also states that a drone radiator, with acoustic and other characteristics similar to the vent mass, can be substituted for an external vent.—NAS

7,162,049

43.38.Ja PORTED LOUDSPEAKER SYSTEM AND METHOD WITH REDUCED AIR TURBULENCE, BIPOLAR RADIATION PATTERN AND NOVEL APPEARANCE

Matthew S. Polk, Jr., assignor to Britannia Investment Corporation
9 January 2007 (Class 381/345); filed 7 January 2003

When a vented loudspeaker system is operated at high levels, air turbulence in the vent can substantially degrade performance. The inventor has



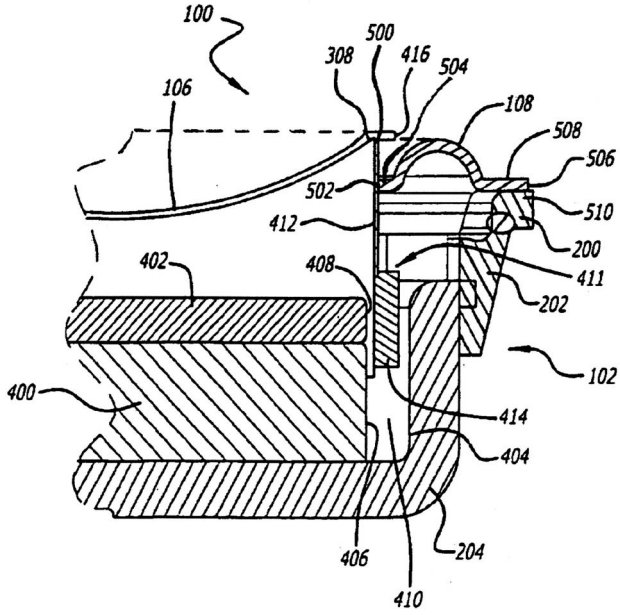
addressed this problem in an earlier patent. The new configuration shown here is said to minimize turbulence and also suppress pipe resonances in the vent.—GLA

7,167,573

43.38.Ja FULL RANGE LOUDSPEAKER

Clayton C. Williamson, assignor to Harman International Industries, Incorporated
23 January 2007 (Class 381/403), filed 2 April 2003

Small edge-driven domes have been used as tweeters for almost half a century. More recently, midrange units and even full-range loudspeakers have adopted this configuration. In the latter two cases, fairly large diaphragm excursions may be required. This patent argues that if the suspension **108** is attached to the voice coil former **412**, rather than to the edge of



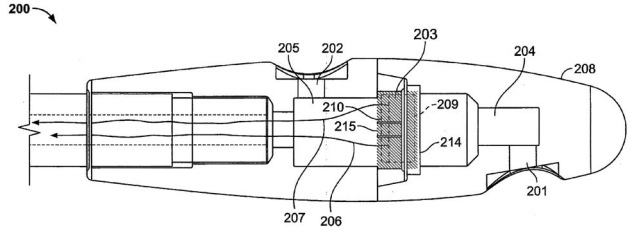
the diaphragm **106**, two advantages can be realized: first, tighter assembly tolerances permit longer excursions and second, spurious high-frequency resonances generated in the suspension are not transmitted directly to the diaphragm.—GLA

7,162,041

43.38.Kb NOISE CANCELING MICROPHONE WITH ACOUSTICALLY TUNED PORTS

Andrew J. Haapapuro and Viorel Drambarean, assignors to Etymotic, Research, Incorporated
9 January 2007 (Class 381/71.7); filed 27 September 2004

Simply stated, a typical noise-canceling microphone is a first-order gradient type whose front and back openings allow distant sounds (noise) to effectively cancel at the diaphragm, while desired signals (voice) are received largely through only the front opening. For the most uniform vocal



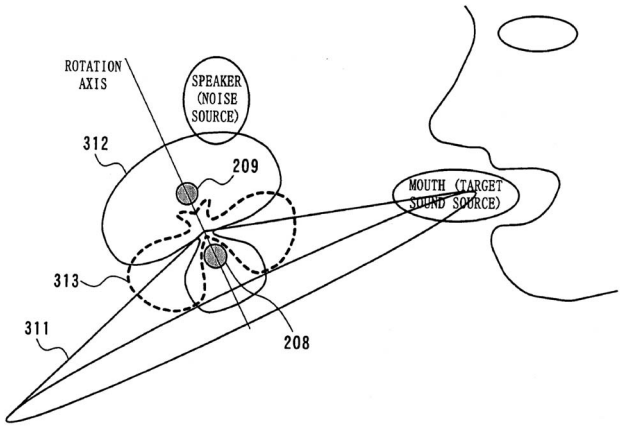
frequency response, the microphone should be positioned at a precise distance from the talker's lips. In this design the front and back openings on the outer shell can be rotated, repositioning them for optimum response at other operating distances.—JME

7,164,620

43.38.Kb ARRAY DEVICE AND MOBILE TERMINAL

Osamu Hoshuyama, assignor to NEC Corporation
16 January 2007 (Class 367/129); filed 7 April 2005

This patent deals with microphones in cell phones and the like. Two omnidirectional microphones are positioned so that one is close to the user's mouth, the other located at some distance. Through conventional delay processing, the two signals can be aligned and summed (beamformed) so that the desired signal is maximized. At the same time the two aligned signals



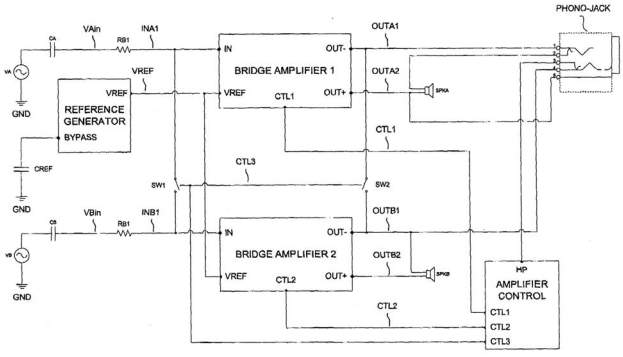
are subtracted and reprocessed to produce a new pattern with a response null in the noise component. The figure shows what is presumed to take place, showing the primary polar pattern **313** for voice pickup and the noise pickup pattern **312**.—JME

7,167,569

43.38.Lc OUTPUT COUPLING CAPACITOR FREE AUDIO POWER AMPLIFIER DYNAMICALLY CONFIGURED FOR SPEAKERS AND HEADPHONES WITH EXCELLENT CLICK AND POP PERFORMANCE

Kazim Seven, assignor to National Semiconductor Corporation
23 January 2007 (Class 381/111); filed 25 October 2000

Low-power audio amplifiers used to drive loudspeakers and headphones are normally powered by a single dc source, requiring a large coupling capacitor at the output. However, if two such amplifiers are connected



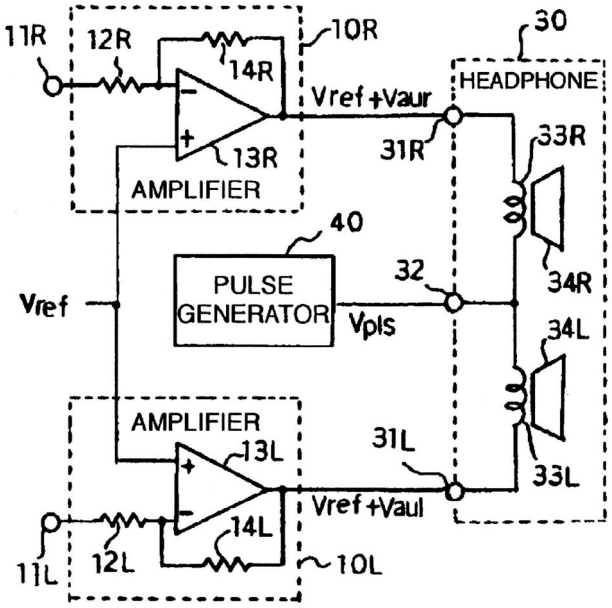
in a bridge configuration, then both outputs have the same dc offset and no coupling capacitor is needed. Unfortunately, the direct-coupled bridge has a tendency to produce loud transient signals during power-up. The circuitry described in this patent suppresses such transients and also allows the connection of loudspeakers or headphones to either or both of the two channels.—GLA

7,171,011

43.38.Lc ACOUSTIC DRIVE CIRCUIT

Masashi Morimoto, assignor to Oki Electric Industry Company, Limited
30 January 2007 (Class 381/120): filed in Japan 24 May 2002

This clever headphone amplifier circuit gets rid of output coupling capacitors through the action of pulse generator **40**. The pulse generator produces a rectangular waveform at an ultrasonic frequency, say, 2 MHz. The pulse duty ratio is adjusted to establish an average offset equal to the



reference voltage V_{ref}. The inductive reactance of voice coils **33R** and **33L** prevents more than a trickle of rf current from actually flowing through the coils.—GLA

7,170,822

43.38.Pf LAMINATED PIEZOELECTRIC TRANSDUCER AND METHOD OF MANUFACTURING THE SAME

Jerry Peck, assignor to Undersea Systems International, Incorporated
30 January 2007 (Class 367/165); filed 7 October 2004

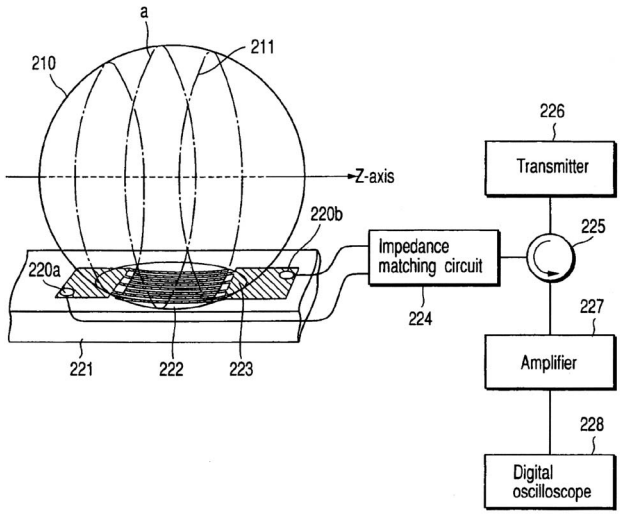
This patent proposes an underwater microphone to mitigate the impedance mismatch caused by pressure change with water depth. Additionally, the author maintains that the microphone is tunable to a given face mask, helmet, or headgear of choice.—GFE

7,170,213

43.38.Rh SURFACE ACOUSTIC WAVE ELEMENT, ELECTRIC SIGNAL PROCESSING APPARATUS USING THE SURFACE ACOUSTIC WAVE ELEMENT, ENVIRONMENT EVALUATING APPARATUS USING THE ELECTRIC SIGNAL PROCESSING APPARATUS, AND ANALYZING METHOD USING THE SURFACE ACOUSTIC WAVE ELEMENT

Kazushi Yamanaka *et al.*, assignors to Toppan Printing Company, Limited
30 January 2007 (Class 310/313 R); filed in Japan 9 October 2001

The authors claim a surface acoustic wave (SAW) device to inspect a spherical (or near-spherical) object **210** by projecting a surface wave into said object. A single-crystal piezoelectric substrate **221** is used, the face of



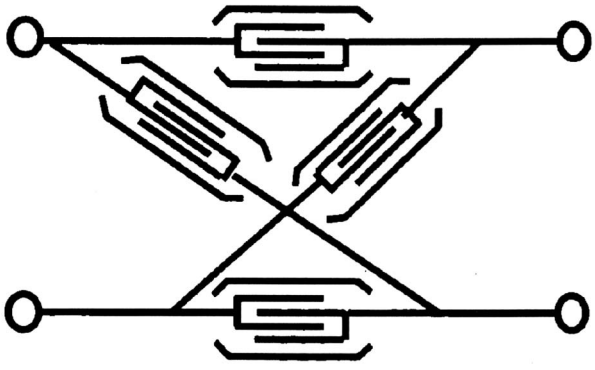
which is concave to match that of body **210**. An alternative SAW generation method using a modulated laser beam that produces modulated thermo-elastic heating directly within the surface of object **210** is also claimed.—AJC

7,170,371

43.38.Rh SURFACE ACOUSTIC WAVE DEVICE WITH A THICKER PARTIAL BUS BAR AREA AND OPTIMAL BUS BAR TO ETECTRODE TIP GAP

Masanori Ueda *et al.*, assignors to Fujitsu Media Devices Limited
30 January 2007 (Class 333/195); filed in Japan 23 April 2003

The authors claim a rf lattice filter arrangement with specified SAW interdigital transducer finger dimensions.—AJC

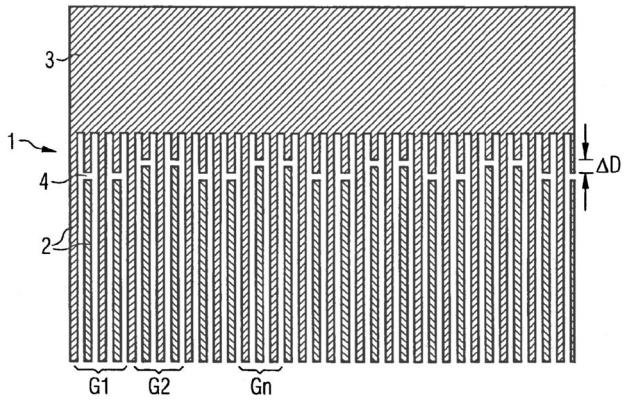


7,170,372

43.38.Rh CONVERTER FOR SURFACE WAVES WITH IMPROVED SUPPRESSION OF INTERFERING EXCITATION

Werner Ruile *et al.*, assignors to EPCOS AG
30 January 2007 (Class 333/196); filed in Germany 24 July 2001

The authors claim an arrangement of fingers **2** of SAW rf filter **1** that will suppress the passage of out-of-band interfering rf signals by randomly varying the distance of insulating gap **4** from bus bar **3** over distance **ΔD**



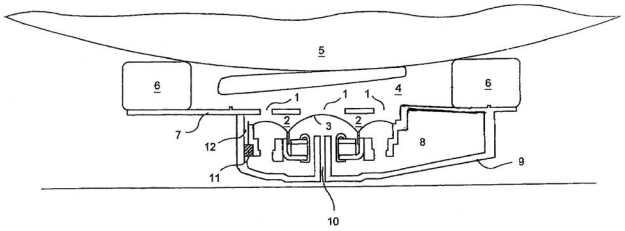
that is within the limit of the size of the distance between two or three fingers. Several slot shapes and distances of slot **4** from bus bar **3** are claimed.—AJC

7,162,051

43.38.Si HEADPHONE

Axel Grell and Konrad Hass, assignors to Sennheiser electronic GmbH & Company KG
9 January 2007 (Class 381/372); filed in Germany 21 October 2002

This patent relates to a type of headphone design that includes a damped vent (acoustic resistor) connecting the front **4** and rear **8** cavities. If such a vent is partially blocked by the ear then response is affected. In the



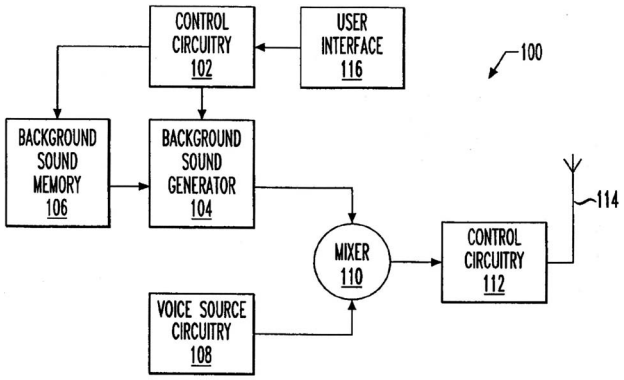
improved design shown here, the vent **11** exits beneath sound openings **1** and is therefore unaffected by changes in the fit of the headphone.—GLA

7,162,212

43.38.Si SYSTEM AND METHOD FOR OBSCURING UNWANTED AMBIENT NOISE AND HANDSET AND CENTRAL OFFICE EQUIPMENT INCORPORATING THE SAME

David J. Bennetts *et al.*, assignors to Agere Systems Incorporated
9 January 2007 (Class 455/114.2); filed 22 September 2003

Suppose that a salesman calls his boss to inform her that he is stuck in traffic and will be late for an important meeting. He may not be believed if she hears music and laughter in the background. This patent describes a



method for storing and electronically injecting random noise, background music, or any other desired masking signal into the communication channel at the command of the user.—GLA

7,171,193

43.38.Si TELECOMMUNICATIONS INTERRUPTION AND DISCONNECTION APPARATUS AND METHODS

Lawrence Andrew Hoffman, assignor to The Hoffman Group LLC
30 January 2007 (Class 455/417); filed 22 March 2004

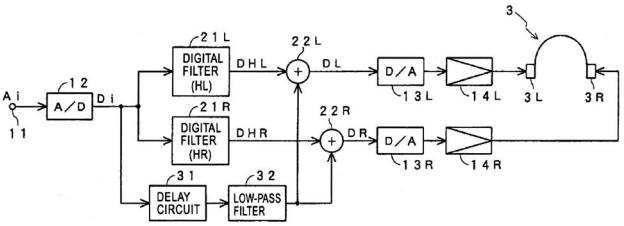
This patent describes a device that falsifies a ''losing connection'' sound so that a person can politely disconnect from an endless/painful telephone conversation. No more faking walkie-talkie hissing noises.—GFE

7,162,047

43.38.Vk AUDIO REPRODUCING APPARATUS

Yuji Yamada and Koyuru Okimoto, assignors to Sony Corporation
9 January 2007 (Class 381/309); filed in Japan 18 March 2002

With sophisticated digital processing it is possible to model interaural crosstalk and head-related transfer functions well enough to create convincing virtual sound sources surrounding the head of a headphone wearer. The processing is computationally intensive and a number of earlier patents have



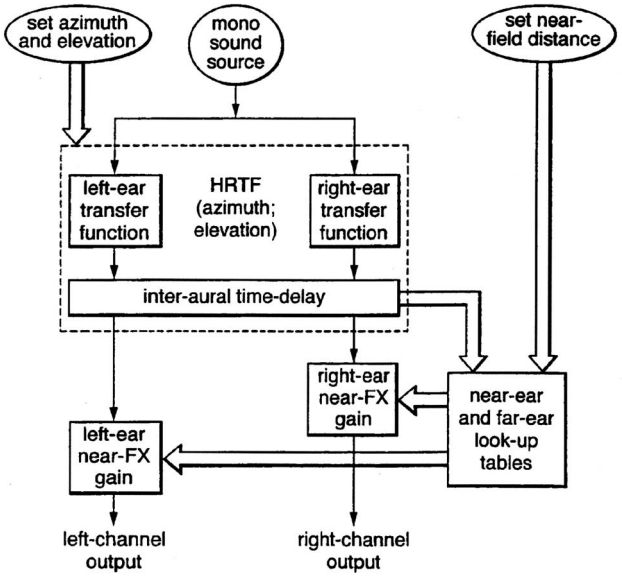
described methods for simplifying the calculations. In the patent at hand, low-frequency information is fed through a side chain consisting of delay **31** and low-pass filter **32**. Since the main FIR filters **21L** and **21R** handle a restricted bandwidth, the number of taps can be greatly reduced.—GLA

7,167,567

43.38.Vk METHOD OF PROCESSING AN AUDIO SIGNAL

Alastair Sibbald *et al.*, assignors to Creative Technology Limited
23 January 2007 (Class 381/17); filed in United Kingdom 13 December 1997

The patent outlines procedures for positioning a mono sound source for binaural presentation at a chosen spatial distance and azimuth angle



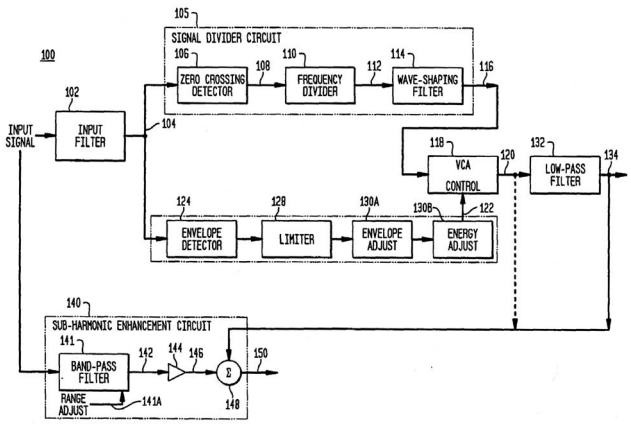
about a listener. The presentation may be via headphones or with loudspeakers via HRTFs and crosstalk cancellation. The process may have application in video games and other related areas.—JME

7,171,002

43.38.Vk METHODS AND APPARATUS FOR SUB-HARMONIC GENERATION, STEREO EXPANSION AND DISTORTION

Elon Ray Coats and Earnest Lloyd Trammell, assignors to Peavey Electronics Corporation
30 January 2007 (Class 381/1); filed 4 October 2004

The patent describes variations on the ''boombox'' of the 70s, which was a method of subharmonic generation for extending the range of mid-low frequencies downward by an octave. The abstract reads: ''Methods and ap-



paratus are disclosed that achieve subharmonic signal processing, Stereo-width expansion, subwoofer signal processing, and tube distortion emulation to achieve various desirable acoustic effects when used to modify an input signal containing, for example, music content.''—JME

7,168,890

43.40.Tm ECCENTRIC VIBRATION SYSTEM WITH RESONANCE CONTROL

Kingsley S. Evarts, assignor to American Piledriving Equipment, Incorporated
30 January 2007 (Class 405/232); filed 19 January 2005

In vibratory pile driving, oscillatory forces that act along the axis of the pile are generated by counter-rotating eccentric weights. As these weights are brought up to speed or slowed down they may produce undesirable resonant vibrations in the drive assembly. The present patent describes a drive assembly that overcomes this problem by including a second set of counter-rotating weights in the drive assembly. The phase of this second set is adjusted relative to that of the first so that the forces produced by the second set cancel those from the first set during start-up and slowdown, and so that they add to those from the first set during steady operation.—EEU

7,164,251

43.40.Vn OSCILLATION ADJUSTER AND OSCILLATION ADJUSTING METHOD

Mimpei Morishita, assignor to Toshiba Elevator Kabushiki Kaisha
16 January 2007 (Class 318/727); filed in Japan 12 March 2002

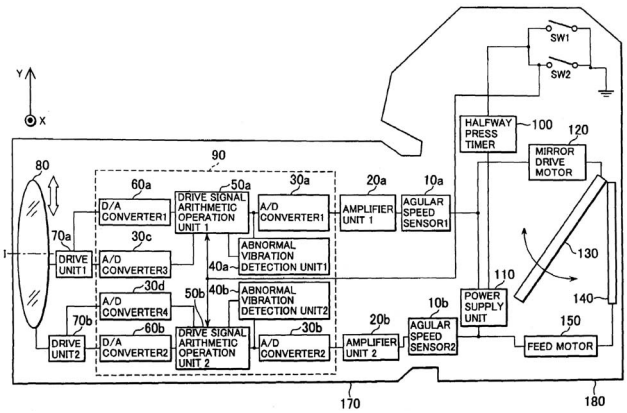
This patent pertains to the suppression of vibrations in elevators and the like, where eccentricity or imbalance in the rotating drive element tends to induce vibrations in the parts (e.g., elevator cabs) that are driven via cables. The oscillations of the cable-supported elements are sensed and are employed via a controller to induce counter-acting rotational oscillations in the drive system.—EEU

7,164,853

43.40.Vn VIBRATION DETECTION DEVICE AND VIBRATION CORRECTING OPTICAL DEVICE

Hiroyuki Tomita, assignor to Nikon Corporation
16 January 2007 (Class 396/55); filed in Japan 25 January 2002

The author claims an angular-vibration camera stabilization system that accommodates panning. Angular motion sensors **10a** (*X* axis) and **10b** (*Y* axis) provide the vibration signals. These signals are fed to servo amplifiers **20a** and **20b**, then to digital arithmetic units **50a** and **50b**. Servomotors **70a** and **70b** drive a compensation lens **80** that properly modifies the optical



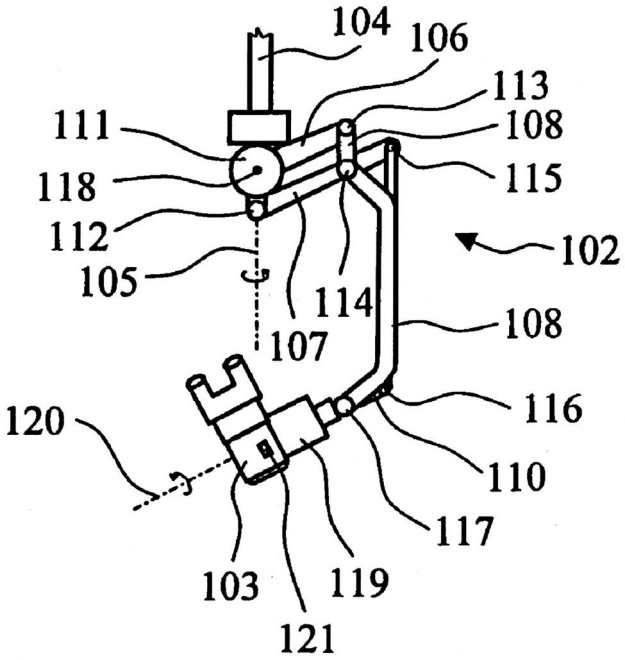
path to film (or CCD). Either of two compensation modes is logically determined by logic **40a** and **40b**; normal (still photography) and pan. A half press of trigger button **SW1** activates the system for stabilization. A full trigger-button press causes **SW2** to close and the photograph to be taken. In a Pentaprism camera, mirror **130** is pulled aside prior to film **140** exposure.—AJC

7,170,250

43.40.Vn HOLDING ARRANGEMENT HAVING A DEVICE FOR ACTIVELY DAMPING VIBRATION

Roland Brenner, assignor to Carl Zeiss Surgical GmbH
30 January 2007 (Class 318/560); filed in Germany 30 June 2003

The author claims a system **102** to stabilize surgical microscope **103**. The system is suspended via ceiling mount **104** and contains pivot **111** with servomotor for axis **118** to control the height via arm **106**. Pivot **119**, also with a servomotor for axis **120**, carries microscope **103**. One or more accel-



erometer sensors **121** on microscope **103** sense the vibration of **103** to provide signals to processing units (not shown) to support the microscope in any fixed position while countering vibrations that may occur. Processing schemes include accommodation of changing accessory weight loads and position shifting speeds required by the operator.—AJC

7,171,292

43.40.Vn VEHICLE POWERTRAIN TORSIONAL PROCESSING SYSTEM

Steven M. Bellinger and David M. Carey, assignors to Cummins, Incorporated
30 January 2007 (Class 701/29); filed 23 March 2004

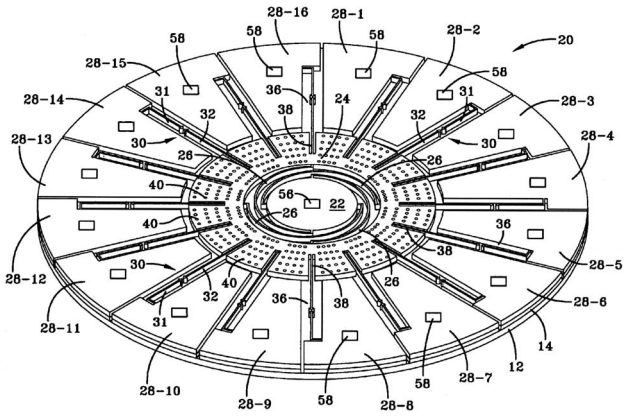
The **N**th-order torsional component of a vehicle's power train vibration is monitored in relation to the engine's rotational speed and compared to a threshold value. A control computer may modify the output torque of the engine to reduce the aforementioned component to below the threshold value.—EEU

7,159,442

43.40.Yq MEMS MULTI-DIRECTIONAL SHOCK SENSOR

Daniel J. Jean, assignor to The United States of America as represented by the Secretary of the Navy
9 January 2007 (Class 73/12.01); filed 6 January 2005

A shipping shock sensor is claimed where mass **24** on springs **26** is free to move in its plane. Latches **30** between arms **31–32** engage when



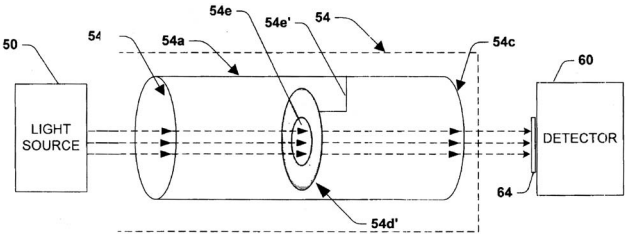
acceleration greater than a predetermined level occurs. Latch levels between 5 and 30 000 g can be accommodated. Readout occurs after retrieval by observing latched arms **30**. Electrical readout is possible by making posts **28** insulated and by considering each latched arm as a switch.—AJC

7,168,323

43.40.Yq SYSTEM AND METHOD FOR OPTICAL VIBRATION SENSING

Frederick M. Discenzo, assignor to Rockwell Automation Technologies, Incorporated
30 January 2007 (Class 73/655); filed 28 September 1999

An explosion-proof vibration sensor immune to acoustic and electric noise is claimed using relative motion to modulate a light beam. A light beam from fixed source **50** passes through aperture **54e**, whose carrier is attached to the machine under observation, to arrive at fixed detector **60**. If structure **54** displaces via vibration, the light intensity reaching the detector



60 is changed or modulated. Such modulations are processed for vibration frequency, amplitude, and direction. Processor memory includes historical vibration analysis data. The processor includes a display for the user and methods to alternate between various sets of display information and to send out alarm notifications.—AJC

7,168,324

43.40.Yq VIBRATION INFORMATION TRANSMISSION APPARATUS AND VIBRATION MONITORING/ANALYZING SYSTEM

Shingo Boda and Hiroyuki Maeda, assignors to Shinkawa Sensor Technology, Incorporated
30 January 2007 (Class 73/660); filed 12 August 2002

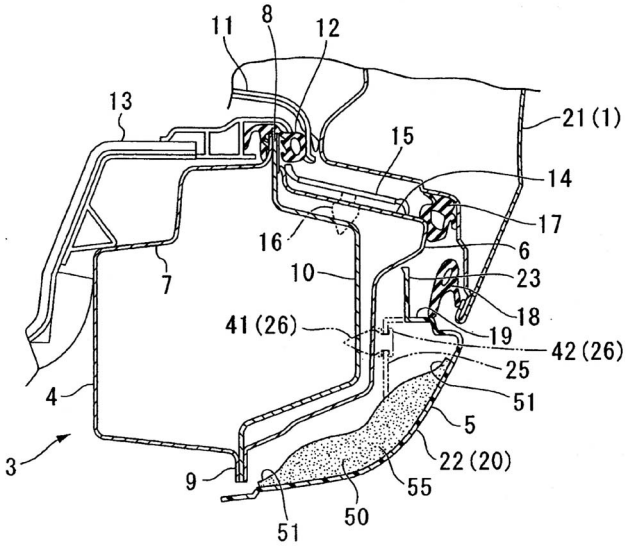
A composite vibration monitoring scheme is claimed where sensor outputs are multiplexed to reduce wiring needed to feed an analysis and display processor. There is little novelty here except for creating a compact arrangement.—AJC

7,168,757

43.50.Gf SIDE SILL GARNISH HAVING SOUND ABSORBING MEMBER WHICH INCLUDES OUTER PERIPHERAL RIGID PORTIONS

Akira Futatsuhashi, assignor to Honda Motor Company, Limited
30 January 2007 (Class 296/209); filed in Japan 28 September 2004

The author claims a vehicle-door-seal sound attenuator **3** (against exterior sound penetration) fashioned into the garnish **5** (''rocker panel''). Vehicle door **21** has a seal **18** to garnish **5**. The vehicle floor is **13** and the body frame is **10**. A sound leak past seal **8** and water barrier **23** will enter the



inner cavity of **5** to pass over absorber **50, 55**. Foam sound absorber **50, 55** is compressed at its edges to form mounting flanges **51**. The weep hole by body flange **9** is apparently small enough to not be of concern.—AJC

7,170,247

43.50.Gf METHOD OF CONTROL OF MAGNETIC-SOUND OF ALTERNATING CURRENT ROTATING MACHINE

Soichi Yoshinaga and Kenichi Wakabayashi, assignors to Denso Corporation
30 January 2007 (Class 318/448); filed in Japan 14 April 2004

The sound-radiating radial vibrations of the rotor and stator of an electric motor are reduced by adding to the driving current several smaller higher-frequency currents properly phased so that their vector sum results in reduction of the radial vibrations. Open-loop control can be used, employing previously determined magnitudes and phases of the noise-reducing currents.—EEU

7,163,962

43.55.Ev MORE CONTROLLABLE ACOUSTIC SPRAY PATCH

John R. Woods, assignor to Spraytex, Incorporated
16 January 2007 (Class 516/11); filed 9 June 2003

This admixture is an applied patch for a damaged acoustic ceiling. It has a popcorn-like finish and is sprayed from a pressurized container.—CJR

7,171,245

43.55.Hy METHOD FOR ELIMINATING MUSICAL TONE FROM BECOMING WIND SHEAR SOUND

Jian-Hueng Chen *et al.*, assignors to Chunghwa Telecom Company, Limited
30 January 2007 (Class 455/570); filed 6 May 2004

Cell phones automatically mute what they judge to be wind noise, but they sometimes mistakenly mute streaming music downloads. The patent alters the streaming music file so as to prevent the phone from thinking that it is receiving wind noise.—GFE

7,162,045

43.55.Lb SOUND PROCESSING METHOD AND APPARATUS

Shigeki Fujii, assignor to Yamaha Corporation,
9 January 2007 (Class 381/94.2); filed in Japan 22 June 1999

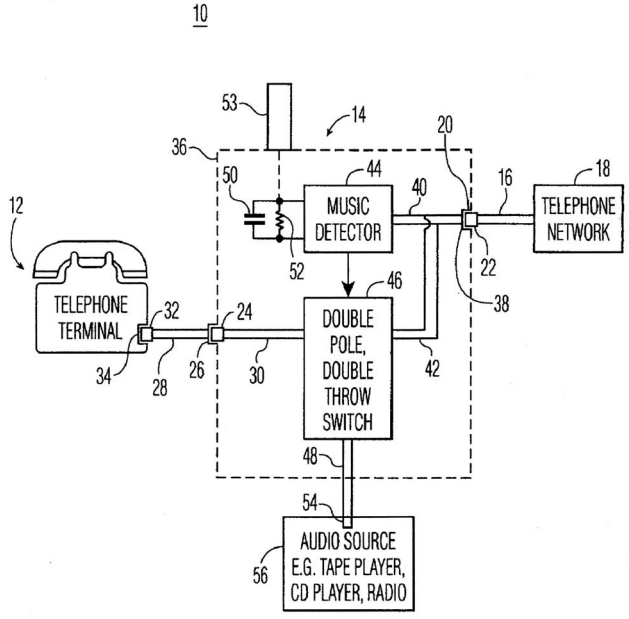
To more faithfully reproduce a listening experience in a concert hall, Yamaha is planning on processing different sound sources individually and combining the results at the end. For instance, the instruments, singers, speech, and even crowds might be subjected to different reflections and reverberation to create a more realistic-feeling hall or chamber. Currently, individual sources are combined and subjected to the same hall sound field, which can generate unnatural results.—GFE

7,136,479

43.60.Dh METHOD OF AND APPARATUS WHEREIN A TELEPHONE USER ON HOLD IS NOT REQUIRED TO LISTEN TO AN OBJECTIONABLE AUDIO SIGNAL

Srinivas Gutta *et al.*, assignors to Koninklijke Philips Electronics N.V.
14 November 2006 (Class 379/266.01); filed 19 June 2002

Music detector **44** determines if music-on-hold from the telephone



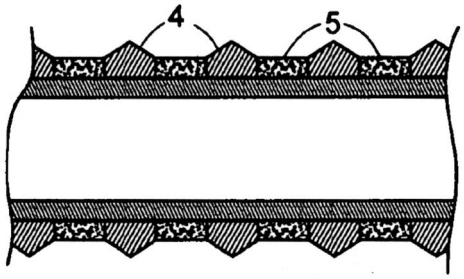
network **18** is present, and if so, operates switch **46** to select music of your preference, or even silence, from audio source **56**.—NAS

7,140,973

43.60.Dh PUTTER GRIP WITH IMPROVED VIBRATION TRANSMISSION TO HANDS

John W. Rohrer, assignor to Rohrer Technologies, Incorporated
28 November 2006 (Class 473/300); filed 8 December 2003

To improve the tactile feedback in a putter, the voids in handle grip **5** are filled with rigid vibration-transmitting protrusions **4**. The patent notes



that this is the opposite from what a driving club grip offers, i.e., enhanced vibration damping.—NAS

7,167,568

43.60.Mn MICROPHONE ARRAY SIGNAL ENHANCEMENT

Henrique S. Malvar *et al.*, assignors to Microsoft Corporation
23 January 2007 (Class 381/66); filed 2 May 2002

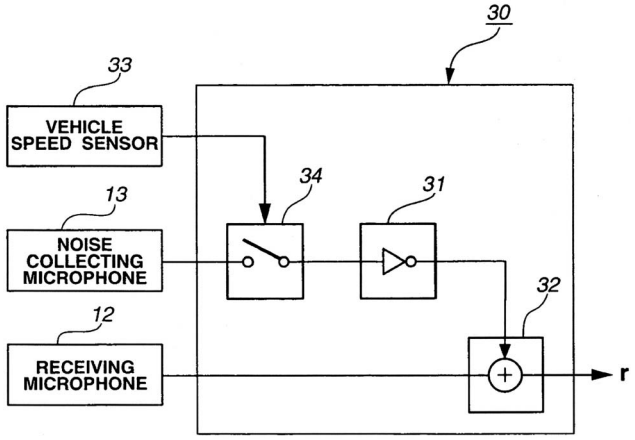
Under its broad title, this patent covers a dazzling universe of signal-processing techniques ostensibly for use with personal computers. The abstract opens as follows: ''A system and method facilitating signal enhancement utilizing an adaptive filter is provided. The invention includes an adaptive filter that filters an input based upon a plurality of adaptive coefficients and modifies the adaptive coefficients based on a feedback output.'' If you can make it this far, then you have a chance of discovering what this patent is all about.—JME

7,171,006

43.60.Mn VOCAL SOUND INPUT APPARATUS FOR AUTOMOTIVE VEHICLE

Makoto Kimura, assignor to Nissan Motor Company, Limited
30 January 2007 (Class 381/86); filed in Japan 7 January 2003

The patent deals with voice pickup in the cab of an automobile. For normal operation of the vehicle there may be adequate signal-to-noise ratio



for normal voice levels. When the vehicle is operated at higher speeds, motor and road noises may reduce the **S/N** ratio. This can be counteracted to

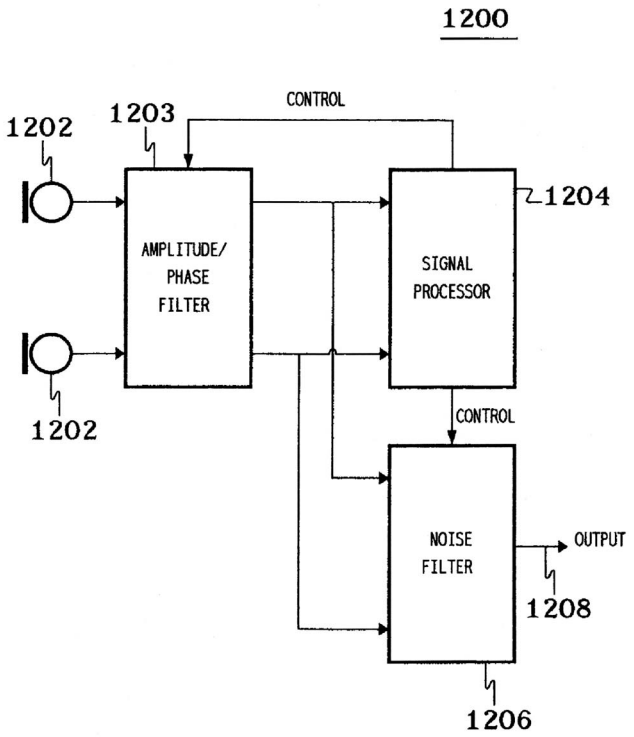
some degree by picking up the interfering noises, processing them, and then using them to cancel noise components in the communication path.—JME

7,171,008

43.60.Mn REDUCING NOISE IN AUDIO SYSTEMS

Gary W. Elko, assignor to MH Acoustics, LLC
30 January 2007 (Class 381/92); filed 12 July 2002

This interesting patent addresses the problems of wind noises often encountered in outdoor microphone applications. Traditionally, mechanical filters, such as screens, foam pop-filters, and the like have been used to reduce these noises at the point of origin. The patent suggests that these measures may not be necessary if the two monopole elements in a first-order



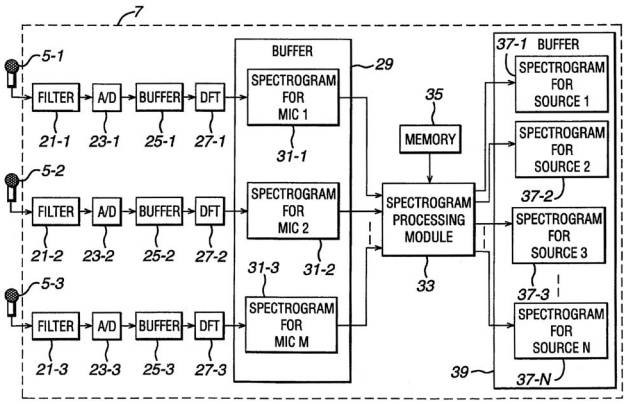
microphone array can be addressed separately, and their outputs analyzed in terms of the wind components (which have much lower propagation speed than audio components). As the abstract states: ''Time-dependent phase and amplitude differences between the microphones can be compensated for based on measurements made in parallel with routine audio systems processing.''—JME

7,171,007

43.60.Mn SIGNAL PROCESSING SYSTEM

Jebu Jacob Rajan, assignor to Canon Kabushiki Kaisha
30 January 2007 (Class 381/92); filed in United Kingdom 7 February 2001

The patent deals with identifying vocal sound sources in a typical conferencing setup. A beamforming microphone array is used to identify the



directions of each talker, along with vocal spectral signatures of each talker.—JME

7,171,009

43.60.Mn METHOD OF CORRECTING SOUND FIELD IN AN AUDIO SYSTEM

Yoshiki Ohta, assignor to Pioneer Corporation
30 January 2007 (Class 381/98); filed in Japan 14 February 2000

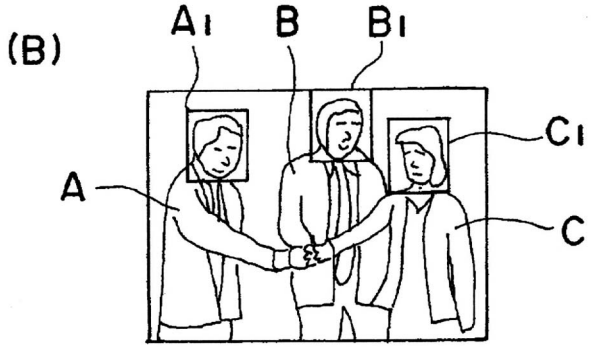
This is one more in a long line of patents that addresses the problems of setting up and calibrating home surround sound systems. Relative loudspeaker levels, equalization, and delay are all rationalized via the system.—JME

7,035,418

43.66.Qp METHOD AND APPARATUS FOR DETERMINING SOUND SOURCE

Hiroshi Okuno *et al.*, assignors to Japan Science and Technology Agency
25 April 2006 (Class 381/310); filed in Japan 11 June 1999

Using this system would involve setting up a multiple-microphone audio recording system as well as a video recorder. Intended for use during a conference or other such conversation, the system would be able to locate and track the specific individual who is speaking at a given moment. The



system seems to rely much more on audio, rather than video, processing to make the source decision, using both multimicrophone beamforming techniques as well as tracking the voice characteristics. Once a particular talker has been located, the video signal can be modified so as to highlight the speaking individual on a display monitor.—DLR

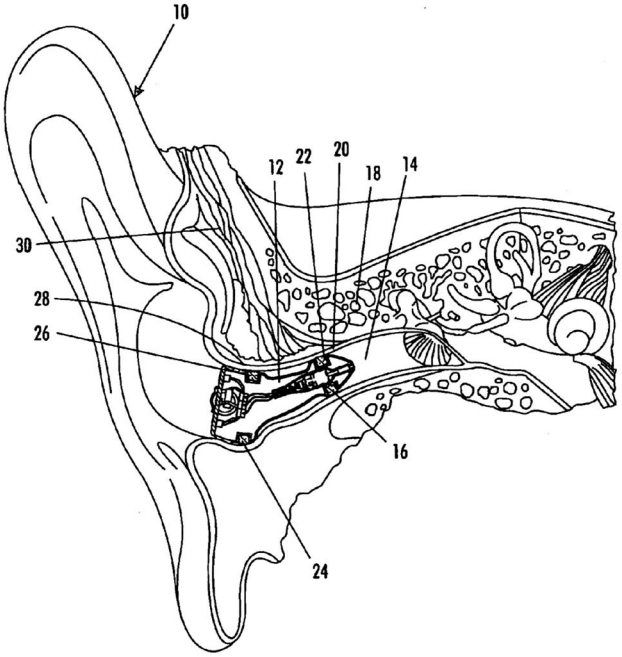
7,164,775

43.66.Ts IN THE EAR HEARING AID UTILIZING ANNULAR RING ACOUSTIC SEALS

John A. Meyer, Fairport and Andrew F. Morabito, Rochester, both of New York

16 January 2007 (Class 381/322); filed 18 November 2004

An acoustic feedback path from hearing aid receiver to microphone, which can produce an oscillatory condition, is created from slit air leaks occurring between the ear canal, which changes shape during jaw motion,



and the hard-shelled body of the hearing aid. Two compliant, elastomer rings **16**, **24** mounted into grooves around the hearing aid body create an acoustic seal and prevent acoustic feedback.—DAP

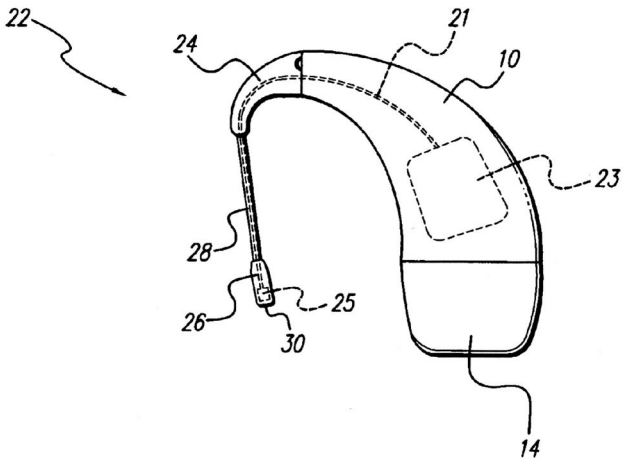
7,167,572

41.66.Ts IN THE EAR AUXILIARY MICROPHONE SYSTEM FOR BEHIND THE EAR HEARING PROSTHTETIC

William Vanbrooks Harrison *et al.*, assignors to Advanced Bionics Corporation

23 January 2007 (Class 381/330); filed 14 April 2004

An in-the-ear microphone location is known to have superior acoustic pickup of telephone signals as compared to an over-the ear microphone location. Consequently, the microphone pickup of a behind-the-ear cochlear implant processor is located within the concha of the wearer's ear via a formable stalk extension from the normal earhook that may be made optionally removable with a connector. When a telephone handset is held over the



wearer's ear, an acoustic seal is said to be formed, which improves low-frequency sound pickup from the telephone.—DAP

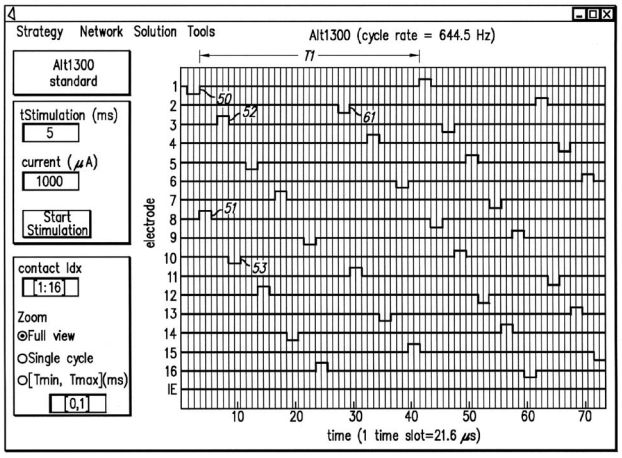
7,167,754

43.66.Ts LOW POWER INVERTED ALTERPHASIC STIMULATION IN A COCHLEAR IMPLANT

Stefaan Peeters and Filiep Vanpoucke, assignors to Advanced Bionics Corporation

23 January 2007 (Class 607/57); filed 19 November 2004

A stimulation strategy for a cochlear implant electrode array is said to reduce power required by inverting the polarity of sequential pulsed stimuli to two electrodes if the electrodes are spaced close together and, if the



electrodes are spaced far apart, the polarity of sequential stimuli to the two electrodes may or may not be inverted.—DAP

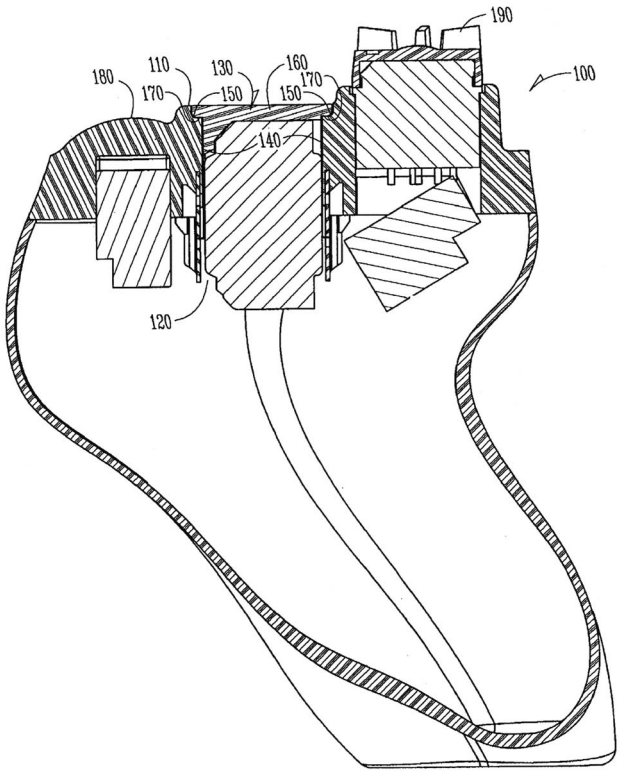
7,171,014

43.66.Ts HEARING AID BATTERY DOOR SEAL

Nelson Morales *et al.*, assignors to Starkey Laboratories, Incorporated

30 January 2007 (Class 381/322); filed 31 October 2002

Acoustic feedback from hearing aid receiver to microphone going through openings created between the battery door and faceplate can cause an oscillatory condition. To prevent this feedback from occurring, a lip projecting from three sides of the battery door makes contact with a stepped-



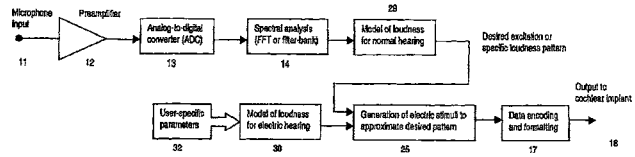
ridge or lap joint on a side wall of the faceplate, thus creating an air seal when the battery drawer is closed. The acoustic seal may be enhanced by placing gaskets on the sides of the battery door and side walls of the faceplate.—DAP

7,171,272

43.66.Ts SOUND-PROCESSING STRATEGY FOR COCHLEAR IMPLANTS

Peter John Blamey *et al.*, assignors to University of Melbourne
30 January 2007 (Class 607/57); filed in Australia 21 August 2000

Currently, the relative loudness of electrically stimulated signals that a cochlear implant wearer perceives may not be the same as what normal-hearing persons hear for the same acoustic input. To improve the control of loudness and to enhance speech intelligibility, a selective-weighting shaping algorithm is applied to the amplitudes of the cochlear implant channels. The



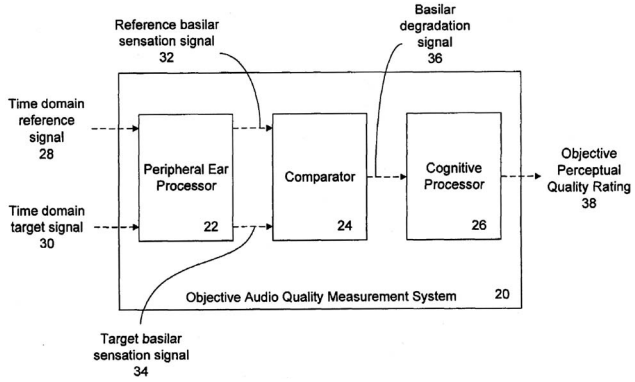
algorithm is based on the relative importance of each channel for speech perception by normal-hearing listeners. Additionally, overall loudness estimates in response to certain stimuli are made for the implant user and are adjusted to be approximately the same as those for normal-hearing persons.—DAP

7,164,771

43.71.Gv PROCESS AND SYSTEM FOR OBJECTIVE AUDIO QUALITY MEASUREMENT

William C. Treurniet *et al.*, assignors to Her Majesty the Queen as Represented by the Minister of Industry through the Communications Research Centre
16 January 2007 (Class 381/56); filed in Canada 27 March 1998

It is desirable to have a system for estimating the sound quality of an audio processing system without having to perform comparative juried listening tests. The proposed objective method first processes both reference



and target signals with a peripheral ear model and then uses a neural network in a cognitive processor to determine sound quality from the basilar signal degradation.—DAP

7,035,790

43.72.Ar SPEECH PROCESSING SYSTEM

Jebu Jacob Rajan, assignor to Canon Kabushiki Kaisha
25 April 2006 (Class 704/201); filed in United Kingdom 2 June 2000

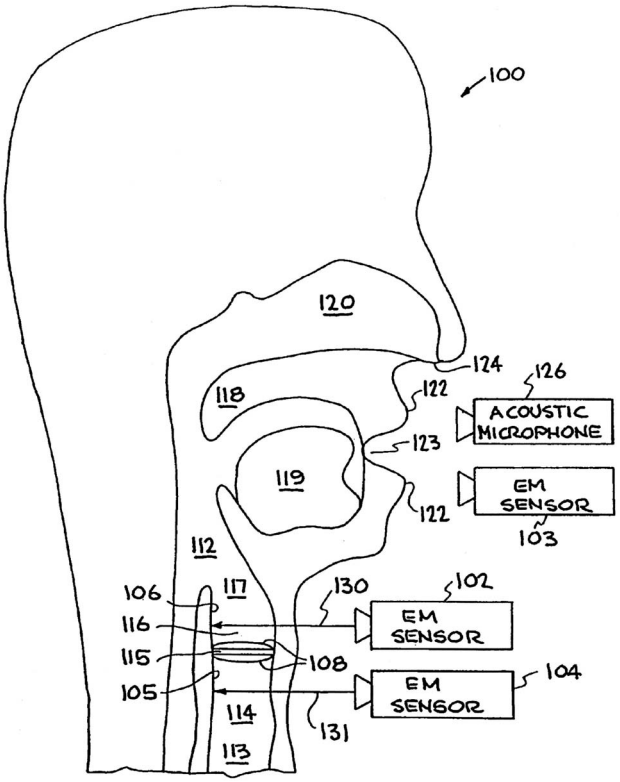
This speech analysis system would apply a variety of analysis methods to each nonoverlapping frame of the speech signal, including linear prediction, moving average, Kalman filtering, Bayes analysis, Gibbs sampling, and, in some cases, applying varying model orders to determine the best representation of the speech signal. The resulting parametrization could then be used to reconstruct a clean version of the signal, rejecting any noise determined to have been present in the original signal. All mathematical processes are clearly described in the well-written patent.—DLR

7,035,795

43.72.Ar SYSTEM AND METHOD FOR CHARACTERIZING VOICED EXCITATIONS OF SPEECH AND ACOUSTIC SIGNALS, REMOVING ACOUSTIC NOISE FROM SPEECH, AND SYNTHESIZING SPEECH

Greg C. Burnett *et al.*, assignors to The Regents of the University of California
25 April 2006 (Class 704/223); filed 8 October 2003

This speech recording and analysis system uses a type of radar detection to monitor and record from glottal structures during speech production. This information is used, together with the acoustic signal, to produce a



detailed analysis of the voicing characteristics. The detailed parameter data also allow the removal of noise from the recorded acoustic signal. The electromagnetic sensor used for vocal-tract motion recording is said to be able to track movements of less than a millimeter in magnitude.—DLR

7,165,033

43.72.Ar APPARATUS AND METHODS FOR DETECTING EMOTIONS IN THE HUMAN VOICE

Amir Liberman, Tzoran, Israel
16 January 2007 (Class 704/270); filed in Israel 12 April 1999

This patent is good for a little amusement, but trying to understand it is another thing altogether. It attempts to describe a "love detector" that uses relative amplitudes of the frequency bands in some sort of spectral

decomposition of the voice signal. Exactly what kind of spectral breakdown is involved here is still a trade secret, it seems.—SAF

7,165,026

43.72.Dv METHOD OF NOISE ESTIMATION USING INCREMENTAL BAYES LEARNING

Alejandro Acero *et al.*, assignors to Microsoft Corporation
16 January 2007 (Class 704/226); filed 31 March 2003

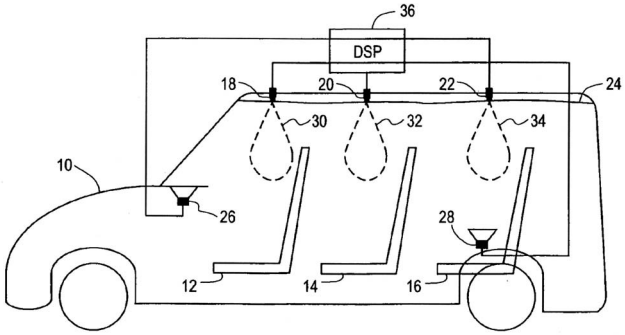
Typical nonlinear noise estimation techniques apply the well-known Bayes formula, compute statistics (mean and variance) of some "prior" noise garnered from a section of test signal known to consist entirely of the noise, and then seek to maximize either the likelihood or the posterior probability to estimate noise at each subsequent frame. This patent proposes to involve complete Bayesian updating, changing the prior noise statistics at each frame to estimate the noise at the current frame using a Gaussian approximation of data likelihood for the current frame and the updated Gaussian approximation of noise in a sequence of prior frames. The methods are sound and well-documented, as is typical of the Microsoft speech technology patents.—SAF

7,171,003

43.72.Ew ROBUST AND RELIABLE ACOUSTIC ECHO AND NOISE CANCELLATION SYSTEM FOR CABIN COMMUNICATION

Saligrama R. Venkatesh and Alan M. Finn, assignors to Lear Corporation
30 January 2007 (Class 381/66); filed 19 October 2000

Just in time for long family road trips, this patent describes a method to increase the clarity and amplification of speech between the driver and passengers. The author asserts that the main problems of in-car amplification are echo feedback and the variability of noise that occupies the bandwidth of human speech. Furthermore, due to the cramped quarters, the voice ampli-



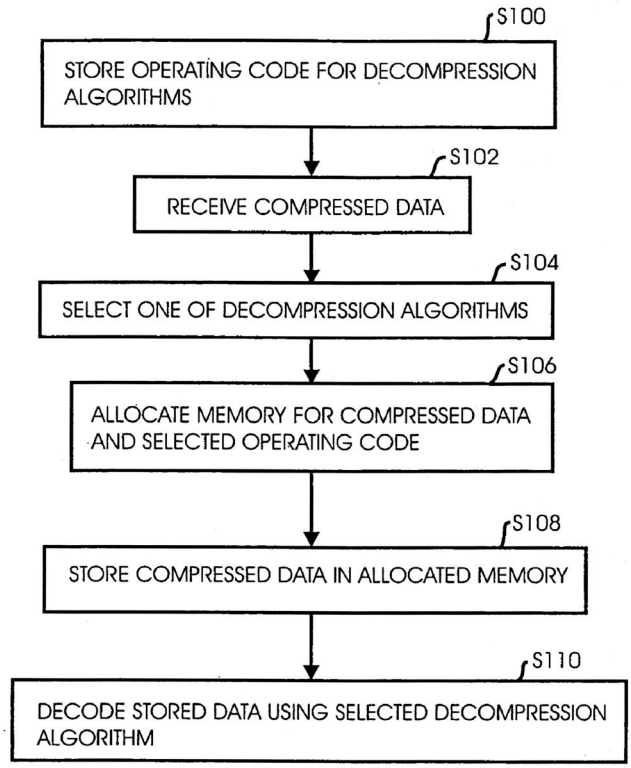
fication system must process the signal in under 20 ms or a perceivable delay will occur. As any back-seat driver can testify, the car is a low signal-to-noise environ; therefore, the authors put forth a processor based on Wiener filtering constrained by the psychoacoustics of speech.—GFE

7,162,416

43.72.Gy COMPRESSED AUDIO STREAM DATA DECODER MEMORY SHARING TECHNIQUES

Paul Morton and Darwin Rambo, assignors to Broadcom Corporation
9 January 2007 (Class 704/201); filed 12 September 2005

Various types of voice decoder algorithms require different amounts of memory to store operating data and code. To reduce the amount of memory required for holding data to be decoded, a processor dynamically allocates



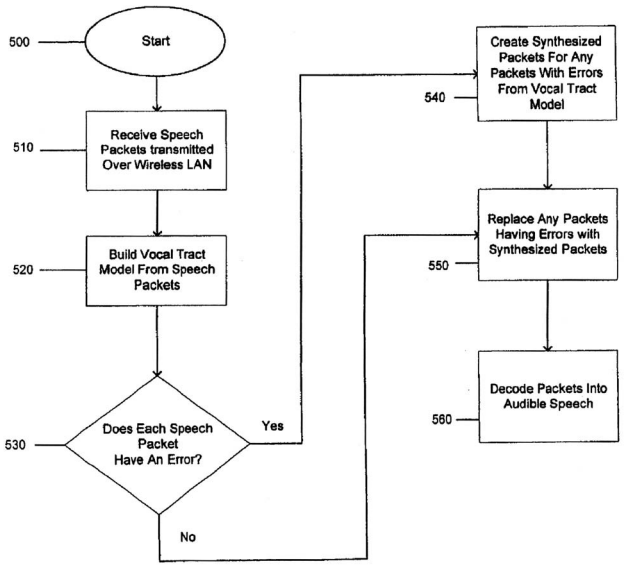
only the required amount of memory to store the compressed data and at least one decompression algorithm.—DAP

7,164,672

43.72.Gy METHOD AND APPARATUS FOR QOS IMPROVEMENT WITH PACKET VOICE TRANSMISSION OVER WIRELESS LANS

Richard Henry Erving and Robert Raymond Miller II, assignors to AT&T Corporation
16 January 2007 (Class 370/338); filed 29 March 2002

To improve the quality of wireless speech transmission, speech packets transmitted over a wireless LAN are monitored for errors by measuring the



signal quality. Any of the received packets found to have errors is replaced with synthesized speech packets created from a vocal-tract model generated from the speech packets that were received with no errors.—DAP

7,031,919

43.72.Ja SPEECH SYNTHESIZING APPARATUS AND METHOD, AND STORAGE MEDIUM THEREFOR

Yasuo Okutani and Masayuki Yamada, assignors to Canon Kabushiki Kaisha
18 April 2006 (Class 704/258); filed in Japan 31 August 1998

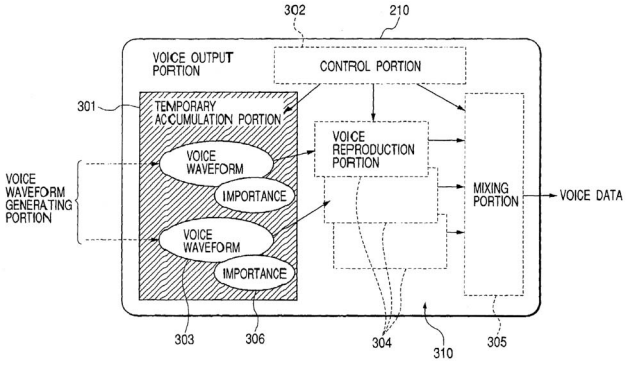
This method for speech synthesis uses a database of phonetic units, stored according to a triphone classification system, and with each rated according to pitch and amplitude values. So far, we have a typical prior-art method. Upon retrieval of a set of possible phonetic units for synthesis, a penalty scoring system allows the selection of the best match for the synthesis context. Several variations on the scoring system are elaborated in the claims.—DLR

7,031,924

43.72.Ja VOICE SYNTHESIZING APPARATUS, VOICE SYNTHESIZING SYSTEM, VOICE SYNTHESIZING METHOD AND STORAGE MEDIUM

Hiroyuki Kimura *et al.*, assignors to Canon Kabushiki Kaisha
18 April 2006 (Class 704/274); filed in Japan 30 June 2000

This is a fairly elaborate speech synthesis control system, organized to manage multiple speech output streams, while providing various speaker attributes to each of the speech streams. The patent text describes methods for controlling the speaker's gender and age, but the claims concentrate on



several methods for adjusting the output volume of each of several talkers when their voice streams are to be mixed. A system of marking the relative importance of each utterance allows a speaker-specific volume control to be adjusted so as to highlight one or the other message.—DLR

7,035,791

43.72.Ja FEATURE-DOMAIN CONCATENATIVE SPEECH SYNTHESIS

Dan Chazan and Ron Hoory, assignors to International Business Machines Corporation
25 April 2006 (Class 704/207); filed 10 July 2001

The patent describes a fairly typical system for speech synthesis by concatenation except that the synthesis reconstructions are done in a feature parameter domain instead of using time waveforms. Several feature types are discussed, with the primary type apparently being Mel-frequency cepstral coefficients. The unit of analysis and concatenation is the lefeme, a unit defined in an earlier IBM patent as "a portion of a phone in a particular left/right context." Judging by a Google search, the term has not caught on among other research groups. A major issue in concatenative synthesis is the nature of methods used to determine the synthesis pitch patterns. Here, the pitch profiles seem to be recovered from the stored segment database, but

this process is not described. The methods of pitch adjustment are, on the other hand, described in great detail.—DLR

7,035,794

43.72.Ja COMPRESSING AND USING A CONCATENATIVE SPEECH DATABASE IN TEXT-TO-SPEECH SYSTEMS

Sudheer Sirivara, assignor to Intel Corporation
25 April 2006 (Class 704/219); filed 30 March 2001

The patent covers a method for partitioning a concatenative speech synthesis system such that parts of the process can run on a server, with other parts running on the handheld client device. Diphone units are coded using the standard G.723 method, providing linear prediction coefficients and residuals for transmission to the client device.—DLR

7,162,417

43.72.Ja SPEECH SYNTHESIZING METHOD AND APPARATUS FOR ALTERING AMPLITUDES OF VOICED AND INVOICED PORTIONS

Masayuki Yamada *et al.*, assignors to Canon Kabushiki Kaisha
9 January 2007 (Class 704/214); filed in Japan 31 August 1998

This patent outlines a simplistic approach to separately altering the amplitude of voiced and unvoiced portions of synthesized speech. The objective is to improve overall quality of synthesized speech at a variety of volume levels, avoiding the excessive magnification of unvoiced noises that yield strange-sounding results. It is implied that amplified speech is better if the voiced sounds are able to be amplified more.—SAF

7,165,030

43.72.Ja CONCATENATIVE SPEECH SYNTHESIS USING A FINITE-STATE TRANSDUCER

Jon Rong-Wei Yi *et al.*, assignors to Massachusetts Institute of Technology
16 January 2007 (Class 704/238); filed 17 September 2001

This interesting document outlines a method for concatenative speech synthesis, which must, of course, concatenate small sound segments of some kind. The main improvement here is the search for a segment sequence, which is performed by finding a path through a finite-state transducer having source utterance segments as input labels and a pronunciation characterization as the output label. Each path is assigned a numerical score that characterizes the quality of the sequence of segments and the search for the best path can be performed with dynamic programming similar to a Viterbi search.—SAF

7,031,920

43.72.Ne LIGHTING CONTROL USING SPEECH RECOGNITION

Kevin J. Dowling and George G. Mueller, assignors to Color Kinetics Incorporated
18 April 2006 (Class 704/270); filed 26 July 2001

This patent describes a basic voice control system and an elementry slot grammar for the possible utterances. Actual recognition hardware is not specified. The grammar allows lighting to be controlled by groups and/or colors.—DLR

7,035,797

43.72.Ne DATA-DRIVEN FILTERING OF CEPSTRAL TIME TRAJECTORIES FOR ROBUST SPEECH RECOGNITION

Juha Iso-Sipila, assignor to Nokia Corporation
25 April 2006 (Class 704/227); filed 14 December 2001

The background section of this patent includes a fairly detailed description of a European standard for a front-end speech analysis procedure to be implemented in a mobile phone device, such as a cell phone, which allows the analysis feature vectors to be transmitted to a server-side recognition process. The patented analysis and transmission system is said to provide improved noise rejection by the addition of frequency-division filtering, relative spectral processing, and principal components analysis stages during the extraction of the Mel-frequency cepstral features.—DLR

7,162,414

43.72.Ne METHOD AND APPARATUS TO PERFORM SPEECH RECOGNITION OVER A DATA CHANNEL

Michael Stanford, assignor to Intel Corporation
9 January 2007 (Class 704/201); filed 7 December 2001

Due to bandwidth limitations, wireless voice transmission may provide lower than desired speech quality, which can reduce the performance of speech recognition systems. To improve transmitted voice quality, data transmission techniques rather than voice transmission is used.—DAP

7,162,421

43.72.Ne DYNAMIC BARGE-IN IN A SPEECH-RESPONSIVE SYSTEM

Torsten Zeppenfeld *et al.*, assignors to Nuance Communications
9 January 2007 (Class 704/233); filed 6 May 2002

A method is outlined for processing speech by a user during spoken prompts, normally called user ''barge-in.'' Typical systems allowing barge-in automatically pass control to the speech recognition section, turning off the voice prompts at that time. If the user actually did not say anything relevant (perhaps they were clearing their throat), the exit from the menu of prompts is not desirable. The patent proposes to instead remain with the spoken prompts, lowering the volume at barge-in detection, while running the user's utterance past the speech recognizer. The control will be passed and the prompt menu exited only if the recognition routine determines that the user has indeed said something relevant.—SAF

7,162,424

43.72.Ne METHOD AND SYSTEM FOR DEFINING A SEQUENCE OF SOUND MODULES FOR SYNTHESIS OF A SPEECH SIGNAL IN A TONAL LANGUAGE

Martin Holzapfel and Jianhua Tao, assignors to Siemens Aktiengesellschaft
9 January 2007 (Class 704/258); filed in Germany 26 April 2001

This patent puts some standard things into action and adds a few tweaks. According to the authors, tone languages are normally synthesized using sound modules which each represent one syllable. The patent proposes to adapt the well-known methods which use triphones to model segments smaller than a syllable. In this way, less-common syllables could be constructed out of subsyllabic segments, perhaps in combination with whole-syllable modules. A suitability function is also defined, which describes the compatibility of two adjacent sound modules for concatenation, and it is

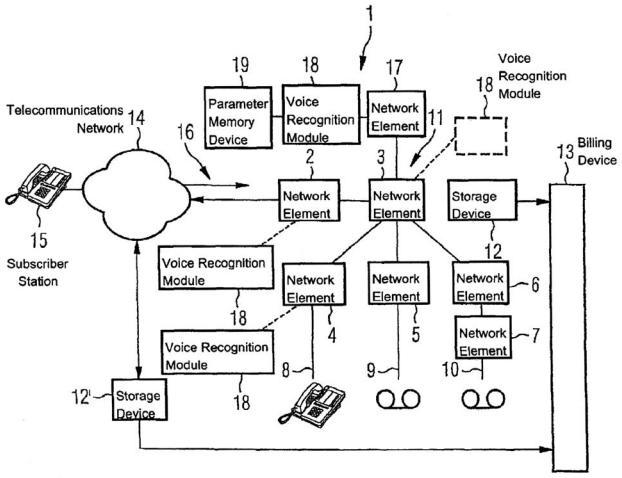
proposed that this function should be less stringent at a syllable boundary and more selective within one syllable.—SAF

7,164,754

43.72.Ne CONTROL OF ACCESS FOR TELEPHONE SERVICE PROVIDERS USING VOICE RECOGNITION

Marian Trinkel *et al.*, assignors to Deutsche Telekom AG
16 January 2007 (Class 379/88.02); filed in Germany 5 July 2000

Speech recognition is used to recognize certain biometric voice parameters for improving security against unauthorized access in a telecommunications network having many subscriber stations. A decision is made within a speech recognition module associated with a network element whether to establish a telecommunications connection from one subscriber station to



another. The decision is based on checking at least one biometric voice parameter against a previously stored record. If the decision made is not to connect, the network automatically connects the denied subscriber station to a network operator.—DAP

7,165,031

43.72.Ne SPEECH PROCESSING APPARATUS AND METHOD USING CONFIDENCE SCORES

David Llewellyn Rees, assignor to Canon Kabushiki Kaisha
16 January 2007 (Class 704/252); filed in United Kingdom 14 February 2002

This patent describes a technique for small-vocabulary speech recognition tasks in which a ''match score'' is computed for each hypothesized term generated by the attempted recognition of an utterance. These match scores are then converted into estimates of the likelihood of observing the signal x assuming that a word w was spoken. In this way, the posterior probability is computed of a recognition hypothesis being correct. The patent is sufficiently documented, although this method seems like a small variation on a very common theme.—SAF

7,171,357

43.72.Ne VOICE-ACTIVITY DETECTION USING ENERGY RATIOS AND PERIODICITY

Simon Daniel Boland, assignor to Avaya Technology Corporation
30 January 2007 (Class 704/231); filed 21 March 2001

This patent introduces a voice activation method based on energy ratio and pitch information. The method calculates two parameters in each successive frame of the signal: (1) the ratio of high-frequency (2400–4000 Hz) to low-frequency (100–2400 Hz) energy, and (2) the pitch period of the signal. The method also computes long-term average values for the above-mentioned parameters. The current frame of the signal is labeled voiced if (a) the difference between average energy ratio and current frame energy ratio exceeds a threshold; (b) the difference between average energy ratio and total energy of current frame exceeds a threshold; (c) the average total signal energy exceeds a minimum average total energy of the signal and voice has been detected in the previous frame; and (d) the average periodicity of the signal is lower than a minimum threshold.—AAD

7,171,360

43.72.Pf BACKGROUND LEARNING OF SPEAKER VOICES

Chao-Shih Huang *et al.*, assignors to Koninklijke Philips Electronics N.V.
30 January 2007 (Class 704/245); filed in the European Patent Office 10 May 2001

Speaker verification systems in security applications are getting more realistically successful, but are still annoying in that, in order to discriminate a reasonably large number of speakers reliably, a large number of utterances (e.g., 20) are required from each person during an initial enrollment. This patent outlines a way to circumvent this process while still providing essentially the same thing to the computer verification system. Listen to speakers blindly (perhaps in a government office setting), and enroll them *en masse*. It turns out, one never needs to know who is who until the process is completed. Once a large enough set of utterances has been collected from each speaker in this haphazard way, classification performance is naturally obtained. The patent shows that the resulting verification performance is competitive with standard systems requiring annoying enrollment setups.—SAF

7,171,362

43.72.Pf ASSIGNMENT OF PHONEMES TO THE GRAPHEMES PRODUCING THEM

Horst-Udo Hain, assignor to Siemens Aktiengesellschaft
30 January 2007 (Class 704/267); filed in Germany 31 August 2000

Training of neural nets for automatic speech recognition systems involves matching patterns from a number of letters in a word to the associated phoneme. Each phoneme is assigned a grouping value that specifies the number of graphemes which produce it. This invention discusses using dynamic time warping to produce patterns for training a neural network with automatic assignment of phonemes to the graphemes producing them.—DAP

7,161,081

43.75.Wx PORTABLE TELEPHONY APPARATUS WITH MUSIC TONE GENERATOR

Tsuyoshi Futamase *et al.*, assignors to Yamaha Corporation
9 January 2007 (Class 84/622); filed in Japan 11 January 1999

Hold on, my ''portable terminal apparatus'' is ''performing style information.'' The author of this Japanese patent maintains that your cellular phone will soon become a traveling karaoke machine that will play music as well as show corresponding text and animations. You can either purchase or make your own karaoke music on your PC and upload it to your phone, or directly make karaoke music on your phone. Some of the document also deals with making songs of your own via mixing simple chord progression tracks similar to the ones preinstalled on electric keyboards, such as ''ethnic'' and ''tropical.''—GFE

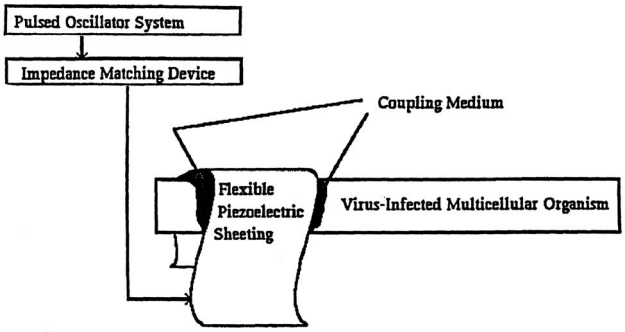
7,165,451

43.80.Sh METHODS FOR USING RESONANT ACOUSTIC AND/OR RESONANT ACOUSTO-EM ENERGY TO DETECT AND/OR EFFECT STRUCTURES

Juliana H. J. Brooks and Albert E. Abel, assignors to GR Intellectual Reserve, LLC

23 January 2007 (Class 73/579); filed 10 September 1999

The authors claim a plethora of acoustic and electromagnetic interactions in fluids such as blood, in living tissues, and in inanimate objects (e.g., explosives). A total of 40 claims includes determining resonances in various body organs that affect regeneration, growth, or healing, and the resonances of cells and growths that are pathological. For instance, the figure shows a method of exposing extra-body blood to ultrasound of frequencies that have



been found to cause resonance and disruption of a virus infecting the body. Tests show 25% for short duration and up to 100% ''disruption'' for long exposures. Typical ultrasound intensities applied are from 100 to 10 000 W/m^2. Other effects include the piezoelectric property of bones and the examination of packages that may contain explosives.—AJC

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.

Loudness reduction induced by a contralateral tone (L)[a)]

Bärbel Nieder
Institute of Hearing, Speech and Language and Department of Speech-Language Pathology & Audiology (106A FR), Northeastern University, Boston, Massachusetts 02115

Søren Buus[b)]
Institute of Hearing, Speech and Language and Communications & Digital Signal Processing Center, Department of Electrical & Computer Engineering, Northeastern University, Boston, Massachusetts 02115

Yves Cazals[c)]
INSERM EPI 9902, Laboratoire d'Otologie et de NeuroOtologie, Université d'Aix-Marseille II, Marseille, France

Bertram Scharf[d)]
Institute of Hearing, Speech & Language and Psychology Department (125 NI), Northeastern University, Boston, Massachusetts 02115

(Received 31 December 2006; revised 31 March 2007; accepted 3 April 2007)

The induced reduction in the loudness (ILR) of a weaker tone caused by a preceding stronger tone was measured with both tones in the same ear (ipsilateral ILR) and also in opposite ears (contralateral ILR). The two tones were always equal in duration and were presented repeatedly over several minutes. When the tone duration was 200 ms, for 24 listeners the loudness reduction averaged 11 dB under ipsilateral ILR and 6 dB under contralateral ILR. When the duration was 5 ms, ILR was 8 dB whether ipsilateral or contralateral. For each duration, ipsilateral and contralateral ILR were strongly correlated (r around 0.80).
[DOI: 10.1121/1.2735107]

PACS number(s): 43.66.Cb, 43.66.Pn [AJO] Pages: 35–37

I. INTRODUCTION

A number of recent papers have shown that the loudness of a weaker tone preceded in a sequence of trials by a stronger tone at the same frequency may be reduced on the average the equivalent of 10 dB (for a recent review of this induced loudness reduction or ILR, see Wagner and Scharf, 2006). All these studies measured ipsilateral ILR, that is, they presented the inducer and test tones to the same ear (or ears, in binaural presentations). The one known exception was a report by Marks (1996) who used magnitude estimation and loudness comparisons to measure the loudness reduction also when the stronger inducer and the weaker test tones were in opposite ears. The measured reduction was about half as great in the contralateral configuration as in the ipsilateral configuration. We now confirm and quantify that result in measurements based on adaptive loudness matching, which has been the paradigm used in most of the recent research on ILR. In addition, we show that 5 ms tones, unlike 200 ms tones, undergo just as much ILR contralaterally as ipsilaterally. Moreover, our measurements reveal, both for 200 and for 5 ms tones, very high correlations between ipsilateral and contralateral ILR.

II. METHOD

A. Procedure

Two sets of measurements were made in two laboratories, one set in Marseille, France, the other in Boston. The same computer programs and types of equipment were used in the two laboratories. Basic procedures were identical; differences are indicated where appropriate.

[a)]Portions of this work were presented at the 2003 meeting of the International Society for Psychophysics, Cyprus and at the 149th Meeting of the Acoustical Society of America in May 2005 [Scharf *et al.*, J. Acoust. Soc. Am. **114**, 2454(A) (2005)].
[b)]Søren Buus collaborated closely in the design and execution of these experiments. His illness and untimely death prevented him from participating in the preparation of the present paper.
[c)]Current address: CNRS UMR 6153, Lab Physiologie Neurovegetative, Université P. Cézanne, Avenue Normandie Niemen, 13397 Marseille, France.
[d)]Electronic mail: scharf@neu.edu

Listeners matched the loudness of 500 Hz test tones and 2500 Hz comparison tones in a two-interval, two-alternative forced-choice paradigm with interleaved adaptive tracks. In the *baseline* condition, only the test and comparison tones were presented. In the *experimental* condition, each test tone was preceded by a 500 Hz inducer tone, which was 10 dB higher than the higher-level test tone. Listeners were told to ignore the inducer. In both conditions, on each trial the listener indicated whether the test or the comparison tone was louder; the level of the comparison tone was varied according to the listener's response to determine the point of equal loudness. In any given session, baseline blocks always preceded experimental blocks.

Each trial comprised two observation intervals. The 500 Hz test tone was always presented in the first interval, usually to the right ear, at either 60 or 70 dB SPL (65 or 75 dB in Marseille). [To avoid the complication of describing an additional set of data and to save space, this paper gives the data only for the higher-level test tone which is unaffected by the presence of the weaker test tone as Nieder *et al.* (2003) suggested and as unpublished experiments from our laboratory fully confirmed.] The 2 500 Hz comparison tone was presented in the second interval. The level of the comparison tone was adjusted according to a simple up-down method (Jesteadt, 1980). Each block of trials comprised four interleaved adaptive tracks, two for the lower-level (60 or 65 dB) test tone and two for the higher-level (70 or 75 dB) test tone. On each trial, the track for that trial was selected at random among those that had not yet ended. A track ended after nine reversals. The equal-loudness level for each track was calculated as the average of the last four reversals.

A block of trials lasted about five minutes. All details of the adaptive procedure may be found in Nieder *et al.* (2003) as well as the reasons for presenting the comparison tone always in the second interval and for varying only its level. The major difference between the current experiments and those described by Nieder *et al.* (2003) was the inclusion of sessions in which the inducer was in one ear and the test and comparison tones were in the other, contralateral ear. [It is to be noted that the contralateral inducers were intense enough to have resulted in some stimulation via bone conduction in the test-tone ear (e.g., Zwislocki, 1953). However, that stimulation would have been too weak to affect the loudness of the test tone, as shown by Mapes-Riordan and Yost (1999).]

B. Stimuli

All tones in a series had the same equivalent rectangular duration, either 5 or 200 ms. Tones were gated on and off with rise/fall times of 6.67 ms so that the durations measured between the half-amplitude points were 1.67 ms longer than those stated. These envelope shapes ensured that almost all the energy of the tones was contained within the critical band centered around 500 Hz (cf. Buus, 1997). Even for the 5 ms tone burst, the energy within the critical band was only 0.8 dB less than the overall energy. Although nothing has been reported about ILR with sounds other than pure tones, it is highly unlikely that the small difference in the spread of energy would be the source of any measured differences in ILR between 5 and 200 ms tones.

TABLE I. Mean amounts of induced loudness reduction measured for 12 listeners in Marseille and for 12 others in Boston. The standard errors are shown in italics. All values are in dB.

	Induced loudness reduction in dB			
	200 ms Tones		5 ms Tones	
	Ipsilateral	Contralateral	Ipsilateral	Contralateral
Marseille	**12.4**	**6.2**		
	2.2	*1.4*		
Boston	**10.4**	**5.5**	**7.8**	**8.5**
	1.5	*1.2*	*1.5*	*1.8*

At the beginning of each experimental block, twelve inducer tones were presented at 80 dB SPL (85 dB in Marseille) separated by 210 ms intervals in an attempt to accelerate ILR. In all subsequent trials, the inducer tone was followed after 1 410 ms by the test tone; the comparison tone came on 560 ms (600 ms in Marseille) after the end of the test tone. The listener's response initiated the next trial after a delay of 100 ms. The relatively long pause between the inducer and the test tone was chosen to help listeners focus on the loudness of the test tone and to be within the time interval, approximately 500–3 000 ms, during which ILR is greatest and stable (Arieh and Marks, 2003). The timing of the baseline condition was identical to that of the experimental condition, with inducer tones replaced by 5 or 200 ms intervals of silence, as appropriate.

C. Apparatus

A PC-compatible computer with a signal processor (TDT AP2) generated the stimuli, recorded the listeners' responses, and executed the adaptive procedure. Nieder *et al.* (2003) provide technical details.

D. Listeners

Twelve relatively experienced listeners were run in each laboratory. All listeners were tested for or reported normal hearing with the exception of two older listeners in Marseille. These listeners are included since Wagner *et al.* (2005) found that ILR was about the same in ears with elevated thresholds as in ears with normal thresholds.

III. RESULTS

For each listener, the amount of ILR was taken as the setting of the comparison tone in the baseline condition minus the setting in the experimental condition. These were the settings at which the comparison tone was judged equal in loudness to the test tone, without (baseline condition) and with (experimental condition) a preceding inducer on every trial. A positive value means the inducer reduced the loudness of the test tone. The mean data are presented in Table I together with the standard errors.

Despite somewhat large variability, with standard errors between 1.2 and 2.2 dB (standard deviations between 4.1 and 7.6 dB), which is usual in measurements of ILR, the ipsilateral results agree closely with those in the literature. For a signal duration of 200 ms, the amount of ipsilateral ILR, 10 and 12 dB, is close to the 10–11 dB reported for ipsilateral signals in several other papers (e.g., Arieh and Marks, 2003; Mapes-Riordan and Yost, 1999; Nieder *et al.*, 2003). For the duration of 5 ms, our mean ipsilateral ILR of nearly 8 dB is close to the 9 dB reported by Nieder *et al.* (2003).

With respect to the contralateral results, the 200 ms data are similar to those reported by Marks (1996) for 1 s tones. By a method of constant stimuli, Marks measured an *ipsilateral* ILR of 9.5 dB compared to ours of 12.4 and 10.4 dB; he measured a *contralateral* ILR of 5.3 dB compared to ours of 6.2 and 5.5 dB. The differences at 200 ms between ipsilateral and contralateral ILR that we measured both in Marseille and Boston are highly significant by t test ($p=0.0002$). In stark contrast, for a signal duration of 5 ms, ipsilateral and contralateral ILR are about the same, both near 8 dB.

For the data with the 70 dB test tones gathered in Boston, we calculated a two-way ANOVA (inducer location $\times$ tone duration) for each listener and condition. The effect of the inducer location was significant ($F_{1,11}=16.2$, $p=0.002$) as well as the interaction of inducer location and tone duration ($F_{1,11}=14.9$, $p=0.003$). Scheffe post hoc tests confirmed the significant differences between the amounts of ILR for the 200 and 5 ms tones; for ipsilateral ILR, $p=0.03$ and for contralateral ILR, $p=0.01$. Despite the difference in their absolute amounts, ipsilateral and contralateral ILR are highly correlated with $r=0.85$ (Marseille) and 0.76 (Boston) for the 200 ms tones and 0.91 for the 5 ms tones (all significant at the 0.01 level). Such high correlations suggest a common underlying mechanism. On the other hand, listeners show only small and insignificant correlations between the ILR at the two signal durations, that is, strong ILR at 200 ms is not associated with strong ILR at 5 ms, neither for the ipsilateral nor for the contralateral type.

IV. DISCUSSION

An intriguing outcome of the present experiment is the large amount of ILR with contralateral 5 ms inducers—more than with contralateral 200 ms inducers, no less than with 5 ms ipsilateral inducers, and not much less than with *ipsilateral* 200 ms inducers. It may be that contralateral ILR depends on signal duration in a different manner than does ipsilateral ILR. As noted above, the interaction between the two is highly significant.

The resemblances between stimulus conditions for ipsilateral and contralateral ILR are like those for induced loudness adaptation (ILA). Induced loudness adaptation is usually measured with a continuous tone, to which is added an intermittent increment or brief tone either to the same ear, for ipsilateral ILA, or to the other ear, for contralateral ILA. Thus, under both ILA and ILR a stronger tone is presented either to the same ear as a weaker tone or to the opposite ear with the result that the loudness of the weaker tone often declines (e.g., Charron and Botte, 1988). Wagner and Scharf (2006) elaborated on the suggestion of Nieder *et al.* (2003) and of Scharf (2001) that ipsilateral ILR and ipsilateral ILA are based on some of the same underlying mechanisms. Contralateral ILR, at least for longer tones, may arise in a similar fashion. However, contralateral ILA is probably based on a different process in part because it involves binaural interactions between simultaneous inducer tones and test tones (cf. Charron and Botte, 1988).

The basis for induced loudness reduction remains unclear. Any proffered hypotheses about the basis for ILR will have to take into account the existence of contralateral ILR and its likely dependence on signal duration.

ACKNOWLEDGMENTS

We thank Mary Florentine for encouraging this research and Maria Ortiz for running many of the subjects. This research was supported by NIH/NIDCD Grant No. R01DC02241.

Arieh, Y., and Marks, L. E. (**2003**). "Time course of loudness recalibration: Implications for loudness enhancement," J. Acoust. Soc. Am. **114**, 1550–1556.

Buus, S. (**1997**). "Auditory masking," in *Encyclopedia of Acoustics*, Vol. 3, edited by M. J. Crocker (Wiley, New York), pp. 1427–1445.

Charron, S., and Botte, M. (**1988**). "Frequency selectivity in loudness adaptation and auditory fatigue," J. Acoust. Soc. Am. **83**, 178–187.

Jesteadt, W. (**1980**). "An adaptive procedure for subjective judgments," Percept. Psychophys. **28**, 85–88.

Mapes-Riordan, D., and Yost, W. A. (**1999**). "Loudness recalibration as a function of level," J. Acoust. Soc. Am. **106**, 3506–3511.

Marks, L. E. (**1996**). "Recalibrating the perception of loudness: Interaural transfer," J. Acoust. Soc. Am. **100**, 473–480.

Nieder, B., Buus, S., Florentine, M., and Scharf, B. (**2003**). "Interactions between test-and inducer-tone durations in induced loudness reduction," J. Acoust. Soc. Am. **114**, 2846–2855.

Scharf, B. (**2001**). "Sequential effects in loudness," in *Fechner Day 2000*, edited by Sommerfeld, E., Kompass, R., and Lachmann, T., (Papst, Berlin), pp. 254–259.

Wagner, E., Scharf, B., and Florentine, M. (**2005**). "Induced loudness reduction in hearing-impaired listeners." Paper presented at Meeting of the Association for Research in Otolaryngology, Abstract 538.

Wagner, E., and Scharf, B. (**2006**). "Induced loudness reduction as a function of exposure time and signal frequency," J. Acoust. Soc. Am. **119**, 1012–1020.

Zwislocki, J. (**1953**). "Acoustic attenuation between the ears," J. Acoust. Soc. Am. **25**, 752–759.

Nonlinear standing waves in a resonator with feedback control (L)

X. Y. Huang,[a] N. T. Nguyen, and Z. J. Jiao
School of Mechanical and Aerospace Engineering, Nanyang Technological University, Singapore 639798, Singapore

(Received 5 February 2007; revised 29 March 2007; accepted 8 April 2007)

An experimental study is presented to demonstrate that nonlinear effect on standing waves in a resonator can be reduced by a feedback loop responding to the second harmonic. The resonator was a cylindrical tube sealed at one end and driven by a horn driver unit at another end. The feedback control loop consisted of a pressure sensor, a frequency filter, a phase shifter, and an actuator. The results show that the waveform distortions can be eliminated and large amplitude sinusoidal pressure oscillations are obtained. A simple model is proposed for a qualitative discussion on the control mechanism, which shows that the feedback loop alters the imaginary part of the complex mode frequency so as to suppress (or enhance) the second harmonic. © *2007 Acoustical Society of America.* [DOI: 10.1121/1.2735808]

PACS number(s): 43.25.Gf, 43.25.Ts [MFH] Pages: 38–41

I. INTRODUCTION

When standing waves are driven to high amplitudes in a resonator, the nonlinear characteristics of gas dynamics will convert part of the dynamic energy from the driving frequency to high order harmonics. As a result, the sinusoidal pressure wave forms responding to the driving frequency will be distorted. Shock waves will ultimately be formed inside the resonator and pressure amplitudes be saturated due to high dissipations.[1–3] Research efforts have recently been made to suppress the nonlinear effect on dynamics pressures in resonators, so that the shock waves will not be formed and pressure amplitudes can be increased. At present, there are three possible methods available for this purpose. The first one is to make resonators with nonuniform cross sections[4–7] the second one to make the resonators dispersive,[8–10] and the third to use active control to suppress the harmonics.[11] Lawrenson *et al.*[4] conducted an experimental investigation to demonstrate that significantly larger amplitude pressure oscillations with more continuous wave forms could be generated in shaped resonators than the saw-tooth wave forms in straight ducts. Ilinskii *et al.*[5,6] developed a one-dimensional model to calculate pressure waves in axisymmetric resonators with varying cross-section areas. The results confirmed major characteristics of shaped resonators, such as wave forms and frequency response. Hamilton *et al.*[2] analytically studied the relationship between the resonance frequencies and resonator shapes, especially the nonlinear interactions of modes in shaped resonators. Sugimoto *et al.*[8,9] introduced dispersion into a straight tube by arranging periodically identical Helmholtz resonators along the tube. The resonance frequency of the Helmholtz resonator was tuned around the second harmonic of the straight tube. They demonstrated that the shock waves were annihilated and that high-amplitude pressure oscillations were generated in the straight tube. The concept was further verified by an improved experiment.[10] An active approach was adopted by Huang *et al.*[11] to suppress the high resonance modes generated in a resonance tube under strong driving. The control system in their study consisted of a computer with a sophisticated software to process the input pressure signals and generate control signals. The results showed that the wave form distortions were eliminated by completely suppressing the second and third harmonics. Since the second harmonic is probably the most crucial mode associated with the nonlinear behaviorof the resonators,[7] it would be interesting to see if the nonlinear resonance could be altered by suppressing only the second harmonic. Besides, the control of a single frequency mode may be achieved by some simple techniques without involving complicated signal processing. The purpose of the present paper is to examine the effect of a simple analog feedback loop responding to the second harmonic on the wave form distortions and overall resonance pressure amplitudes.

II. EXPERIMENT

A. Apparatus

The experimental setup is illustrated in Fig. 1, and consists of a cylindrical resonator, a driver unit, a pressure sensor, a feedback loop, and other electronic instruments. The cylindrical resonator was made of brass with an inner diameter of 25 mm and an outer diameter of 45 mm. One end of the resonator was screwed onto the driver unit and the other end was sealed. The air column length from the opening of the driver unit to the sealed end inside the resonator was about 164 mm. The driver unit was a 150 W horn driver (VOICETEK, NSU-150) driven by a power amplifier (Crown XLS202). The pressure transducer (PCB 103B12) was fixed at the sealed end of the cylindrical resonator. The signals from the pressure transducer were received by a signal conditioner (PCB 483 A) and analyzed by a signal analyzer (ONO SOKKI CF-5220Z). The outputs from the signal analyzer were collected in two forms after correction by the

[a]Electronic mail: mxhuang@ntu.edu.sg

 0001-4966/2007/122(1)/38/4/$23.00

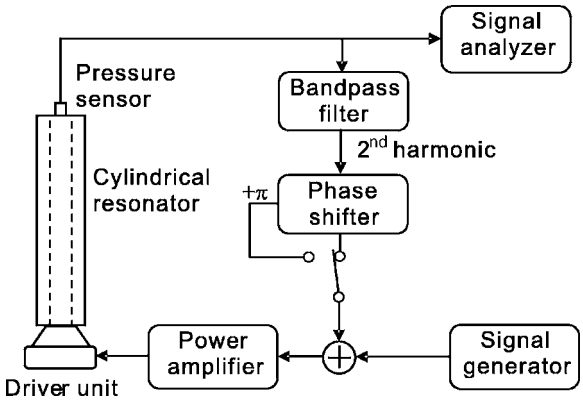


FIG. 1. A schematic diagram for the cylindrical resonator and the feedback loop used in the experiment.

transducer sensitivity (75 mV/kPa). One form was in time domain to give pressure wave forms, and another was in frequency domain to show pressure spectra in decibel scales which were defined as $20\log(P_{\mathrm{rms}}/1\ \mathrm{kPa})$, with P_{rms} being the rms value in kilopascal. The signals were also input to a bandpass frequency filter which allowed the second harmonics to pass to an in-house made analog phase shifter. The phase shifter can provide a phase shift to signals between the output and input from 0 to 2π, with an option to have additional shift π. This additional phase shift π was useful in observing different control effects with the phase shift being inversed from one to another. The output of the phase shifter, after being added to the primary driving signals from a HP 3312A wave form generator, was sent to the power amplifier to generate controlled driving in the cylindrical resonator, so as to close the feedback loop. During the experiment, the voltage input to the driver unit was monitored and used to indicate the driving strength.

B. Results and discussions

The fundamental resonance frequency of the resonator was found to be 1044 Hz. The experiments were conducted by driving the resonator at the fundamental resonance frequency with different driving strengths: moderate driving (driving voltage <5 V), strong driving (5 V$<$ driving voltage <30 V), and limited driving (driving voltage >30 V, which was close to the power limit of the driver unit). The results at various driving strengths, with and without the control, are presented in both wave forms and frequency spectra in Fig. 2. At moderate strength, the pressure wave forms are still sinusoidal waves without distortions. Typical pressure wave forms at the driving voltage of 2.7 V are shown in Fig. 2(a). The spectra in Fig. 2(b) show that the second harmonic has an amplitude of 24 dB lower than the fundamental mode. When the voltage input to the amplifier was increased to the strong driving range, wave form distortions were observed and the control effect was studied. The driving voltage was 15 V for the results presented in Figs. 2(c)–2(h). The resonance results without the control are shown in Figs. 2(c) and 2(d). The wave form is distorted and becomes asymmetric from positive to negative. The corresponding spectra plotted in Fig. 2(d) show that the second harmonic in this case has

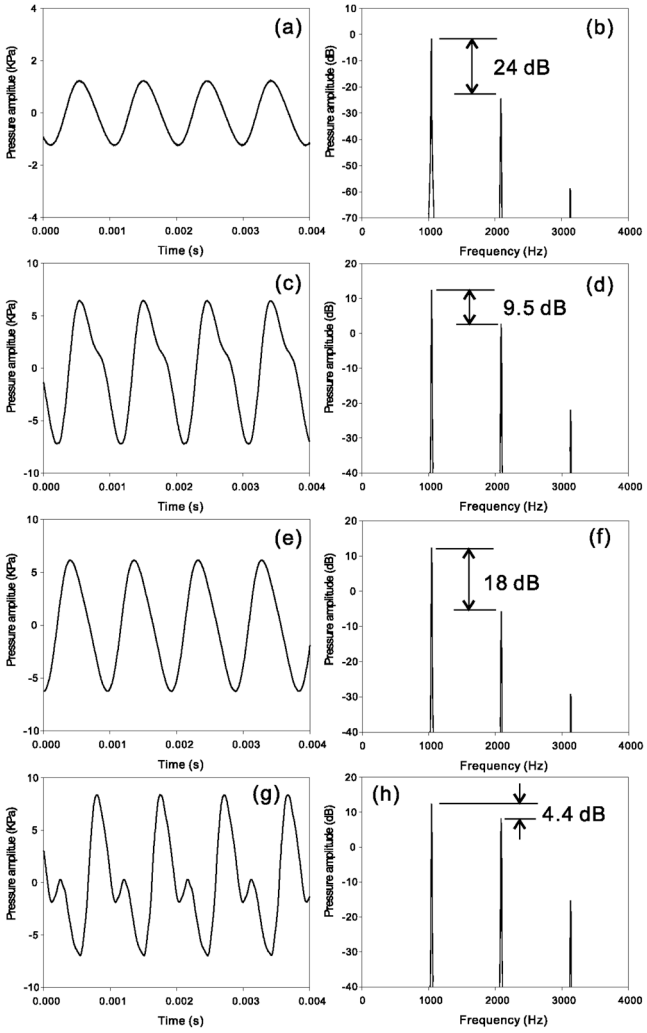


FIG. 2. Experimental pressure wave forms and spectra in the cylindrical resonator driven at the fundamental frequency 1044 Hz. (a) Pressure wave forms and (b) pressure spectra under the driving voltage 2.7 V without control. (c)–(h) are the results under the driving voltage 15 V. (c) Pressure wave forms and (d) pressure spectra without control. (e) Pressure wave forms and (f) pressure spectra with the control at the optimal phase and gain. (g) Pressure wave forms and (h) pressure spectra with the control at the phase shift inversed from the optimal setting.

the amplitude of only 9.5 dB below the fundamental mode. There is a 14.5 dB increase compared to the moderate driving results in Fig. 2(b), indicating that the second harmonic is getting more energy from the fundamental mode through the nonlinearity of gas dynamics. The control was actuated by closing the feedback loop and adjusting the gain and phase shift until the maximum amplitude reduction on the second harmonic was observed. The gain and phase shift setting in this case is called optimal control setting to distinguish other combinations of the gain and phase shift, which could also reduce the second harmonic. The control effect can be seen from Figs. 2(e) and 2(f). The wave form distortions are almost suppressed and the amplitude difference between the second harmonic and fundamental mode is increased to 18 dB. The control has no significant effect on the pressure amplitude at the fundamental mode in this case. The control effect was further examined by reversing the phase shift from the optimal setting by means of a switch in the

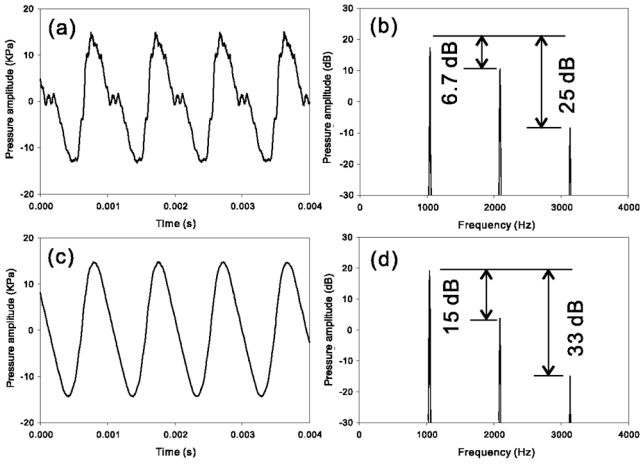


FIG. 3. Pressure wave forms with and without the control under the same driving voltage 32 V. (a) Pressure wave forms and (b) pressure spectra without control. (c) Pressure wave forms and (d) pressure spectra with the control at the optimal phase and gain.

phase shifter. Figure 2(h) show that the second harmonic has been increased to become 4.4 dB below the fundamental mode, and the wave forms, plotted in Fig. 2(g), are now strongly distorted. The controller in this case has apparently enhanced the second harmonic instead of suppressing it. With the control at the optimal setting, the driving strength was further increased to approach to the limit of the driver. The results at the driving voltage of 32 V are plotted in Fig. 3. It is seen that the pressure wave form without the control [Fig. 3(a)] is further distorted, compared to Fig. 2(c), and there are high frequency components on the curves, which are probably caused by high harmonics. The spectrum plotted in Fig. 3(b) shows that the second harmonic is now stronger with an amplitude only 6.7 dB lower than the fundamental mode. The third harmonic is increased to have an amplitude about 25 dB below the fundamental mode, comparing to 34 dB in Fig. 2(d) (not labeled there), which may contribute to additional wave form distortions, together with other high harmonics. Figures 3(c) and 3(d) show the results of the resonance under the same driving but with the control. The pressure wave form becomes sinusoidal without much distortion. The spectra plotted in Fig. 3(d) show that the control has substantially suppressed the second and third harmonics, respectively. A further comparison between the fundamental modes in Figs. 3(b) and 3(d) reveals that with the control the amplitude of the fundamental mode has increased about 2 dB. This suggests that some of the energy has been retained at the fundamental mode by the control, which may otherwise be transferred to the second and third harmonics due to the nonlinear mode-to-mode interaction. Although the present study is not focused on producing large amplitude pressure oscillations, it is interesting to note that the pressure oscillations obtained in Fig. 3(c) are of peak-to-peak amplitudes about 28 kPa. This is quite meaningful considering that the wave forms with such amplitudes are free from distortions and are generated in a cylindrical resonator.

III. CONTROL MECHANISM

A simple model is proposed here to highlight possible mechanism of the feedback control on the second harmonic.

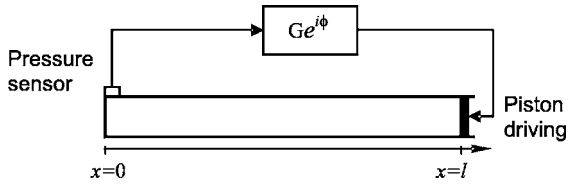


FIG. 4. A schematic drawing for a cylindrical resonator with a feedback loop used in the analytical model.

Figure 4 is a schematic drawing of the controlled resonator used for the model. The resonator is sealed at one end ($x=0$) and is driven by a piston at other end ($x=l$). The piston is responding to the control signals associated with the second-harmonic sound pressures picked by a sensor at $x=0$. The linear acoustic wave equation is used in this model for the second harmonic, which is generally weak compared to the fundamental mode. The wave equation and boundary conditions are written as

$$\frac{\partial^2 p_2}{\partial t^2} - c^2 \frac{\partial^2 p_2}{\partial x^2} = 0, \tag{1}$$

$$\left.\frac{\partial p_2}{\partial x}\right|_{x=0} = 0, \quad \left.\frac{\partial p_2}{\partial x}\right|_{x=l} = \frac{1}{l} G e^{i\phi} |p_2(x=0)| e^{i\omega_2 t}, \tag{2}$$

where p_2 is the pressure for the second harmonic and c is the speed of sound. The factor $1/l$ is introduced to make the control gain G nondimensional. ϕ is the phase shift. ω_2 is the mode frequency for the second harmonic, which will be determined from the boundary conditions. The solution for p_2 to fit the homogeneous boundary condition at $x=0$ can be found as

$$p_2(x,t) = A \cos(kx) e^{i\omega_2 t}, \tag{3}$$

with

$$k = \frac{\omega_2}{c}. \tag{4}$$

By applying the boundary condition at $x=l$, one obtains

$$-Ak \sin(kl) = \frac{1}{l} G e^{i\phi} |p_2(x=0)|. \tag{5}$$

Since $|p_2(x=0)| = A$ [according to Eq. (3)], Eq. (5) can be written as

$$-Ak \sin(kl) = \frac{1}{l} G e^{i\phi} A, \tag{6}$$

or

$$kl \sin(kl) = -Ge^{i\phi}. \tag{7}$$

Equation (7) is the characteristic equation to determine wave number k, and thus ω_2. If the controller is switched off, i.e., $G=0$, Eq. (7) will be reduced to $\sin(kl)=0$ and we have $kl = 2\pi$ for the second harmonic. In this case,

$$kl = \frac{\omega_2 l}{c} = 2\pi,$$

so that

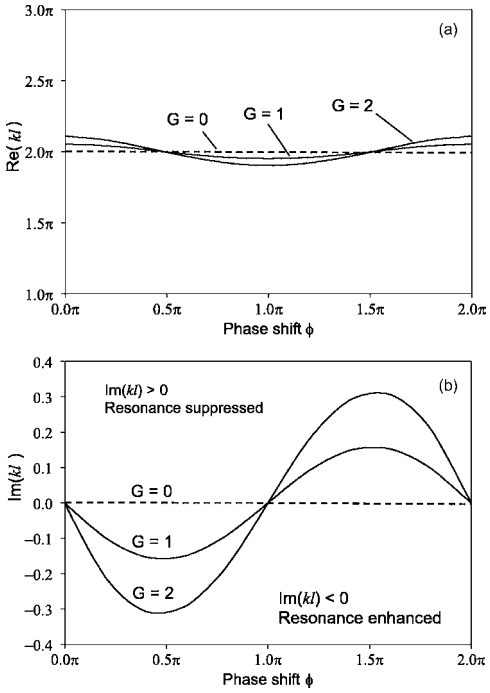


FIG. 5. The real parts and imaginary parts of kl calculated from Eq. (7) vs the control phase shift ϕ at different values of the control gain.

$$\omega_2 = \frac{2\pi c}{l}. \tag{8}$$

When the controller is switched on, kl will be solved from Eq. (7) and will generally be a complex number, depending on the control parameters G and ϕ. We may set

$$kl = \mathrm{Re}(kl) + i\,\mathrm{Im}(kl) = \frac{\omega_2 l}{c}, \tag{9}$$

so that

$$\omega_2 = \frac{c\,\mathrm{Re}(kl)}{l} + i\frac{c\,\mathrm{Im}(kl)}{l}. \tag{10}$$

The time factor for the second harmonic, according to Eq. (10), will become

$$\exp(i\omega_2 t) = \exp\left[i\frac{c\,\mathrm{Re}(kl)}{l}t - \frac{c\,\mathrm{Im}(kl)}{l}t\right]. \tag{11}$$

Equation (11) shows that the second harmonic under the control will have a mode frequency $c\,\mathrm{Re}(kl)/t$ and a time-dependent amplitude. If $\mathrm{Im}(kl) > 0$, the amplitude will decay exponentially with time, indicating that the resonance cannot be sustained. On the other hand, if $\mathrm{Im}(kl) < 0$, the amplitude will increase exponentially with time, indicating an unsteady mode. The behavior of the second harmonic resonance mode is now determined by the control, i.e., the gain G and phase shift ϕ. Variations of $\mathrm{Re}(kl)$ and $\mathrm{Im}(kl)$ with the control phase shift are calculated from Eq. (7) and plotted, respectively, in Figs. 5(a) and 5(b) for $G=1$ and 2. It is seen that the control effect has modified $\mathrm{Re}(kl)$ slightly around 2π, which is the value for kl without the control. The control has indeed introduced imaginary parts in kl and the sign of $\mathrm{Im}(kl)$ is dependent on the phase shift, as shown by Fig. 5(b). It is seen that $\mathrm{Im}(kl)$ is negative when ϕ is between 0 and π, indicating that the second harmonic will be enhanced. $\mathrm{Im}(kl)$ is positive when ϕ is between π and 2π, indicating that the second harmonic will be suppressed. The maximum and minimum of $\mathrm{Im}(kl)$ are out of phase. These results may qualitatively explain the feedback control loop used in the experiment.

IV. CONCLUSIONS

An experimental study has been conducted on nonlinear resonances in a straight cylinder with a feedback control to the second harmonic. The results show that, with the control set at the optimal phase and gain, the pressure wave form distortions which occurred at strong driving have been suppressed and large amplitude sinusoidal pressure oscillations without distortions are obtained. The controller, which is designed for the second harmonic, is found to have effects on the third harmonic to make the pressure wave forms smooth. A qualitative explanation to the control mechanism has been provided by a simple model, showing that the feedback control loop used in the experiment can weaken or enhance the second harmonic, depending on the phase shift. The results produced by this study may help to deepen our understanding on mode-to-mode interactions associated with nonlinear resonances in cavities. The simple controller based on the analog phase shifter may be incorporated with other techniques, such as the shaped resonators, to generate high amplitude pressures.

[1]R. Saenger and G. Hudson, "Periodic shock waves in resonating gas columns," J. Acoust. Soc. Am. **32**, 961–970 (1960).
[2]D. B. Cruikshank, Jr., "Experimental investigation of finite-amplitude acoustic oscillations in a closed tube," J. Acoust. Soc. Am. **52**, 1024–1036 (1972).
[3]J. Jimenez, "Nonlinear gas oscillations in pipes. I. Theory," J. Fluid Mech. **59**, 23–46 (1973).
[4]C. C. Lawrenson, B. Lipkens, T. S. Lucas, D. K. Perkins, and T. W. Van Doren, "Measurement of macrosonic standing wave in oscillating closed cavities," J. Acoust. Soc. Am. **104**, 623–636 (1998).
[5]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).
[6]Y. A. Ilinskii, B. Lipkens, and E. A. Zabolotskaya, "Energy losses in an acoustical resonator," J. Acoust. Soc. Am. **109**, 1859–1870 (2001).
[7]M. F. Hamilton, Y. A. Ilinskii, and E. A. Zabolotskaya, "Linear and nonlinear frequency shifts in acoustical resonator with varying cross section," J. Acoust. Soc. Am. **110**, 109–119 (2001).
[8]N. Sugimoto, M. Masuda, T. Hashiguchi, and T. Doi, "Annihilation of shocks in forced oscillations of an air column in a closed tube," J. Acoust. Soc. Am. **110**, 2263–2266 (2001).
[9]N. Sugimoto, M. Masuda, and T. Hashiguchi, "Frequency response of nonlinear oscillations of air column in a tube with an array of Helmholtz resonators," J. Acoust. Soc. Am. **114**, 1772–1784 (2003).
[10]M. Masuda and N. Sugimoto, "Experiments of high-amplitude and shock-free oscillations of air column in a tube with array of Helmholtz resonators," J. Acoust. Soc. Am. **118**, 113–123 (2005).
[11]P. T. Huang and J. G. Brisson, "Active control of finite amplitude acoustic waves in a confined geometry," J. Acoust. Soc. Am. **102**, 3256–3267 (1997).

Sounds recorded in the presence of Blainville's beaked whales, *Mesoplodon densirostris*, near Hawai'i (L)

Shannon Rankin[a] and Jay Barlow
Southwest Fisheries Science Center, 8604 La Jolla Shores Drive, La Jolla, California 92037

(Received 16 March 2007; revised 17 April 2007; accepted 1 May 2007)

During a combined visual and acoustic cetacean survey of the Hawaiian Islands in 2002, four midfrequency sounds were recorded in close proximity to a group of Blainville's beaked whales, *Mesoplodon densirostris*. These sounds included one frequency-modulated whistle, and three frequency- and amplitude-modulated pulsed sounds, with energy between 6 and 16 kHz. Until recently, little was known of the acoustic behavior of beaked whales, and early descriptions of audible sounds made by beaked whales are incomplete [Caldwell and Caldwell, Cetology **4**, 1–5 (1971); Lynn and Reiss, Marine Mammal Sci. **8**(3), 229–305 (1992); T. C. Poulter, "Marine mammals," in Animal Communication; Techniques of Study and Results of Research, edited by T. A. Sebeok (Indiana University Press, Bloomington, 1968)]. Recent recordings of high-frequency clicks (>20 kHz, [Johnson *et al.*, Proc. R. Soc. London, Ser. B (Suppl.) **271**, 5383–5386 (2004); Zimmer *et al.*, J. Acoust. Soc. Am. **117**(6), 3919–3927.]) were above the frequency response of our equipment, and therefore not detected. Sound production within the midfrequency range of sounds described here suggests that the hearing of *M. densirostris* is sensitive at frequencies used in some types of active sonars. © *2007 Acoustical Society of America.* [DOI: 10.1121/1.2743159]

PACS number(s): 43.80.Ka, 43.30.Sf [WWA] Pages: 42–45

I. INTRODUCTION

The connection between mass strandings of beaked whales with U.S. Navy midfrequency sonar has led to an increased interest in the acoustic behavior and physiology of beaked whales (Cook *et al.*, 2006; Johnson *et al.*, 2004; Madsen *et al.*, 2005; Zimmer *et al.*, 2005). At present, the cause of strandings remains unclear (Cox *et al.*, 2006). Sounds produced by beaked whales provide insight into their hearing sensitivity and frequency bands important for communication and foraging. Visual and acoustic monitoring for the presence of beaked whales can minimize exposure to sonar and other anthropogenic sounds; however, acoustic monitoring of beaked whales requires an understanding of their vocal repertoire (Barlow *et al.*, 2006; Barlow and Gisiner, 2006).

Until recently, little has been known of the acoustic behavior of beaked whales. Recordings of a stranded Blainville's beaked whale (*Mesoplodon densirostris*) consisted of pulsed and tonal calls in the frequency range of 1–6 kHz (Caldwell and Caldwell, 1971). Poulter (1968) described *Mesoplodon* calls as including "roars" and "lowing and sobbing groans." Lynn and Reiss (1992) recorded broadband pulsed sounds and midfrequency whistles (1.4 to 10.7 kHz) from stranded *M. carlhubbsi* held in captivity. Recently, research using acoustic recording tags has increased our understanding of echolocation of beaked whales (Johnson *et al.*, 2004; Zimmer *et al.*, 2005). Johnson *et al.* (2004) did not detect sounds with significant energy below 20 kHz in either *M. densirostris* or *Ziphius cavirostris*, and no calls of any kind were detected when the animals were within 200 m of the surface. In this paper, we present sounds recorded in the presence of *M. densirostris* during a recent cetacean survey off the Hawaiian Islands.

II. METHODS

The 2002 Hawaiian Island Cetacean Ecosystem and Abundance Survey (HICEAS) survey (Fig. 1), aboard the R/V *David Starr Jordan* and *R/V McArthur*, included the exclusive economic zone of the Hawaiian Island chain and transit to and from San Diego, CA (Barlow *et al.*, 2004). This research project combined visual and acoustic detection of cetaceans using standard protocol designed by Southwest Fisheries Science Center (SWFSC). Visual observation of cetaceans was conducted during daylight hours and consisted of six experienced visual observers rotating between two "big eye" binocular (25×150) stations and one station observing with 7× binoculars and unaided eye (Kinzey *et al.*, 1999).

The acoustics team was on the *Jordan* only, and consisted of two rotating acoustic technicians monitoring a towed hydrophone array and a hull-mounted hydrophone aurally and visually using ISHMAEL real-time spectrogram display (Mellinger, 2001). A hydrophone was mounted to the bow of the *Jordan*, to detect bow-riding dolphins and other cetaceans in close proximity to the bow of the ship, where the towed array has limited detection ability. The bow hydrophone has a limited acoustic range due to high ambient noise levels, and during this survey they were occasionally monitored when animals were near the bow. The frequency response of the bow hydrophone was 500 to 25 kHz (±10 dB). Output from the bow hydrophone was recorded with the output from the towed hydrophone array. The hydrophone array consisted of two elements separated by 3 m, with a fre-

[a]Electronic mail: shannon.rankin@noaa.gov

 0001-4966/2007/122(1)/42/4/$23.00

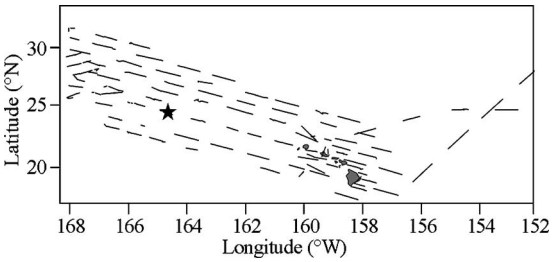


FIG. 1. Map of sighting and acoustic detection of *Mesoplodon densirostris* northwest of the Hawaiian Islands. Tracklines for the HICEAS survey are shown as solid lines, with the *M. densirostris* shown as a star.

quency response of 500 to 25 kHz (±10 dB). The array was towed 200 m behind the vessel at an average speed of 10 knots and an average depth of 6 m. Hydrophone output was passed through a Mackie CR1604-VLZ sound mixer for high-pass filtering of low-frequency noise (cut below 70 Hz). All recordings were made using a Tascam DA-38 digital recorder, sampling at 48 kHz.

III. RESULTS

At 15:57 on 17 October 2002, a group of six *M. densirostris* was detected 200 m in front of the ship 116 km southeast of Laysan Island at 24° 48.13 N, 171° 17.89 W (Fig. 1). The only other sightings within 30 miles were two sightings of sperm whales and two sightings of Bryde's whales. No other cetaceans were seen within three miles of the *M. densirostris*. Beaufort 0 weather conditions provided exemplary views of this group of three cow/calf pairs. During this encounter, 5–7 surfacings were observed before the animals dove within 50 m of the bow at approximately 15:58 local time. At 15:59, four calls were detected on the bow hydrophone with low signal-to-noise ratio (SNR). There was interference from wave noise on the bow and from the anchor knocking against the hull of the ship. No sounds were detected on the hydrophone array. Despite the poor SNR of these recordings, they may prove to be important in understanding the acoustic behavior of *M. densirostris*.

The first of the four calls was a frequency-modulated downsweep, with harmonics [Fig. 2(a)]. The other three calls exhibited both frequency- and amplitude-modulation [Fig. 2(b), Fig. 3]. Due to the low SNR, the pulsed nature of these sounds could not be detected in the waveform, but is evident in the banding found in the spectrogram. The poor quality required that approximations of the call characteristics were made using the spectrogram instead of the waveform (48 kHz sampling rate, 90% overlap, Hann window, 2048 FFT size).

The first two calls are of extremely poor quality, with wave interference separating what appears to be the first and the second call. The first call is a downsweep of at least 1.0 s duration and ranging from 12 to 6.3 kHz [Fig. 2(a)]. The wave noise appears to mask both the lowest frequency and end of the first call as well as the beginning of the second call. The second call appears to be pulsed in nature, with an interclick time interval of 1,474 s^{-1} and duration over 1.0 s [Fig. 2(b)]. The third and fourth calls are both 0.9 seconds in length, and have similar frequency-and amplitude-

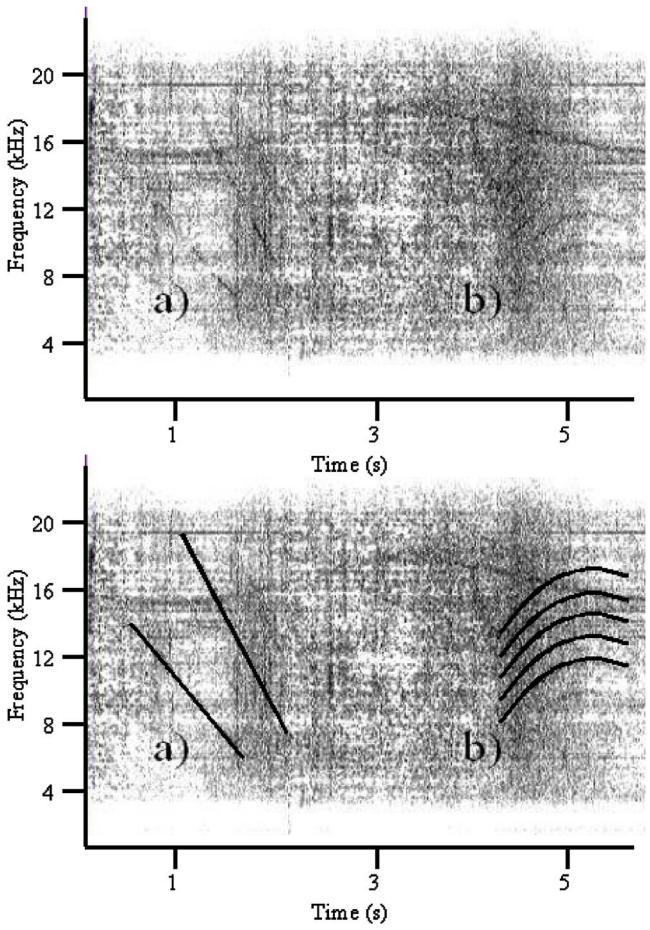


FIG. 2. (Color online) First and second of four vocalizations recorded in the presence of *M. densirostris*, with an outline of the interpreted signals shown in the bottom spectrograph (48-kHz sampling rate, Hann window, 512 FFT size). The first call is a frequency-modulated downsweep (A), and the second call exhibits both frequency- and amplitude modulation (B).

modulations to each other [Figs. 3(a) and 3(b)]. Both calls are pulsed sounds, with the greatest energy between 9.4 and 15.5 kHz. The pulse repetition rate is approximately 711 and 752 s^{-1} for the third and fourth calls, respectively. Due to the poor signal quality, it was impossible to obtain accurate measurements, and these values are intended as a general description of these calls.

IV. DISCUSSION

The sounds described here were produced by a group of three cow/calf pairs in close proximity to the research vessel. The animals dove within 50 m of the bow at the approximate time of this recording, and no other cetaceans were detected in the near vicinity (3 nmi). Although a reasonable comparison cannot be made with the basic descriptions of published accounts of *M. densirostris* vocalizations (Caldwell and Caldwell, 1971; Poulter, 1968), these sounds do appear to be similar to some sounds produced by the highly vocal Baird's beaked whales, *Berardius bairdii* (Dawson *et al.*, 1998; Rankin and Barlow). Recent research using acoustics tags on Blainville's and Cuviers' beaked whales (*Z. cavirostris*) has detected click sounds while the animals were at depth (Johnson *et al.*, 2004). These animals appeared to be feeding,

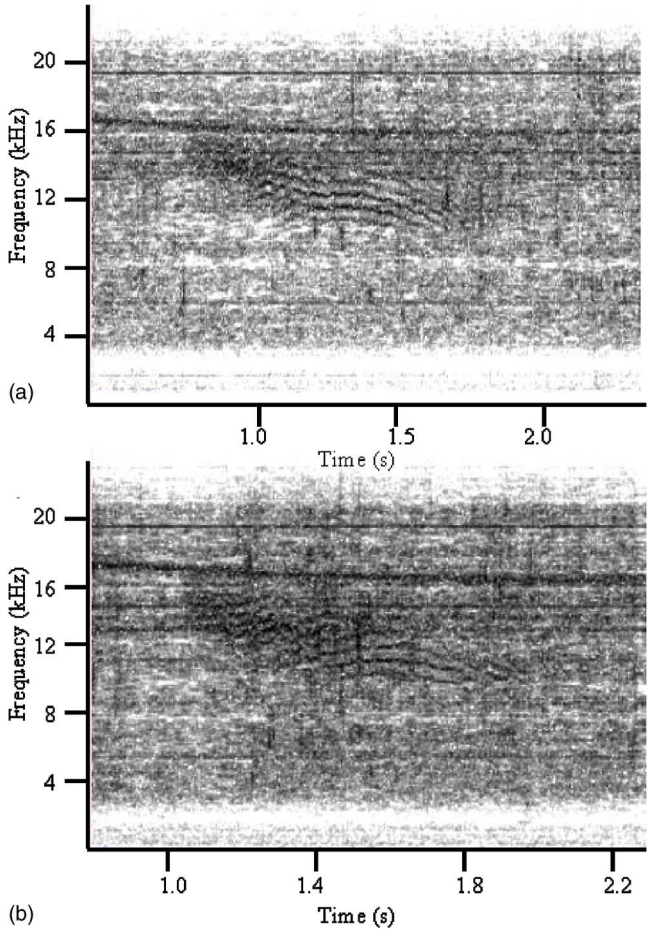


FIG. 3. (Color online) Third (A) and fourth (B) calls, exhibiting both frequency- and amplitude modulation, recorded in the presence of *M. densirostris* (48-kHz sampling rate, Hann window, 512 FFT size).

and no other sound types were recorded from these animals. While it is likely that the predominant sounds of *Mesoplodon* species are clicks produced while foraging, this study shows that they may at least occasionally produce midfrequency tonal sounds like those sounds that have been associated with communication in other species.

In five seasons of combined visual and passive acoustics cetacean surveys, including 51 964-km survey effort, sounds from only 2 of 72 beaked whale sightings have been detected on a towed hydrophone array (Barlow and Rankin). Both of these detections were of *B. bairdii* in Alaskan waters. The sounds of *M. densirostris* presented here were detected on bow hydrophones, and were not detected on the towed hydrophone array. Detection of sounds on the hydrophone array within 15 deg of the bow is negatively affected by the physical presence of the ship, and sounds detected with this range tend to have a high SNR (Rankin and Barlow). Beaked whales spend a relatively short time in surface waters, and most calls are high-frequency echolocation clicks produced at depth, far below the depth and above the frequency range of our towed hydrophones. We were unable to determine the source level of the sounds detected on the bow hydrophone; however, the poor SNR on the bow hydrophone and the fact that they were not detected on the hydrophone array indicate that they were likely low in intensity.

Despite the paucity of data concerning the hearing sensitivity of beaked whales, it is reasonable that they should have sensitive hearing at those frequencies that they utilize. Thus, it is expected that Blainvilles' beaked whales should have sensitive hearing in these midfrequency ranges. Cook *et al.* (2006) found that *M. europaeus* could minimally detect sounds between 5 and 80 kHz, with a similar hearing sensitivity at 5 kHz as the bottlenosed dolphin, *Tursiops truncatus*. The pulsed sounds presented in this paper provide verification that beaked whales do, indeed, use midrange frequencies which overlap in frequency with some naval sonar applications.

ACKNOWLEDGMENTS

Many thanks to the officers and crew of the R/V *David Starr Jordan* for their continued hard work and dedication. This research could not have been accomplished without the quick reactions of our marine mammal observers, including Holly Fearnbock, Erin LaBrecque, Liz Mitchell, Cornelia Oedekoven, Richard Rowlett, Juan Carlos Salinas, and Suzanne Yin. We are also grateful to the cruise leader Sarah Mesnick, and to Jenna Borberg for her assistance with acoustics. Sofie VanParijs, Jessica Burtenshaw and two anonymous reviewers provided helpful critiques on an early version of this manuscript. Funding was provided by the U.S. Navy.

Barlow, J., and Rankin, S. (**2007**). Manuscript available from Shannon Rankin, SWFSC, 8604 La Jolla Shores Drive, La Jolla, CA 92937.

Barlow, J., and Gisiner, R. (**2006**). "Mitigating, monitoring, and assessing the effects of anthropogenic sound on beaked whales," J. Cetacean Res. Manage. **7**(3), 239–249.

Barlow, J., Rankin, S., Zele, L., and Appler, J. (**2004**). Marine mammal data collected during the Hawaiian Islands cetacean and ecosystem assessment survey (HICEAS) conducted aboard the NOAA ships *McArthur* and *David Starr Jordan*, July–December, 2002. NOAA Technical Memorandum NOAA-TM-NMFS-SWFSC-362. Available from NOAA Southwest Fisheries Science Center, 8604 La Jolla Shores Dr., La Jolla, CA 92037. 32pp.

Barlow, J., Ferguson, M. C., Perrin, W. F., Ballance, L., Gerrodette, T., Joyce, G., Macleod, C. D., Mullin, K., Palka, D. L., and Waring, G. (**2006**). "Abundance and densities of beaked whales and bottlenose whales (family *Ziphiidae*)," J. Cetacean Res. Manage. **7**(3), 263–270.

Caldwell, D. K., and Caldwell, M. C. (**1971**). "Sounds produced by two rare cetaceans stranded in Florida," Cetology **4**, 1–5.

Cook, M. L. H., Varela, R. A., Goldstein, J. D., McCulloch, S. D., Bossart, G. D., Finneran, J. J., Houser, D., and Mann, D. A. (**2006**). "Beaked whale auditory evoked potential hearing measurements," J. Comp. Physiol., A **192**, 489–495.

Cox, T. M., Ragen, T. J., Read, A. J., Vos, E., Baird, R. W., Balcomb, K., Barlow, J., Caldwell, J., Cranford, T., Crum, L., D'Amico, A., D'Spain, G., Fernández, A., Finneran, J., Gentry, R., Gerth, W., Gulland, F., Hildebrand, J., Houser, D., Hullar, T., Jepson, P. D., Ketten, D., MacLeod, C. D., Moore, S., Mountain, D., Palka, D., Ponganis, P., Rommel, S., Rowles, T., Taylor, B., Tyack, P., Wartzok, D., Gisiner, R., Mead, J., and Benner, L. (**2006**). "Understanding the impacts of anthropogenic sound on beaked whales," J. Cetacean Res. Manage. **7**, 177–187.

Dawson, S., Barlow, J., and Ljungblad, D. (**1998**). "Sounds recorded from Baird's beaked whales, *Berardius bairdii*," Marine Mammal Sci. **14**(2), 335–344.

Johnson, M., Madsen, P. T., Zimmer, W. M. X., Aguilar de Soto, N., and Tyack, P. L. (**2004**). "Beaked whales echolocate on prey," Proc. R. Soc. London, Ser. B **271**, S383–S386.

Kinzey, D., Gerrodette, T., Barlow, J., Dizon, A., Perryman, W., Olson, P., and Von Saunder, A. (**1999**). Marine mammal data collected during a

survey in the Eastern tropical Pacific Ocean aboard the NOAA ships *McArthur* and *David Starr Jordan* and the UNOLS ship *Endeavor* 31 July–9 December, 1998. U.S. Department Commer., S., NOAA Technical Memo., NOAA-TM-NMFS-SWFSC-283. 113 p.

Lynn, S. K., and Reiss, D. L. (**1992**). "Pulse sequence and whistle production by two captive beaked whales Mesoplodon species," Marine Mammal Sci. **8**(3), 299–305.

Madsen, P. T., Johnson, M., Aguilar de Soto, N., Zimmer, W. M. X., and Tyack, P. (**2005**). "Biosonar performance of foraging beaked whales (*Mesoplodon densirostris*)," J. Exp. Biol. **208**, 181–194.

Mellinger, D. K. (**2001**). ISHMAEL 1.0 User's Guide. NOAA Technical Memorandum OAR-PMEL-120, available from NOAA/PMEL, 7600 Sand Point Way, NE, Seattle, WA 98115-6349.

Poulter, T. C. (**1968**) "Marine Mammals," in *Animal Communication; Techniques of Study and Results of Research*, edited by T. A. Sebeok (Indiana University Press, Bloomington), pp. 405–465.

Rankin, S., and Barlow, J. (**2007**). "Localization of a stationary sound source using a two-element towed hydrophone array." Manuscript available from Shannon Rankin, SWFSC, 8604 La Jolla Shores Drive, La Jolla, CA 92937 (20 pages).

Zimmer, W. M. X., Johnson, M. P., Madsen, P. T., and Tyack, P. L. (**2005**). "Echolocation clicks of free-ranging Cuvier's beaked whales (*Ziphius cavirostris*)," J. Acoust. Soc. Am. **117**(6), 3919–3927.

Acoustic scattering by a sphere with a hemispherically split boundary condition

Bradley E. Treeby,[a] Jie Pan, and Roshun M. Paurobally
Centre for Acoustics, Dynamics and Vibration, School of Mechanical Engineering, The University of Western Australia, 35 Stirling Highway, Crawley, WA 6009, Australia

(Received 22 December 2006; revised 4 April 2007; accepted 13 April 2007)

A general analytical model is developed for the scattering of sound by a sphere with a nonuniform impedance boundary condition that is divided into two uniformly distributed hemispheres. In addition to the overall solution for the time harmonic pressure, the analytical result gives insight into the modal contributions and coupling for different cases of source incidence and boundary impedance. Modal cross coupling is shown to exist between incoming and scattered wave modes of equi-order and nonequal degree when the degrees are opposite in parity (odd-even or even-odd coupling). This cross coupling is strongest between modes of adjacent degree, and decreases as the degrees become dissimilar. The overall magnitude of the cross coupling is dependent on the extent of the impedance mismatch between the two surface hemispheres. Simulation and discussion are given for several specific cases of source incidence and impedance (each hemisphere is given a different constant impedance value). These results are consistent with expectations from the scattering of sound by a sphere with a uniformly distributed surface boundary. The broad scattering characteristics of the hemispherically divided sphere are shown to be analogous to connecting the appropriate sectors from the corresponding uniformly distributed spheres. © *2007 Acoustical Society of America.* [DOI: 10.1121/1.2736784]

PACS number(s): 43.20.Fn, 43.20.El, 43.66.Pn [LLT] Pages: 46–57

I. INTRODUCTION

The scattering of sound by spherical objects has received considerable analytical treatment in the past century since Lord Rayleigh first discussed the solution for the scattering by a single rigid sphere.[1] Studies have been extended to penetrable and locally reacting spheres,[2–4] inhomogeneous spheres,[5] scattering in the presence of nearby reflecting boundaries,[6] scattering by multiple spheres,[7] and a multitude of other variations. Morse and Ingard[8] give a particularly thorough treatment of the sphere scattering problem. The strong historical interest in sphere scattering is due to the variety of physical phenomena it can approximate, including current applications in fluid mechanics, electro-magnetics, and virtual acoustics. With the advent of numerical methods, the interest in such analytical solutions has understandably somewhat declined. Investigations using boundary-element or finite-element methods are not constrained to simple shapes or uniformly distributed boundary conditions; limits which certainly afflict any analytic analysis attempting to provide solutions that avoid numerical integration. Additionally, investigation using previously established numerical methods generally does not require a detailed knowledge of mathematics or physical phenomenon. As such, these studies may be undertaken by a much larger number of researchers. However, while providing a convenient route to a final solution, such approaches are not particularly well suited to giving detailed insight into the underlying acoustical phenomena. In this regard analytical solutions enable investigation into modal contributions, modal coupling, and simplified expressions for limiting behavior. In some applications, analytical solutions also provide significantly reduced costs in CPU time and memory usage, particularly when information at high frequencies is required.[7]

This paper presents an analytical solution for the scattering of sound by a sphere with a hemispherically divided, locally reacting impedance boundary condition. Such a nonuniform surface impedance distribution was first discussed by Brungart, who presented a solution for the sphere scattering problem for a locally reacting surface boundary not necessarily uniformly distributed about the sphere surface.[9] This formulation was reliant on the numerical evaluation of a spherical harmonic integral dependent on the impedance distribution. For the boundary formulation discussed here, an analytic solution for the integral is developed. This utilizes the integral result of an associated Legendre function product of nonequal degree over a nonstandard interval presented by Hulme.[10] The complete analytic term embodies the modal coupling between the incident and scattered waves which results from the nonuniform impedance distribution. The solution is examined for two specific cases of wave incident direction in relation to the hemispherically divided boundary. Discussion is given in relation to the mathematical degeneration for certain angles of source incidence, modal contributions and coupling, and the pressure both on the sphere surface and in the vicinity of the scatterer.

While this paper approaches the scattering problem in a general manner, the boundary condition distribution utilized has particular relevance to studies of human hearing and virtual acoustics. In such work the head is commonly approximated as a spherical scatterer,[11,12] and it seems reasonable to assume that the surface boundary condition may be split

[a]Electronic mail: treebs@mech.uwa.edu.au

 0001-4966/2007/122(1)/46/12/$23.00

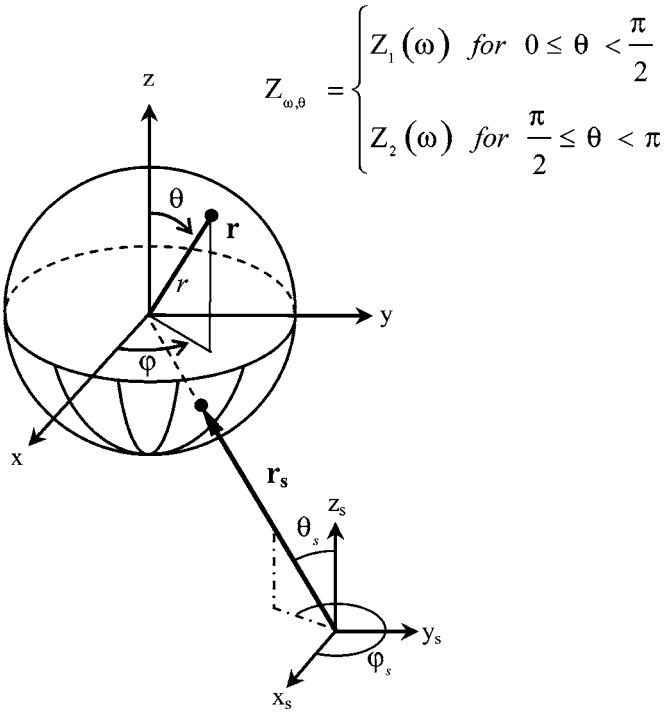


FIG. 1. General spherical coordinate system with the scatterer located at the polar origin. The location of the source and the direction of source propagation, relative to the polar origin, are characterized by $\mathbf{r}_s$. For a plane wave the source displacement from the scatterer is considered infinite and the source direction given by (θ_s, φ_s). The surface boundary condition is divided hemispherically, with the division being coincident with the *x*-*y* plane.

hemispherically into two parts, one covered in hair and one in skin. Previously, the effects of surface impedance (hair in particular) on such scattering have only been investigated using models assuming a uniformly distributed boundary condition,[12] or empirically using numerical and experimental studies.[13,14]

II. ANALYTICAL DERIVATION

A. Problem formulation

The sound field surrounding an object in a uniform and nonviscous medium can be decomposed into an incident and a scattered field, both of which are governed by the scalar wave equation (or Helmholtz equation if harmonic time dependence is assumed). For a given incident wave approaching a locally reactive sphere in an infinite medium, the scattered field is entirely determined by the boundary condition imposed on the surface of the sphere. Solutions of the incident and scattered waves are generally expressed as modal sums, with each mode having a characteristic angular dependence. Conceptually, the analytical treatment of a scattering problem attempts to evaluate the relative contribution of the mode shapes in the scattered wave based on a known boundary condition. Using the spherical coordinate system shown in Fig. 1, a general solution to the Helmholtz equation can be expressed as

$$\Psi(\mathbf{r}) = \sum_{n=0}^{\infty} \sum_{m=-n}^{n} G_n(kr) Y_n^m(\theta,\varphi). \tag{1}$$

Here, Ψ represents the complex pressure for a monochromatic wave $p(\mathbf{r},t)=\Psi(\mathbf{r})e^{-i\omega t}$, $\mathbf{r}=(r,\theta,\varphi)$ characterizes the evaluation position, k is the wave number, and G_n represents any general solution to the spherical Bessel differential equation. The latter constitutes a linear combination of spherical Bessel, spherical Neumann, and spherical Hankel functions chosen to satisfy the appropriate radiation conditions. Y_n^m represents the spherical harmonic function of degree n and order m, where

$$Y_n^m(\theta,\varphi) = y_n^m P_n^m(\cos\,\theta) e^{im\varphi}, \tag{2}$$

and the Condon-Shortley phase $(-1)^m$ has been included with the associated Legendre function P_n^m. The y_n^m term denotes the nonangular dependent normalization,

$$y_n^m = \sqrt{\frac{2n+1}{4\pi}\frac{(n-m)!}{(n+m)!}}. \tag{3}$$

Physically, the spherical harmonic functions represent spatially orthogonal wave types or "mode" shapes in an infinite medium. The characteristic angular dependence of each mode is analogous to the basic radiation patterns from the corresponding simple harmonic source (e.g., monopole, dipoles, quadrupoles, etc).[15] A modal sum can therefore be used to represent the scattering by, and radiation from a spherical object in free space. Note this definition of mode shape differs slightly from traditional vibration modeling, where a mode shape is associated with a particular physical resonance that defines the particle vibration amplitude ratios at the corresponding resonant frequency. Spherical harmonic functions are generally classified into three categories dependent on the shape of the division of an arbitrary spherical surface by the locus of the associated Legendre function.[16] Spherical harmonics of the form Y_n^0 are deemed *zonal* harmonics, $Y_n^{m<n}$ *tesseral* harmonics, and Y_n^n *sectorial* harmonics, as they divide the surface into zones, quadrilaterals (tessera), and sectors, respectively. Spherical harmonics of arbitrary degree and order form a complete orthogonal system over the spherical surface,

$$\int_0^{\pi}\int_0^{2\pi} Y_{n_1}^{m_1*}(\theta,\varphi) Y_{n_2}^{m_2}(\theta,\varphi) d\varphi \sin\,\theta\, d\theta = \delta_{n_1 n_2}\delta_{m_1 m_2}. \tag{4}$$

Both the incident and scattered waves can be expressed as incoming and scattered modal sums of spherical harmonics using Eq. (1). Staring with the incoming wave and following the general source approach taken by Gumerov and Duraiswami,[17] the incident wave expression is

$$\Psi_{in}(\mathbf{r}) = \sum_{n=0}^{\infty} \sum_{m=-n}^{n} E_n^m j_n(kr) Y_n^m(\theta,\varphi). \tag{5}$$

Here, j_n is the spherical Bessel function of the first kind (satisfying origin nonsingularity), and E_n^m is a general source term characterizing the source type and location. The source terms for a plane wave and a simple harmonic point source with source strength Q are, respectively,

$$E_n^m = Q4\pi i^n(-1)^m Y_n^{-m}(\theta_s,\varphi_s), \tag{6}$$

$$E_n^m = Qikh_n(kr_s)(-1)^m Y_n^{-m}(\theta_s,\varphi_s), \tag{7}$$

where $h_n \equiv h_n^{(1)}$ is the spherical Hankel function of the first kind, and the direction of wave propagation and source location is characterized by $\mathbf{r}_s=(r_s,\theta_s,\varphi_s)$, as illustrated in Fig. 1.

When the incident wave reaches the scatterer, secondary waves are produced that are outgoing from the scattering boundary. The corresponding expression for the scattered wave is

$$\Psi_{\text{scat}}(\mathbf{r}) = \sum_{n=0}^{\infty}\sum_{m=-n}^{n} A_n^m h_n(kr) Y_n^m(\theta,\varphi), \tag{8}$$

where h_n represents a solution to the Bessel equation that satisfies the outgoing (Sommerfeld) radiation condition (origin singularity implies the sphere must have a finite size). The complete expression is then simply the sum of the modal contributions from the incident and scattered waves,

$$\Psi(\mathbf{r}) = \sum_{n=0}^{\infty}\sum_{m=-n}^{n} Y_n^m(\theta,\varphi)[E_n^m j_n(kr) + A_n^m h_n(kr)]. \tag{9}$$

The unknown component of the modal contribution from the scattered wave A_n^m is solved based on the surface boundary condition. These boundary conditions are typically expressed in terms of acoustic impedance Z; the complex ratio of surface pressure to the normal radial velocity.[8] The velocity is produced either by the flow of the surrounding fluid through porous surface openings, or by motion of the surface boundary itself.[18] The impedance is dependent on the macroscopic material properties of the surface which are assumed to be time invariant. If the surface is locally reactive the boundary condition on the surface of a sphere of size $r=a$ can be expressed as

$$\left[\frac{\partial\Psi(\mathbf{r})}{\partial r} + \left(\frac{i\rho_0\omega}{Z(\omega,\theta,\varphi)}\right)\Psi(\mathbf{r})\right]_{r=a} = 0. \tag{10}$$

Here, ρ_0 is the mean medium density, and ω the angular frequency. Combining Eq. (9) and Eq. (10) gives the corresponding expression to solve for the unknown coefficients,

$$\sum_{n=0}^{\infty}\sum_{m=-n}^{n} Y_n^m(\theta,\varphi)\left\{[E_n^m kj_n'(ka) + A_n^m kh_n'(ka)] + \frac{i\rho_0\omega}{Z(\omega,\theta,\varphi)}[E_n^m j_n(ka) + A_n^m h_n(ka)]\right\} = 0, \tag{11}$$

where the $'$ operator denotes the radial derivative.

The form of the solution for the unknown scattered modal amplitudes A_n^m depends on the source term E_n^m and the distribution of the impedance term over the sphere surface. The solution can be obtained by multiplying Eq. (11) by the conjugate of a spherical harmonic with alternate degree and order, and then integrating over the sphere surface. For the product of two spherical harmonic functions with no other angular-dependent terms attached to it, this integral conveniently reduces using Eq. (4). This gives the general boundary condition for a single sphere with an arbitrarily distributed locally reactive impedance boundary condition,

$$\sum_{n=0}^{\infty}\sum_{m=-n}^{n}\left\{\delta_{nl}\delta_{mm'}[E_n^m kj_n'(ka) + A_n^m kh_n'(ka)] + \int Y_n^m(\theta,\varphi)Y_l^{m'^*}(\theta,\varphi)\frac{1}{Z(\omega,\theta,\varphi)}d\Omega \times i\rho_0\omega[E_n^m j_n(ka) + A_n^m h_n(ka)]\right\} = 0. \tag{12}$$

The remaining integral depends on the form of the impedance distribution. For an incident plane wave Eq. (12) is equivalent to the formulation discussed by Brungart,[9] who evaluates this integral numerically.

B. Solution of the spherical harmonic integral

If the impedance is assumed to be hemispherically divided and the x-y plane aligned with the hemispherical boundary as shown in Fig. 1, the boundary condition remains continuous with rotations in the spherical angle φ. Using this distribution of impedance [where $Z(\omega,\theta,\varphi)\equiv Z_{\omega,\theta}$] and expanding the remaining spherical harmonic integral in Eq. (12) using Eq. (2) leaves the integral of the product of associated Legendre functions of different degree over two nonstandard intervals,

$$\begin{aligned}\int Y_n^m(\theta,\varphi)Y_l^{m'^*}(\theta,\varphi)\frac{1}{Z_{\omega,\theta}}d\Omega &= y_n^m y_l^{-m'}(-1)^m\int_0^{\pi} P_n^m(\cos\theta)P_l^{-m'}(\cos\theta)\frac{1}{Z_{\omega,\theta}} \times \int_0^{2\pi} e^{i\varphi(m-m')}d\varphi\sin\theta\,d\theta \\ &= y_n^m y_l^m 2\pi\delta_{mm'} \times\left[\frac{1}{Z_1}\int_0^{\pi/2} P_n^m(\cos\theta)P_l^m(\cos\theta)\sin\theta\,d\theta + \frac{1}{Z_2}\int_{\pi/2}^{\pi} P_n^m(\cos\theta)P_l^m(\cos\theta)\sin\theta\,d\theta\right].\end{aligned} \tag{13}$$

Due to the orthogonality of associated Legendre functions over the sphere surface, the integral expressions evaluated for $n\neq l$ and $\cos\theta=\pm1$ will be zero. It is thus only the solution for $\cos\theta=0$ that is of interest and the integral results over the two intervals will be equal in magnitude but opposite in sign,

$$\begin{aligned}T_{n,l}^m &\equiv \int_0^{\pi/2} P_n^m(\cos\theta)P_l^m(\cos\theta)\sin\theta\,d\theta \\ &= -\int_{\pi/2}^{\pi} P_n^m(\cos\theta)P_l^m(\cos\theta)\sin\theta\,d\theta.\end{aligned} \tag{14}$$

The required analytic solution to Eq. (14) is provided by Hulme[10] in the context of fluid mechanics. The derivation reformulates the problem using the associated Legendre dif-

ferential equation and then utilizes the special results of $P_n^m(\mu)$ and its functional derivative for $\mu=0$. This yields

$$T_{n,l}^m = \frac{2^{2m+1}}{\pi(n-l)(n+l+1)}(H_{n,l}^m - H_{l,n}^m), \tag{15}$$

where

$$H_{n,l}^m = \frac{\left(\frac{n}{2}+\frac{m}{2}\right)!\left(\frac{l}{2}+\frac{m}{2}-\frac{1}{2}\right)!}{\left(\frac{n}{2}-\frac{m}{2}-\frac{1}{2}\right)!\left(\frac{l}{2}-\frac{m}{2}\right)!} \times \sin\left[\frac{1}{2}\pi(n+m)\right]\cos\left[\frac{1}{2}\pi(l+m)\right]. \tag{16}$$

An alternate closed-form solution to Eq. (14) derived independently by the authors is contained within the Appendix . This utilizes Rodrigues formula and integration by parts to yield a solution that is reliant on a finite summation of factorial terms rather than trigonometric functions. The solution provided by Hulme, however, is rather more elegant and yields a simpler final expression. Mathematically, the results are equivalent and give identical numerical results for any given degree and order satisfying $n \neq l$ and $|m| \leq \min(n,l)$.

Returning to Eq. (13), when $n=l$ the solution is significantly simpler. Due to the symmetry of $[P_n^m(\cos\theta)]^2$ about $\cos\theta=0$, the integral over the intervals $[0,\pi/2]$ and $[\pi/2,\pi]$ can be deduced from the orthogonality of the associated Legendre function over the domain $[0,\pi]$.[19] This gives

$$\int_0^{\pi/2}[P_n^m(\cos\theta)]^2\sin\theta\,d\theta = \int_{\pi/2}^{\pi}[P_n^m(\cos\theta)]^2\sin\theta\,d\theta = \frac{1}{2n+1}\frac{(n+m)!}{(n-m)!}. \tag{17}$$

Alternatively, the same result can be obtained analogous to the derivation for $n \neq l$ shown in the Appendix . Combining the integral results for $n \neq l$ and $n=l$ and simplifying, the complete closed-form solution to the spherical harmonic integral Eq. (13) can now be written as

$$\int Y_n^m(\theta,\varphi)Y_l^{m'^*}(\theta,\varphi)\frac{1}{Z_{\omega,\theta}}d\Omega = y_n^m y_l^m 2\pi\delta_{mm'}\left(\left(\frac{1}{Z_1}+\frac{1}{Z_2}\right)\frac{1}{2n+1}\frac{(n+m)!}{(n-m)!}\delta_{nl} + \left(\frac{1}{Z_1}-\frac{1}{Z_2}\right)T_{n,l}^m(1-\delta_{nl})\right) = \bar{\Upsilon}\delta_{nl}\delta_{mm'} + 2\pi y_n^m y_l^m \Delta\Upsilon T_{n,l}^m(1-\delta_{nl})\delta_{mm'}, \tag{18}$$

where $\bar{\Upsilon}$ is the admittance spatial average,

$$\bar{\Upsilon} = \frac{1}{2}\left(\frac{1}{Z_1}+\frac{1}{Z_2}\right), \tag{19}$$

and $\Delta\Upsilon$ is the admittance spatial disparity:

$$\Delta\Upsilon = \left(\frac{1}{Z_1}-\frac{1}{Z_2}\right). \tag{20}$$

C. Overall pressure solution

Using the general boundary condition expressed in Eq. (12), the overall expression for the hemispherically divided boundary condition is now

$$\sum_{n=0}^{\infty}\sum_{m=-n}^{n}\left\{\delta_{nl}[E_n^m k j_n'(ka) + A_n^m k h_n'(ka)] + [\bar{\Upsilon}\delta_{nl} + 2\pi y_n^m y_l^m \Delta\Upsilon T_{n,l}^m(1-\delta_{nl})] \times i\rho_0\omega[E_n^m j_n(ka) + A_n^m h_n(ka)]\right\}\delta_{mm'} = 0. \tag{21}$$

To explicitly solve for the unknown scattered modal amplitudes when the surface impedance characteristics of the two hemispheres are not equal, it is convenient to rearrange Eq. (21) in terms of the complex amplitude A_n^m of the $(n,m)^{th}$ scattering mode,

$$\sum_{n=0}^{\infty}\sum_{m=-n}^{n}\alpha_{l,n}^{m',m}A_n^m = \beta_l^{m'}. \tag{22}$$

In matrix form this becomes

$$\begin{bmatrix} \alpha_{0,0}^{0,0} & \alpha_{0,1}^{0,-1} & \alpha_{0,1}^{0,0} & \alpha_{0,1}^{0,1} & \cdots & \alpha_{0,n}^{0,m} \\ \alpha_{1,0}^{-1,0} & \alpha_{1,1}^{-1,-1} & & & & \alpha_{1,n}^{-1,m} \\ \alpha_{1,0}^{0,0} & & \alpha_{1,1}^{0,0} & & & \alpha_{1,n}^{0,m} \\ \alpha_{1,0}^{1,0} & & & \alpha_{1,1}^{1,1} & & \alpha_{1,n}^{1,m} \\ \cdots & & & & \cdots & \cdots \\ \alpha_{l,0}^{m',0} & \alpha_{l,1}^{m',-1} & \alpha_{l,1}^{m',0} & \alpha_{l,1}^{m',1} & \cdots & \alpha_{l,n}^{m',m} \end{bmatrix}\begin{bmatrix} A_0^0 \\ A_1^{-1} \\ A_1^0 \\ A_1^1 \\ \cdots \\ A_n^m \end{bmatrix} = \begin{bmatrix} \beta_0^0 \\ \beta_1^{-1} \\ \beta_1^0 \\ \beta_1^1 \\ \cdots \\ \beta_l^{m'} \end{bmatrix}, \tag{23}$$

and the simplified explicit coefficients can be written as

$$\alpha_{l,n}^{m',m} = k h_n'(ka)\delta_{nl}\delta_{mm'} + i\rho_0\omega h_n(ka) \times [\bar{\Upsilon}\delta_{nl} + 2\pi y_n^m y_l^m \Delta\Upsilon T_{n,l}^m(1-\delta_{nl})]\delta_{mm'}, \tag{24}$$

and

$$\beta_{l,m'} = -E_l^{m'}(k j_l'(ka) + i\rho_0\omega j_l(ka)\bar{\Upsilon}) - 2\pi i\rho_0\omega\Delta\Upsilon\sum_{n=0}^{\infty}E_n^{m'}j_n(ka)y_n^{m'}y_l^{m'}T_{n,l}^{m'}(1-\delta_{nl}). \tag{25}$$

The α matrix is square and the truncation number of outer sum is chosen to satisfy the convergence of Eq. (9) for a particular scattering frequency. The $\alpha_{l,n}^{m',m}$ terms represent the coupling coefficients between the $(l,m')^{th}$ incoming and the $(n,m)^{th}$ scattered modal components. The $(l=n,m'=m)$ terms give the autocoupling between identical modes (autocoupling coefficients), and the $(l\neq n,m'=m)$, $(l=n,m'\neq m)$,

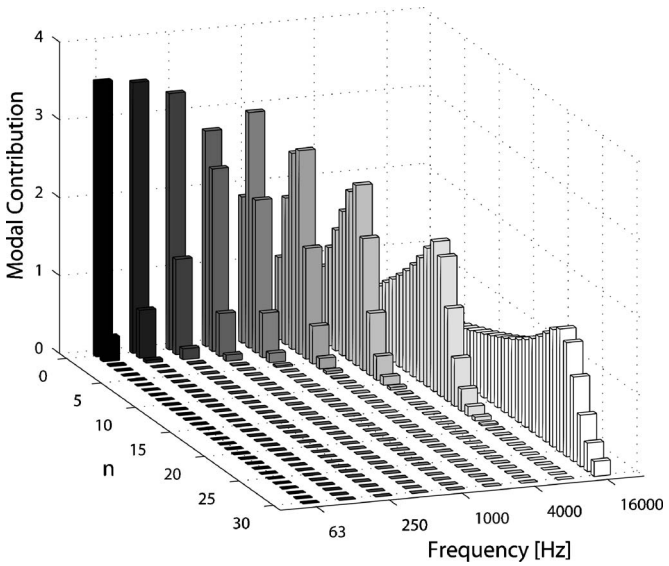


FIG. 2. The modal contribution $|E_n^0 j_n(ka)+A_n^0 h_n(ka)|$ of the first 30 zonal harmonic mode shapes to the total surface pressure of a rigid sphere.

and $(l \neq n, m' \neq m)$ terms give the cross coupling across modes of different degree and order (cross-coupling coefficients). The $\beta_{l,m'}$ terms give the $(l,m')^{th}$ incoming modal component which may excite the $(n,m)=(l,m')$ scattering mode directly through autocoupling, and the $(n,m) \neq (l,m')$ scattering modes through cross coupling.

D. Solution convergence

The convergence of Eq. (9) with n is of particular importance to the current problem, as the solution for A_n^m takes considerably longer as the size of the matrices used in the formulation becomes large. Brungart[9] gives a considerable proof of convergence for both arbitrarily distributed and axially symmetric boundary condition distributions. Convergence with n is consistent regardless of whether the boundary is uniform or hemispherically divided, or if the value of the surface impedance is changed. Figure 2 illustrates the relative modal contribution $|E_n^0 j_n(ka)+A_n^0 h_n^{(1)}(ka)|$ of the first 30 zonal harmonic mode shapes on the total surface pressure of a rigid sphere. Modal contributions are shown for octave band frequencies from 63 to 16 000 Hz for a plane-wave source propagating in the z direction $(\theta_s=\varphi_s=0)$ using a sphere radius of $a=0.0875$ m (this sphere radius is used in all the subsequent simulations throughout this paper). For the frequencies commonly used in sphere scattering problems (below 5000 Hz) the solution converges very quickly. For very low frequencies the response can be approximated by the first two modal contributions as discussed by Morse and Ingard.[8] At locations further from the sphere surface, the modal contribution is flattened and lengthened—at a distance of 1 diameter from the sphere surface the number of modal sums required for convergence is approximately doubled. For the simulations presented here, a sufficient number of summations are used to satisfy this convergence.

E. Coupling of incoming and scattered modes

When the sphere has a uniformly distributed boundary there is no coupling between incoming and scattered modes of unequal degree or order. The modal energy of the incoming wave is only used to excite identical modes in the scattered wave, and any energy loss is dependent only on the impedance boundary. This is also apparent upon inspection of the matrix formulation of the A_n^m solution given in Sec. II C. If $Z_1=Z_2$, then $\alpha_{l,n}^{m',m}=0$ for $l \neq n$ and/or $m \neq m'$ (the cross-coupling coefficients are all zero). This forces the coupling coefficient matrix in Eq. (23) to become diagonal. As a result the solution of A_n^m is dependent only on the corresponding incoming mode of equal degree and order.

When the impedance is not uniformly distributed about the sphere surface, each incoming mode no longer exclusively excites the identical scattering mode. Rather each scattering mode may be excited by many incoming modes. The corresponding scattered modal amplitudes A_n^m are thus dependent on the properties of the modal coupling coefficients in addition to the excitation strength of the incoming modes. Inspection of the coupling coefficient terms given by Eq. (24) illustrates that there is no coupling between modes of different order (i.e., the cross-coupling coefficients for $m \neq m'$ are always zero). This is an intuitive expectation given that the chosen impedance distribution ensures that the boundary is continuous (axially symmetric) with rotations in φ.

The remaining cross coupling is characterized by the spherical harmonic integral (coupling integral) given in Eq. (18). Table I gives the values of this integral for lower degree terms of zero order. For nonequal degree the coupling integral is only nonzero when the degrees are opposite in parity (only odd-even or even-odd cross-coupling exists). The cross coupling is strongest between modes of adjacent degree, and

TABLE I. Zero-order spherical harmonic integral (coupling integral) for a hemispherically divided boundary condition evaluated using Eq. (18). The nonzero off-diagonal terms $(n \neq l)$ characterize the relative coupling between the incoming and scattered radiation mode shapes of different degree.

	$\int Y_n^0(\theta_r,\varphi_r) Y_l^0(\theta_r,\varphi_r) \frac{1}{Z_{\omega,\theta}} d\Omega$					
	$n=0$	1	2	3	4	5
$l=0$	$\overline{\Upsilon}$	$0.4330\Delta\Upsilon$	0	$-0.1654\Delta\Upsilon$	0	$0.1036\Delta\Upsilon$
1	$0.4330\Delta\Upsilon$	$\overline{\Upsilon}$	$0.2421\Delta\Upsilon$	0	$-0.0541\Delta\Upsilon$	0
2	0	$0.2421\Delta\Upsilon$	$\overline{\Upsilon}$	$0.3698\Delta\Upsilon$	0	$-0.1448\Delta\Upsilon$
3	$-0.1654\Delta\Upsilon$	0	$0.3698\Delta\Upsilon$	$\overline{\Upsilon}$	$0.2790\Delta\Upsilon$	0
4	0	$-0.0541\Delta\Upsilon$	0	$0.2790\Delta\Upsilon$	$\overline{\Upsilon}$	$0.3498\Delta\Upsilon$
5	$0.1036\Delta\Upsilon$	0	$-0.1448\Delta\Upsilon$	0	$0.3498\Delta\Upsilon$	$\overline{\Upsilon}$

decreases as the degrees become dissimilar. The overall magnitude of the coupling integral terms for nonequal degree (cross coupling) is dependent on the extent of the impedance mismatch between the two surface hemispheres. The coupling integral terms for higher modal order are consistent with the lower order (i.e., the coupling between equi-order modes does not decrease as the order increases). The coupling integral terms for nonequal degrees of varying parity (e.g., $n=l-1$, $n=l-3$, etc.) approach finite limits as l becomes large. These limits can be deduced from the oscillatory behavior of the coupling integral, even for the small number of terms given in Table I. Again, the coupling integral limits for higher-order terms are consistent with those for the zero-order modes.

F. Axially incident source

When the incident wave is travelling in the z direction ($\theta_s=\varphi_s=0$), the source terms [Eqs. (6) and (7)] reduce significantly as

$$Y_n^m(0,0) = \sqrt{\frac{2n+1}{4\pi}}\delta_{m0}. \tag{26}$$

E_n^m will thus only be nonzero for zero order. As cross coupling does not exist between incoming and scattered modes of different order, the solution for A_n^m can be simplified accordingly as only zonal harmonics will be excited in the scattered wave. This result is not surprising considering the symmetry of this problem; only the zonal harmonic mode shapes are symmetrical for any arbitrary rotation about the z axis. The formulation of the unknown scattered modal amplitudes, dependent now only on zero-order terms, is reduced from a system of $(N+1)^2$ equations and unknowns to a system of $N+1$ equations and unknowns for a given sum truncation $n_{\max}=N$.

III. CHANGES IN THE SCATTERED SOUND FIELD

A. Descriptive analysis

For a uniformly distributed boundary, both the magnitude and phase characteristics of the complex surface impedance dictate changes to the scattered wave.[12] As the impedance magnitude is reduced from ∞ towards the characteristic impedance of the propagation medium, the magnitude of the scattered wave is decreased. Consequently, both the surface and proximal pressure are also reduced, particularly in the posterior. When the surface impedance is complex, the impedance phase angle dictates additional changes. For a negative phase angle the surface and posterior pressure are further decreased, and the curved anterior interference pattern (resulting from the interaction of the incident and scattered waves) is slightly shifted due to the phase difference in the scattered wave. A positive phase angle conversely causes the scattered wave to become more focused, and the posterior bright spot to be increased. The sensitivity of the surface pressure to changes in the impedance phase angle is dependent on both frequency and the overall impedance magnitude. As the impedance magnitude is reduced towards the characteristic impedance of the propagation medium, the range of increased sensitivity is translated to lower frequencies.

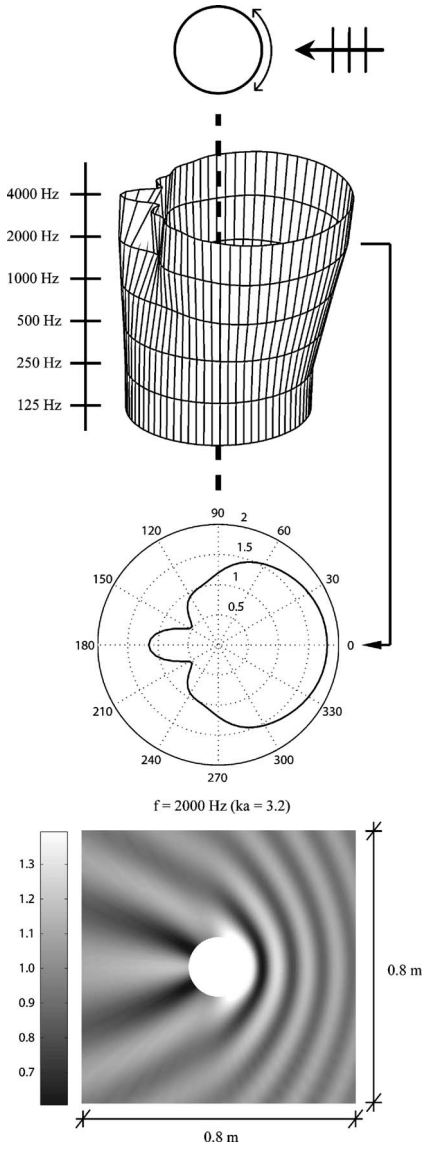


FIG. 3. Scattering characteristics for a rigid sphere for $a=0.0875$ m. The top panel shows a stacked polar plot of the surface pressure magnitude, the middle panel shows a polar plot of the surface pressure magnitude at 2000 Hz, and the bottom panel shows the pressure magnitude in the proximal region surrounding the sphere at the same frequency.

The scattering characteristics for a uniformly rigid and uniformly absorbent sphere are shown in Figs. 3 and 4. The corresponding characteristics for three specific arrangements of source incidence and circumferential rotation angle for a hemispherically divided surface impedance are shown in Figs. 5–7. The impedance distributions are denoted in each figure by the shaded (impedance) and unshaded (rigid) semicircular regions. The adjacent arrows show the source direction and circumferential angular rotation on the sphere surface that corresponds to the polar angle (with 0° always coinciding with the incident direction). For each case three plots are given; a stacked polar plot of the surface pressure magnitude (with each horizontal slice corresponding to octave band frequencies in the range from 125 to 4000 Hz), a

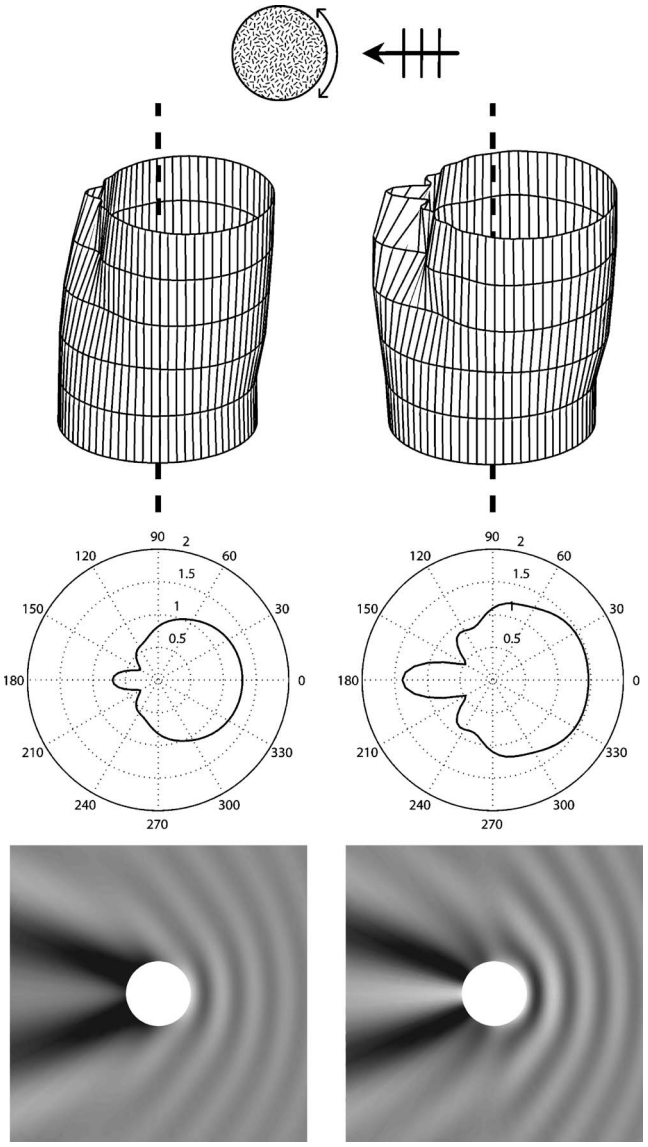


FIG. 4. Scattering characteristics (as in Fig. 3) for a uniformly absorbent sphere. The left panels correspond to an impedance value of $|\zeta|=2$, $\angle\zeta=0°$ and the right panels $|\zeta|=2$, $\angle\zeta=+45°$.

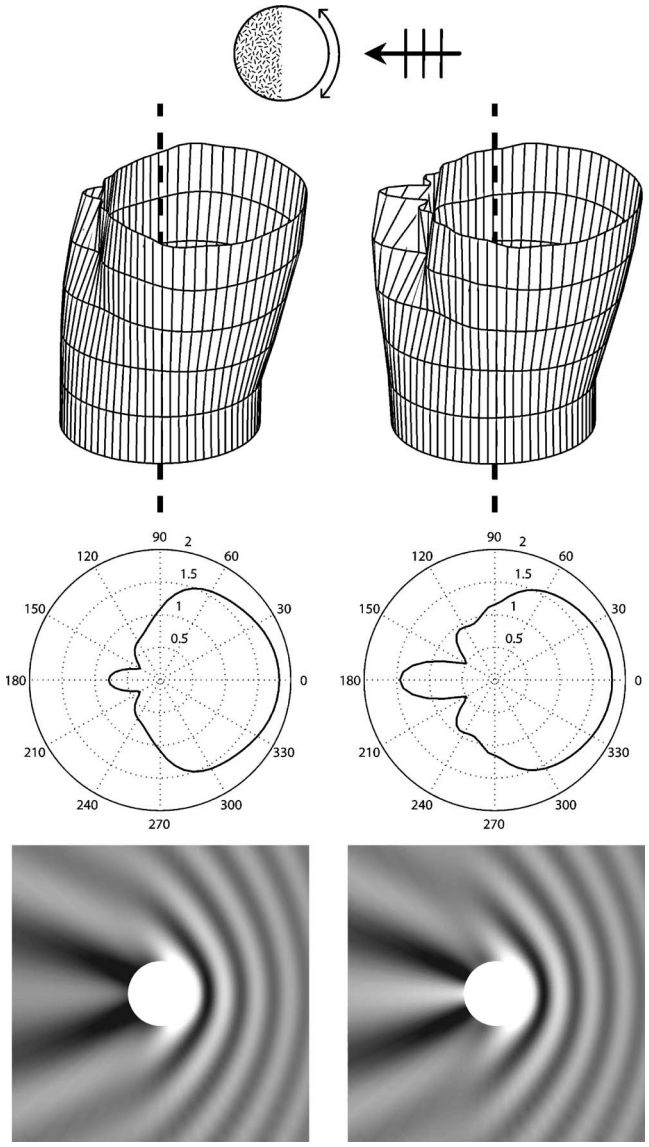


FIG. 5. Scattering characteristics (as in Fig. 3) for a sphere with a rigid anterior hemisphere and an absorbent posterior hemisphere. The left panels correspond to an impedance value of $|\zeta|=2$, $\angle\zeta=0°$ and the right panels $|\zeta|=2$, $\angle\zeta=+45°$.

polar plot of the surface pressure magnitude for $f=2000$ Hz ($ka=3.2$), and a proximal plot of the pressure magnitude over a 0.8×0.8-m region again for $f=2000$ Hz. The incident wave is plane and approaches from the right, and the dashed line on the stacked polar plot indicates the polar origin. The response at low frequencies is approximately unity and this provides a convenient reference for examining the trends at higher frequencies. In Figs. 4–7 the left panels correspond to an impedance value of $|\zeta|=2, \angle\zeta=0°$ and the right panels $|\zeta|=2, \angle\zeta=+45°$. Here, ζ is the specific acoustic impedance defined as $\zeta=Z/\rho_0c_0=|\zeta|e^{i\angle\zeta}$, with ρ_0c_0 the characteristic impedance of the propagation medium. The extraneous changes due to a negative impedance phase angle are not discussed as this case is less likely to occur considering the impedance coverings and frequency range generally of interest (most materials will be governed by their stiffness properties and hence have a positive impedance phase angle). However, these changes can be considered parallel to the relationship between the hemispherical and uniform coverings for a positive impedance phase angle.

Figure 5 shows the effect of reducing the impedance of the posterior hemisphere of a rigid sphere (relative to the incident wave) for a source alignment that causes the impedance distribution to be symmetric about the incident wave direction. In the rigid anterior hemifield the surface pressure remains remarkably consistent with a uniformly rigid sphere (shown in Fig. 3). On the surface there is a wide angular region of increased pressure, and in the proximal region the curved plane-incident and spherical-scattered wave interference pattern remains almost unperturbed. In the absorbent posterior hemifield, when the impedance is purely resistive (left panels), the surface pressure is reduced in a manner similar to the posterior from a uniformly covered sphere with the same impedance. The regions of sound shadow are widened and the bright spot is reduced. Direct comparison of the

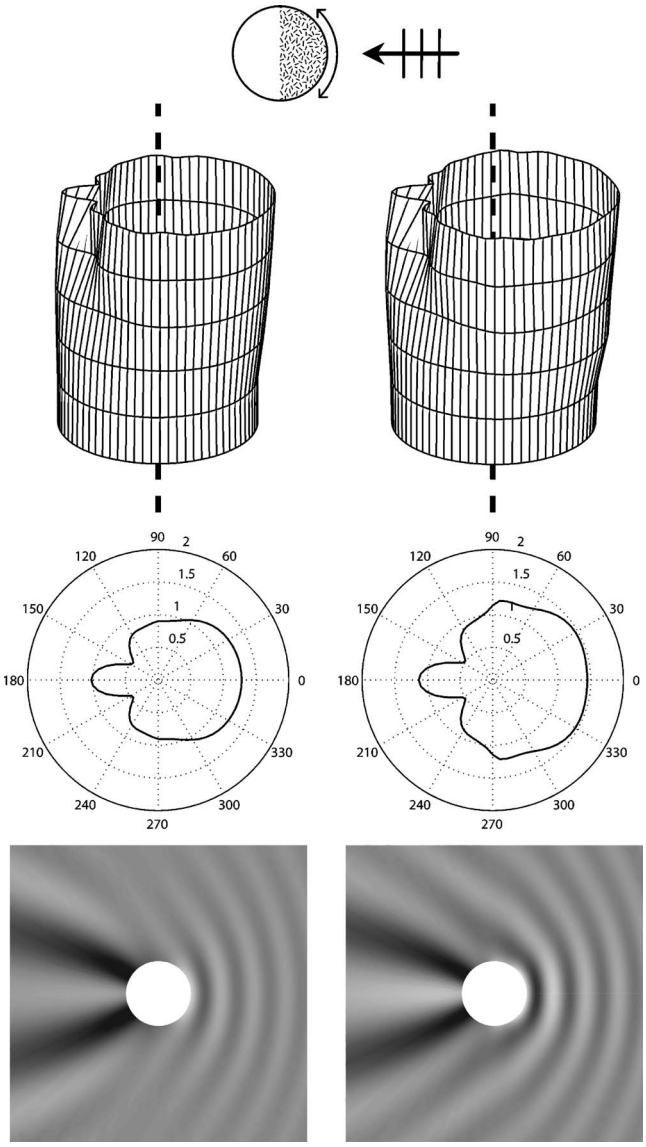


FIG. 6. Scattering characteristics (as in Fig. 3) for a sphere with a rigid posterior hemisphere and an absorbent anterior hemisphere. The left panels correspond to an impedance value of $|\zeta|=2$, $\angle\zeta=0°$ and the right panels $|\zeta|=2$, $\angle\zeta=+45°$.

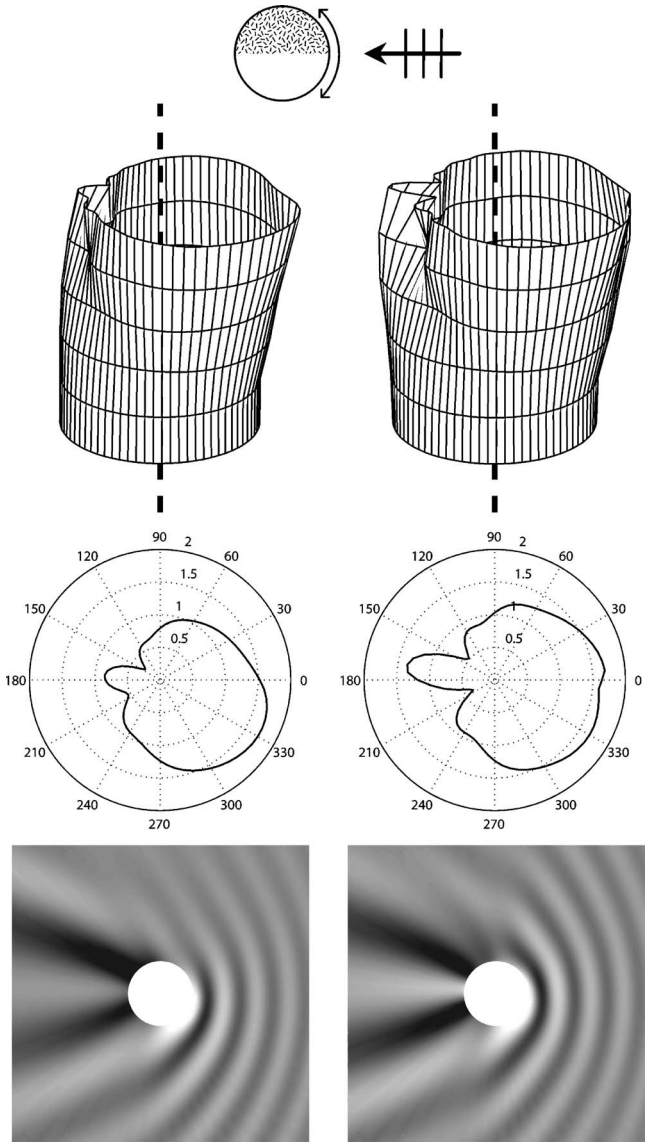


FIG. 7. Scattering characteristics (as in Fig. 3) for a sphere with a rigid lower hemisphere and an absorbent upper hemisphere. The left panels correspond to an impedance value of $|\zeta|=2$, $\angle\zeta=0°$ and the right panels $|\zeta|=2$, $\angle\zeta=+45°$.

bright spot with the uniformly absorbent sphere of the same impedance (shown in Fig. 4) reveals that this reduction is slightly less when the frontal hemisphere is rigid. This is due to the cumulative contribution of the anterior scattered waves which combine with the incident waves diffracted to the rear surface. The transition between the frontal and rear sectors is intuitive with smoothly joining the anterior of a uniformly rigid sphere with the posterior of a uniformly covered absorbent sphere.

For a uniform sphere, when the surface impedance is complex the effect of the impedance phase angle is most obvious in the posterior region. The surface and proximal pressure in the posterior region for the complex impedance and hemispherical distribution shown in Fig. 5 (right panels) are thus again very similar to the corresponding uniformly absorbent sphere. There is an increase in the magnitude of the dominant posterior bright spot lobe, and an exemplification of the ancillary lobes. The additional sidelobes of decreased pressure extending into the proximal region are also evident, although not as prominent as that for the uniform sphere.

Figure 6 illustrates the effect of reducing the impedance of the anterior hemisphere of a rigid sphere, again for a source alignment that causes the impedance distribution to be symmetric about the incident wave direction. The pressure in the posterior and anterior regions matches closely with the corresponding sectors from the appropriate uniformly covered spheres. For a purely resistive impedance (left panels), the proximal frontal region exhibits a decrease in the extent of the curved interference pattern due to the reduced magnitude of the scattered wave. In the posterior the surface pressure remains similar to the uniformly rigid case; although a small reduction is evident particularly in the bright spot. This is again due to the cumulative effect of the reduced scattered waves which combine with the incident waves diffracted to the rear surface. When the surface im-

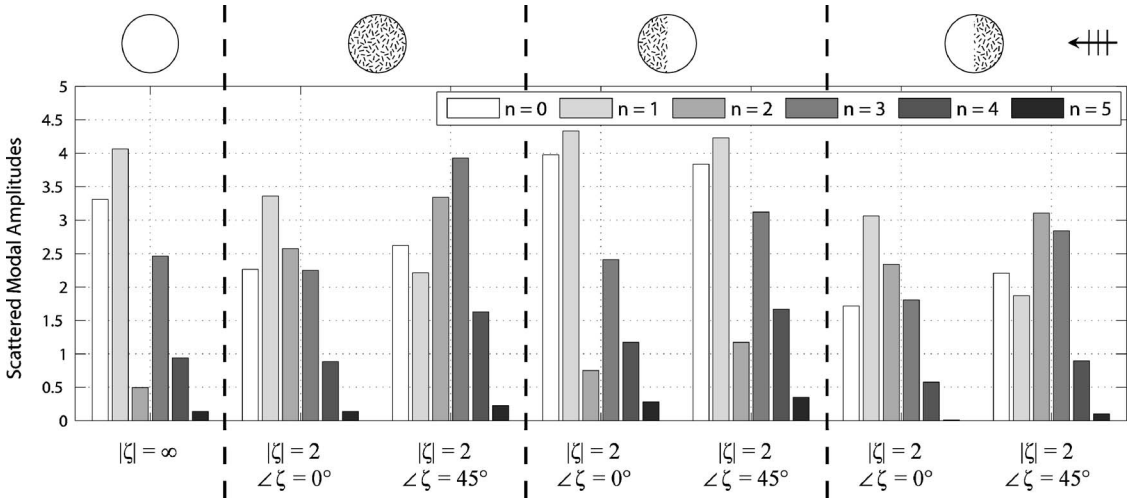


FIG. 8. Magnitudes of the zonal scattered modal amplitudes $|A_n^0|$ for impedance distributions symmetric about the incoming wave direction for $f=2000$ Hz.

pedance is complex and has a positive phase angle (right panels), the interference pattern in the proximal anterior region becomes more concentrated around the frontal median axis, and is slightly shifted due to the phase difference in the scattered wave. The pressure in the posterior is also slightly increased.

The final case shown in Fig. 7 illustrates the scattering characteristics for a source incident on the hemispherical boundary. For the circumferential rotation angle shown, the pressure distribution is no longer symmetric. Again, the broad characteristics of the scattering pattern are analogous to connecting the appropriate sectors from the corresponding uniformly distributed spheres. For a resistive impedance value (left panels) the curved interference pattern evident in the frontal near-field region is most prominent external to the rigid side. In the posterior the bright spot lobes become asymmetric as the interaction between symmetrically diffracted waves arriving in phase about the two hemispheres becomes more complex. The previously symmetric regions of sound shadow are noticeably wider and deeper adjoining the absorptive hemifield. When the upper hemisphere has a complex surface impedance with a positive impedance phase angle (right panels), the posterior bright spot is significantly increased in magnitude and is again asymmetrical, favoring the absorptive hemisphere. The sound shadow becomes deepest on the rigid side adjacent to the bright spot. The additional sidelobe of decreased pressure visible in Fig. 4 at around 105° and 255° is now only evident adjacent to the absorptive hemisphere. In the frontal near-field region the response becomes more symmetrical due to the increase in frontal pressure external to the impedance hemisphere. For the same incident wave alignment, the pressure for a circumferential rotation angle around the hemisphere boundary (not shown) can be considered analogous to a uniformly distributed sphere with an impedance value slightly higher in magnitude than that of the absorptive hemisphere.

Previous studies of the acoustic scattering by nonrigid spheres have been primarily concerned with mathematical treatment, and consequently there has been little discussion on the effect of impedance on scattering characteristics. In this regard Brungart[9] gives the most comprehensive discussion and examines several cases of impedance distribution for specific values of $ka=1$ with $kr=2$ and 20, and $ka=7$ with $kr=8$ and 20. For the sphere radius used in the present study (0.0875 m) this corresponds to evaluating the scattering pattern at 624 Hz at radii of 0.175 and 1.75 m, and at 4367 Hz at radii of 0.1 and 0.25 m. For uniformly distributed impedance boundaries the near-field frontal pressure shows an increase and the rear exhibits the dominant bright spot and associated sidelobes of decreased pressure. In the far field the anterior curved interference pattern is evident, with the magnitude of the fluctuations decreasing at distances further from the sphere. For a complex impedance value several additional "diffraction rings" are noted in the posterior hemifield. These features are consistent with the discussion and explanations given here. Similar results are also presented for hemispherical boundary orientations equivalent to Fig. 6 and Fig. 7. For the data given, the response over the two hemispheres appears to match the corresponding region from the appropriate uniform case, particularly in the near field. Again, this is consistent with the present discussion.

B. Modal analysis

The changes in the characteristics of the scattered waveform with impedance can also be explicated via analysis of the scattered modal amplitudes A_n^m. For convenience, it is easiest to investigate arrangements where the impedance distribution is symmetric about the incident wave direction. For such an alignment only zonal harmonics contribute to the response and the analysis is made somewhat simpler. The magnitudes of the complex scattered modal amplitudes for zero order are shown in Fig. 8 for the incident wave and boundary conditions shown in Figs. 3–6. When the surface is uniformly distributed, the individual modal contributions in the scattered wave are excited only by the corresponding incoming mode E_n^0. In the absence of cross coupling $|A_n^0|$ is always less than $|E_n^0|$. For the particular midfrequency discussed in detail throughout this paper, the scattered response for the rigid sphere is dominated by the zonal monopole (n

$=0$), dipole ($n=1$), and octopole modes ($n=3$). The dipole and octopole (odd degree) zonal modes feature major lobes that are opposite in sign and thus create the regions of increased anterior and decreased posterior pressure. The monopole, quadrupole, and other higher even degree zonal modes feature major lobes that are equal in sign and thus contribute to the posterior bright spot in addition to the anterior pressure. The width of the bright spot and the extent of additional ancillary perturbations in the posterior region are dependent on the degree of the dominant modes. Both even and odd zonal harmonic mode shapes become more directive for higher modal numbers, with the major lobes becoming thinner and gaining additional lobes of alternating sign. An increased contribution from higher modes thus produces a "focusing" effect in both anterior and posterior regions as well as additional features, particularly in posterior regions away from the median axis. For the rigid boundary and midfrequency case discussed, low degree terms dominate the response. This accounts for the regularity of the anterior waveform over a wide angle, and the relatively wide posterior bright spot exhibited in the corresponding response shown in Fig. 3.

When the sphere is no longer rigid, the impedance of the surface dictates the dissipation of sound power (absorption) in addition to a relative phase variation in the scattered wave. The absorption is at a maximum when the surface velocity is in phase with the forcing pressure oscillations. The absorption of a particular mode is dependent on the surface integral of the time-averaged power (squared complex pressure Ψ over impedance) for that mode. Overall, the modal contributions must combine such that the scattered wave conforms to the wave equation given the boundary condition present on the sphere surface. For a resistive surface impedance value of $\zeta=2$, the contribution of the monopole and dipole zonal modes is decreased, and the contribution of the zonal quadrupole mode increased. This corresponds with the overall decrease in the surface pressure, the thinner bright spot, and the slight focusing of the proximal anterior interference pattern. For the uniformly distributed absorbent sphere, at 2000 Hz a positive impedance phase angle significantly increases the contribution of the quadrupole and octopole modes, and decreases the contribution of the dipole mode. This change accounts for the increased focusing of the anterior interference pattern and the exemplification of the primary and ancillary posterior bright spot lobes (the dominant even mode is now a quadrupole rather than a monopole).

When the sphere boundary is hemispherically divided, modal contributions are additionally dependent on the coupling that exists between modes of varying degree. This coupling is dependent on the disproportion between the impedance of the two hemispheres, and only exists between equi-order modes that have degrees opposite in parity. Under certain conditions the modal power absorption may be positive as the coupling facilitates a scattered modal amplitude greater than the corresponding source mode. For example this situation arises for the monopole mode at 2000 Hz for the hemispherically divided sphere with a rigid frontal surface for either impedance value. The relative modal contributions for the axially symmetric cases of a hemispherically divided surface are shown in the right panels of Fig. 8. The contributions of lower degree modes are consistent with the corresponding uniform case with matching frontal impedance. Given the preceding modal discussion, the additional modal trends can be mapped to the changes evident in both the anterior and posterior pressure in a reasonably straightforward manner. For hemispherical distributions that are not symmetric about the incident wave, tesseral and sectorial mode shapes also contribute to the response. Whilst displaying more complex symmetries (analogous to the multiple radiation patterns for simple sources), the relative contribution of these modes in relation to the overall response can be considered analogous to the discussion already given.

IV. CONCLUSION AND DISCUSSION

Analytical treatments of absorptive sphere scattering have traditionally considered only a uniformly distributed boundary condition. This work extends the available analytical results to include scattering when the sphere has a hemispherically divided boundary condition. In addition to the overall solution for the time harmonic pressure, the analytical result allows insight into the modal contributions and coupling for different cases of source incidence and boundary impedance. Modal cross coupling is shown to exist between incoming and scattered wave modes of equi-order and nonequal degree when the degrees are opposite in parity (odd-even or even-odd coupling). This cross coupling is strongest between modes of adjacent degree, and decreases as the degrees become more disparate. The overall magnitude of the coupling is dependent on the extent of the impedance mismatch between the two surface hemispheres. Cross coupling does not exist between modes of different order.

Considerable discussion is also given in relation to the surface and near-field pressure as the impedance of the hemispheres is altered in relation to the incident wave direction. These changes are explicated through both descriptive and modal analysis. The broad scattering characteristics of the hemispherically divided sphere are analogous to connecting the appropriate sectors from the corresponding uniformly distributed spheres. The scattering characteristics are consistent with expectations from previous investigations using a uniformly distributed boundary condition which have been experimentally validated,[12] and discussion given for similar scattering problems evaluated using numerical integration.[9] The analytical formulation has also been experimentally validated using several hemispherical impedance coverings on a rigid sphere. These supplementary results will form part of an additional paper focused on the application of the hemispherical solution.

Implementation of the presented solution is relatively straightforward, particularly for researchers who are familiar with traditional scattering solutions formed as infinite sums of eigenfunctions. No additional mathematical functions are required beyond those needed for the uniformly distributed impedance problem (except perhaps the associated Legendre function in place of the Legendre polynomial). However, the unknown scattered modal amplitudes must be solved using Gaussian elimination rather than formulated explicitly. No

attempt is made to address the question of efficiently implementing the hemispherical boundary solution, as this issue is one of computational rather than physical mathematics. It is left up to the future researcher to decide upon an appropriate implementation; however, considerable discussion on efficient algorithm development already exists in the literature.[17]

The current analytical treatment has particular relevance to studies of human hearing mechanisms and binaural sound synthesis, as the hemispherically divided surface boundary condition matches the broad anthropometric composition of a human head with hair. The widespread exploitation of the sphere scattering problem however, in electro-magnetics and fluid mechanics for example, raises the possibility that the solution may also find use in other applications.

ACKNOWLEDGMENTS

B.E.T. would like to acknowledge the support of the Robert and Maude Gledden and F.S. Shaw Memorial Postgraduate Scholarships. The authors would also like to thank the anonymous reviewers and associate editor for drawing attention to the mathematical treatment by Hulme.

APPENDIX: ALTERNATE DERIVATION OF LEGENDRE FUNCTION INTEGRAL

Starting with the integral expressed in Eq. (14), the associated Legendre functions can be expanded using Rodrigues' formula,[19] making the substitution $\mu=\cos\theta$. Integrating by parts $l+m$ times assuming without loss of generality that $l>n$ gives

$$\int_{\gamma 1}^{\gamma 2} P_n^m(\mu)P_l^m(\mu)d\mu = \frac{(-1)^{n+l}}{2^{n+l}n\,!\,l!}\left[\sum_{v=0}^{l+m-1}\frac{d^v}{d\mu^v}\left((1-\mu^2)^m\left(\frac{d^{n+m}}{d\mu^{n+m}}(1-\mu^2)^n\right)\right)\times(-1)^v\left(\frac{d^{l+m-v-1}}{d\mu^{l+m-v-1}}(1-\mu^2)^l\right)\right]_{\gamma 1}^{\gamma 2}. \quad \text{(A1)}$$

The first derivative component can be further expanded using Leibniz' formula[19] for multiple derivatives of a product,

$$\int_{\gamma 1}^{\gamma 2} P_n^m(\mu)P_l^m(\mu)d\mu = \frac{(-1)^{n+l}}{2^{n+l}n\,!\,l!}\left[\sum_{v=0}^{l+m-1}\left(\sum_{q=0}^{v}\frac{v!}{(v-q)\,!\,q!}\frac{d^{v-q}}{d\mu^{v-q}}(1-\mu^2)^m \times\frac{d^{n+m+q}}{d\mu^{n+m+q}}(1-\mu^2)^n\right)(-1)^v\left(\frac{d^{l+m-v-1}}{d\mu^{l+m-v-1}}(1-\mu^2)^l\right)\right]_{\gamma 1}^{\gamma 2}. \quad \text{(A2)}$$

The expression is to be evaluated over the intervals $[0,1]$ and $[-1,0]$. The term vanishes for ± 1, so only the solution for $\mu=0$ is of interest, and the integral over the two intervals will be opposite in sign. After some considerable investigation the general equation for the sth derivative of the term $(1-\mu^2)^p$ evaluated for $\mu=0$ is obtained as

$$\left[\frac{d^s}{d\mu^s}(1-\mu^2)^p\right]_{\mu=0} = \frac{1+(-1)^s}{2}\frac{i^s p\,!\,s!}{(p-s/2)\,!\,(s/2)!}. \quad \text{(A3)}$$

Letting

$$s_1 = l+m-v-1, \quad s_2 = v-q, \quad s_3 = n+m+q, \quad \text{(A4)}$$

and simplifying the substitution of Eq. (A3) into Eq. (A2) for $\mu=0$ and $l>n$ gives the final solution of the integral for the case of $l>n\geq m\geq 0$,

$$T_{n,l}^m = \frac{i^{3l+3n+2m+1}m!}{2^{n+l+3}}\sum_{v=0}^{l+m-1}\frac{(-1)^v v\,!\,(1+(-1)^{s_1})s_1!}{\left(l-\frac{s_1}{2}\right)!\left(\frac{s_1}{2}\right)!} \times\sum_{q=0}^{v}\frac{(1+(-1)^{s_2})(1+(-1)^{s_3})s_2\,!\,s_3!}{(v-q)\,!\,q\,!\left(m-\frac{s_2}{2}\right)!\left(\frac{s_2}{2}\right)!\left(n-\frac{s_3}{2}\right)!\left(\frac{s_3}{2}\right)!}. \quad \text{(A5)}$$

Although the solution for $T_{n,l}^m$ stated here is only valid for $l>n\geq m\geq 0$, the initial assumption of $l>n$ is arbitrary and thus solutions of the integral for $l<n$ can be computed by $T_{l,n}^m$. Using the properties of associated Legendre functions, results for negative order can be obtained by

$$T_{n,l}^{-m} = \frac{(n-m)!}{(n+m)!}\frac{(l-m)!}{(l+m)!}T_{n,l}^m. \quad \text{(A6)}$$

The closed-form integral result obtained here utilizes a alternate mathematical approach to that presented by Hulme.[10] While the derivation provided by Hulme yields a simpler final expression, the two solutions generate identical numerical results.

[1] J. W. Rayleigh, *The Theory of Sound* (Dover, New York, 1945).
[2] M. Lax and H. Feshbach, "Absorption and scattering for impedance boundary conditions on spheres and circular cylinders," J. Acoust. Soc. Am. **20**, 108–124 (1948).
[3] R. W. Hart, "Sound scattering of a plane wave from a nonabsorbing sphere," J. Acoust. Soc. Am. **23**, 323–329 (1951).
[4] G. Kear, "The scattering of waves by a large sphere for impedance boundary conditions," Ann. Phys. **6**, 102–113 (1959).
[5] P. A. Martin, "Acoustic scattering by inhomogeneous spheres," J. Acoust. Soc. Am. **111**, 2013–2018 (2002).
[6] K. M. Li and W. K. Lui, "The diffraction of sound by an impedance sphere in the vicinity of a ground surface," J. Acoust. Soc. Am. **115**, 42–56 (2004).
[7] N. A. Gumerov and R. Duraiswami, "Computation of scattering from *N* spheres using multipole reexpansion," J. Acoust. Soc. Am. **112**, 2688–2701 (2002).
[8] P. M. Morse and K. U. Ingard, *Theoretical Acoustics* (McGraw-Hill, New York, 1968).
[9] D. L. Brungart, *Acoustic Scattering from a Sphere with Nonuniform Surface Impedance* (Aerospace Medical Research Laboratory, Wright Patterson Air Force Base, OH, 1974).
[10] A. Hulme, "The wave forces acting on a floating hemisphere undergoing forced periodic oscillations," J. Fluid Mech. **121**, 443–463 (1982).
[11] V. R. Algazi, R. O. Duda, R. Duraiswami, N. A. Gumerov, and Z. Tang, "Approximating the head-related transfer function using simple geometric models of the head and torso," J. Acoust. Soc. Am. **112**, 2053–2064 (2002).
[12] B. E. Treeby, R. M. Paurobally, and J. Pan, "The effect of impedance on interaural azimuth cues derived from a spherical head model," J. Acoust. Soc. Am. **121**, 2217–2226 (2007).
[13] B. F. G. Katz, "Boundary element method calculation of individual head-related transfer function. II. Impedance effects and comparison to real

measurements," J. Acoust. Soc. Am. **110**, 2449–2455 (2001).
[14]K. A. J. Reiderer, "HRTF analysis: Objective and subjective evaluation of measured head-related transfer functions," Ph.D. dissertation, Helsinki University of Technology, Espoo, 2005.
[15]J. Pan, "A note on sound radiation from a confined source," J. Sound Vib. **237**, 163–172 (2000).
[16]E. W. Hobson, *The Theory of Spherical and Ellipsoidal Harmonics* (University Press, Cambridge, 1931).
[17]N. A. Gumerov and R. Duraiswami, *Fast Multipole Methods for the Helmholtz Equation in Three Dimensions* (Elsevier, Oxford, 2004).
[18]P. M. Morse and R. H. Bolt, "Sound waves in rooms," Rev. Mod. Phys. **16**, 69–150 (1944).
[19]G. B. Arfken and H. J. Weber, *Mathematical Methods for Physicists* (Elsevier, Burlington, 2005).

Two-dimensional modeling of wave propagation in materials with hysteretic nonlinearity

Sigfried Vanaverbeke and Koen Van Den Abeele[a)]
Laboratory for Nonlinear Wave Spectroscopy, Interdisciplinary Research Center, K.U.Leuven Campus Kortrijk, Etienne Sabbelaan 53, 8500 Kortrijk, Belgium

(Received 25 August 2006; revised 29 March 2007; accepted 22 April 2007)

A multiscale model for the two-dimensional nonlinear wave propagation in a locally microdamaged medium is presented, and numerical simulations are analyzed in view of nondestructive testing applications. The multiscale model uses a statistical distribution of hysterons and upscales their microscopic stress-strain relations to a mesoscopic level. Macroscopic observations are then predicted by finite integration techniques. The influence of a small region with hysteretic nonlinearity on the generation of harmonics is investigated, and numerical results for different amplitudes of the input signal and different analysis techniques of the response signal are presented. Second, a study is conducted on the interaction of a Rayleigh wave with a microdamaged zone with hysteretic nonlinearity at the surface of an otherwise linear body, and the influence of the microdamaged zone on the surface wave velocity and on the generation of harmonics is examined. It is found that the effect of hysteresis on the Rayleigh wave propagation can be barely seen in the surface wave velocity measurement, but shows up nicely in the wave spectrum. The potential of a nonlinearity based depth profiling technique is explored by evaluating the nonlinear responses at different frequencies for a vertically stratified medium with spatially varying hysteresis properties. [DOI: 10.1121/1.2739803]

PACS number(s): 43.25.Dc, 43.25.Fe, 43.25.Ba [MFH] Pages: 58–72

I. INTRODUCTION

Nature accommodates bizarre and unique elastic systems. Rocks and microcracked solids, for example, display an extremely nonlinear response in quasistatic stress-strain tests, including hysteresis and end-point memory. Several experiments performed in the last decade have provided abundant evidence that the dynamics of these materials is also affected in a sizable way by the manifestation of nonlinearity.[1–5]. Generally, for materials in a certain stage of damage, the effects of the nonuniqueness of the stress-strain relation is more pronounced than the classical nonlinearity which is linked to the potential anharmonicity.[3–5]

Modeling wave propagation in microcracked solids with proper consideration of nonlinearity, and in particular hysteresis, presents significant challenges. Over the years a large variety of numerical models have been developed to study wave propagation in *linear* microinhomogeneous solids. The numerical techniques which are used in these models are well developed and documented, and include finite element methods, boundary element methods, finite difference methods, and finite integration methods. Commercially available computer codes can now efficiently and routinely handle mode conversions, anisotropy, inhomogeneities, inclusions, and boundaries in locally inhomogeneous media, but are usually limited to the case of *linear* wave propagation and resonance simulations.

Including the presence of nonunique stress-strain relations for the simulation of elastodynamic wave propagation into the existing codes has been attempted so far only by a few researchers and most efforts in this field are limited to one-dimensional simulations. We mention here in particular the work of Nazarov *et al.*,[1,6] Delsanto and Scalerandi,[7] Scalerandi *et al.*,[8] Gusev and Aleshin,[9] and Van Den Abeele *et al.*[10,11] In higher dimensions, the modeling of the behavior of every individual inhomogeneity on the microlevel puts very high demands on the calculation time on current computer systems.[12] It is therefore advantageous to introduce an intermediate level that is bridging the microscopic and the macroscopic (or sample) level.[11] This intermediate (or mesoscopic) level can be regarded as the element or cell level defined in finite element or difference codes for wave propagation. At the mesoscopic level, we consider each element as a statistical ensemble of hysterons. Hysterons are microscopic features with an elementary hysteretic stress-strain relation allowing only two states. The nonlinear and hysteretic stress-strain relation at the mesoscopic level can then be computed by means of the Preisach–Krasnosel'skii–Mayergoyz (PKM) model.[13–15] Once the procedure to update the stress-strain relation at the mesoscopic level has been successfully implemented, one can proceed to compute the macroscopic response with standard finite difference, finite integration, or finite element methods. In Ref. 11, Van Den Abeele *et al.* proposed a multiscale model of this kind for one-dimensional (1D) nonlinear wave propagation. The predictions at the macroscopic level were based on the elastodynamic finite integration technique (EFIT), which was originally developed by Fellinger *et al.*[16] and is now well established for two- and three-dimensional (2D and 3D) simulations in *linear* solids.[17–19] Van Den Abeele *et al.* coupled the 1D EFIT model to the general description of stress-strain hysteresis using a Preisach model[11] within the

[a)]Electronic mail: koen.vandenabeele@kuleuven-kortrijk.be

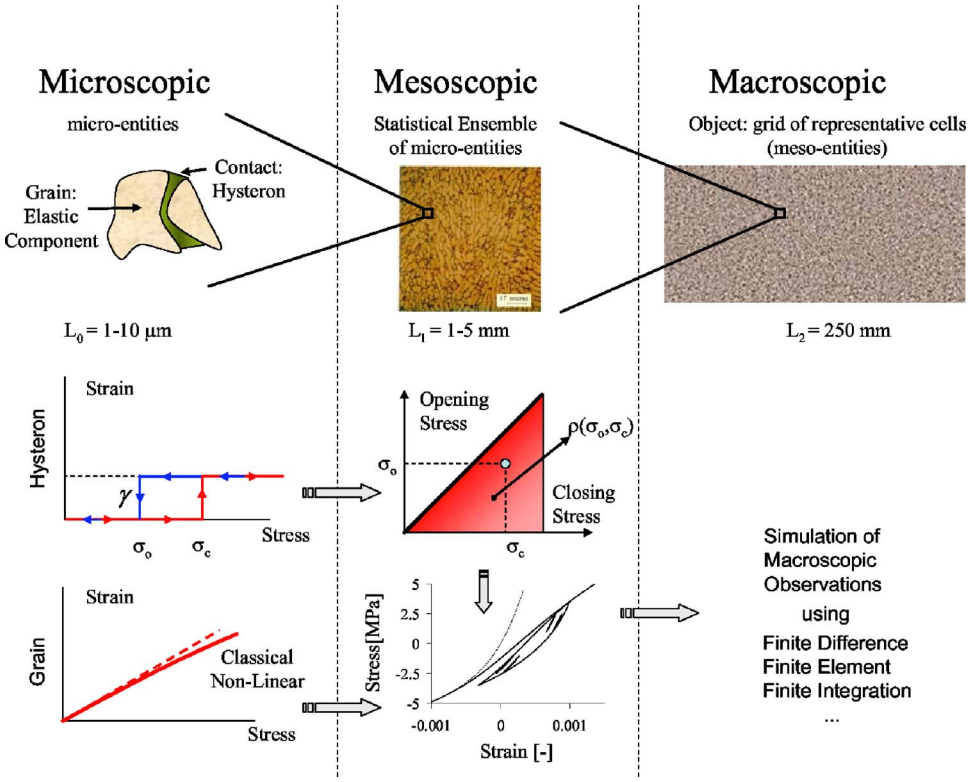


FIG. 1. (Color online) Illustration of the multiscale concept used in the simulations, with the macroscopic, mesoscopic, and microscopic levels and their corresponding constitutive relations.

above-described multiscale approach. Along the same lines, Bou Matar *et al.*[20] have recently developed a similar approach based on a pseudospectral wave solver. In addition, a multiscale version of the local interaction simulation approach (LISA) model for nonlinear wave propagation in 1D and higher dimensions has been reported[21,22] for the investigation of the influence of localized damage on the nonlinear behavior of resonance modes in 1D and on the wave propagation characteristics in 2D.

In this paper, we report simulation results obtained with the extension of our previously reported multiscale nonlinear and hysteretic 1D-EFIT approach to the problem of 2D wave propagation. In Sec. II, we upscale the microscopic state relation to the mesoscopic level by applying a generalization of the Preisach-space model for multidimensional hysteretic elasticity based on the work of Helbig,[23,24] and introduce the staggered grid formulation to predict the macroscopic response to an arbitrary excitation signal. In Sec. III, two types of numerical simulations are performed. First, we consider the problem of an in-plane pulse propagation and investigate the influence of a small microdamaged zone with hysteretic nonlinearity on the generation of harmonics in the signal. In a second illustration, we focus on the interaction of a Rayleigh wave propagating along the surface of a solid body with a microcracked region, represented by a zone with hysteretic stress-strain relations. We investigate the influence of the microdamaged zone on the surface wave velocity, and on the generation of harmonics.

The justification for this research lays in the search and development of appropriate quality control techniques for nondestructive testing of microdamaged materials. Durability and safety is a high priority issue in our current society, for instance in aeronautics. In our opinion nonlinear ultrasonic methods based on bulk and Rayleigh wave propagation will increase the sensitivity for detecting incipient cracks at early stages of deterioration and fatigue processes. Numerical methods can assist in a variety of tasks, such as interpreting the results, defining new methodologies and optimizing detector and source positions.

II. 2D MULTISCALE APPROACH FOR SIMULATING WAVE PROPAGATION IN THE PRESENCE OF HYSTERETIC NONLINEARITY

The multiscale concept used in this paper is schematically illustrated in Fig. 1. We consider a macroscopic sample built up by a finite number of representative mesoscopic material cells or elements. Each cell is composed of a large number of microscale units or hysterons which represent individual grains with grain contacts, bonding system, interphase zones, microcracks, etc. The microscopic units have a typical length of $0.1-100\ \mu$m. Typical mesoscopic cell lengths are of the order of a few millimeters. Macroscopic sizes are no real issue. Our multiscale model involves two upscaling tasks: First, the upscaling of the strain response of the individual hysterons at the microscopic level to the mesoscopic (or element) level, and second, the upscaling of the stress-strain relation at the mesoscopic level to the macroscopic (or sample) level. These two steps are described in Secs. II A and II B.

A. Scalar PKM-space model for hysteretic stress-strain behavior

The quasistatic stress-strain relation of a wide class of microinhomogeneous materials generally reveals a nonunique strain response to stress increase and stress decrease,

even in the purely elastic (nonplastic) regime. As a consequence of this type of mechanical hysteresis, interior loops can be observed with evidence for end-point memory.[24] As mentioned before and supported by experiments,[1–5] the nonuniqueness of the stress-strain relation for microcracked materials generally dominates the classical nonlinearity which is due to the potential anharmonicity of the individual constituents. Since the description of the hysteretic phenomena during a quasi-static experiment forms the basis of our dynamic multiscale model, we review briefly the key elements of the underlying phenomenological model. The model is based on the work by McCall and Guyer,[25–27] who translated the ideas of Preisach,[13] Krasnosel'skii,[14] and Mayergoyz[15] to the field of rock elasticity mechanics.

Let us consider a mesoscopic cell containing a large number of microscopic units. At the microscale, we model the strain response of each individual unit by a combination of a classical nonlinear stress-strain relation, and a nonclassical contribution attributed to hysteresis effects (i.e., nonuniqueness in the stress-strain relation):

$$\varepsilon = \varepsilon_C + \varepsilon_H. \tag{1}$$

The classical strain component ε_C can be regarded as the traditional reversible nonlinear power law relation between stress σ and strain ε which is used, for example, to describe classical nonlinearity in liquids and the potential anharmonicity in single crystals.[28] In this case, the modulus K_C can be expressed as a power series of stress: $K_C(\sigma)=K_0(1+\beta\sigma+\delta\sigma^2+\cdots)$, where K_0 is the linear modulus value and β and δ are combinations of second-, third-, and fourth-order elastic constants defining the classical nonlinearity. Upon an arbitrary pressure change ΔP (pressures P are assumed positive in compression, stresses σ are negative, and $\sigma=-P$), the classical contribution to the overall strain ε is given

$$\Delta\varepsilon_C = \int_{-P-\Delta P}^{-P} \frac{dP'}{K_C(-P')} = \int_{\sigma}^{\sigma+\Delta\sigma} \frac{d\sigma'}{K_C(\sigma')}. \tag{2}$$

To quantify the hysteretic strain component $\Delta\varepsilon_H$, we assume a simple phenomenological state relation in which the strain contribution equals either zero or a (finite) constant value γ depending on the actual pressure value P (or stress value σ) and its history. This behavior can be considered as the formal representation of the instantaneous and discontinuous closure and opening of a crack or a bond system between grains.[26,27] More physical based models for hysterons in dynamic and static elasticity attribute the behavior to the adhesion hysteresis of individual asperities and the friction of crack surfaces.[29–32]

While increasing the pressure, the strain in the hysteron is zero for $P<P_c$ ($\sigma>\sigma_c$) ("open" state), and $-\gamma$ for $P>P_c$ ($\sigma<\sigma_c$) ("closed" state). When decreasing the pressure, the strain remains $-\gamma$ for $P>P_o$ ($\sigma<\sigma_o$) ("closed" state) and reverts to zero for $P<P_o$ ($\sigma>\sigma_o$) ("open" state). Naturally, we have that $P_c>P_o$. For simplicity we assume that only the parameters P_o and P_c (or σ_o and σ_c) can vary from hysteron to hysteron. We further assume that all other parameters remain constant within a material cell. This means that, if a mesoscopic cell contains N hysterons, the four constants K_0, β, δ, and $\hat{\gamma}=\gamma N$ can be considered as "effective" constants at the mesoscopic level. As a consequence, each hysteron within a representative cell can be represented in a mesoscopic stress-stress space according to its associated values of P_o and P_c. This representation is commonly termed "P(K)M-space," and can be dealt with mathematically by its density distribution $\rho(P_c,P_o)$.[10,11,25–27] The four constants K_0, β, δ, and $\hat{\gamma}$ together with the PKM density distribution form a unique signature of each material cell, which may obviously differ from cell to cell. The hysteretic strain contribution from a mesoscopic cell with N hysterons in response to a pressure change ΔP can then be expressed as follows:

$$\begin{aligned}\Delta\varepsilon_H &= -\hat{\gamma}\,(f_c(P+\Delta P)-f_c(P)) \\ &= \hat{\gamma}\,(f_c(-\sigma)-f_c(-\sigma-\Delta\sigma)),\end{aligned} \tag{3}$$

with $f_c(P)$ the fraction of hysterons in the closed state at pressure P. This fraction can be calculated as the integral of the density distribution over the area of closed hysterons in PKM space. Note that $\hat{\gamma}$ corresponds to the hysteretic strain contribution of the cell when all hysterons are in the closed state.

In this paper we will restrict our attention to the effects of the hysteretic nonlinearity on wave propagation and the constants β and δ will be assumed to be zero.

For a given stress protocol (sequence of pressure changes) and a known PKM density distribution, it is now straightforward to predict the strain response of a cell using the infinitesimal $(\Delta\sigma,\Delta\varepsilon)$ relation:

$$\begin{aligned}\Delta\varepsilon &= \Delta\varepsilon_C + \Delta\varepsilon_H \\ &= \int_{\sigma}^{\sigma+\Delta\sigma} \frac{d\sigma'}{K_C(\sigma')} + \hat{\gamma}\,(f_c(-\sigma)-f_c(-\sigma-\Delta\sigma)).\end{aligned} \tag{4}$$

The procedure involves a careful administration of the open and closed hysterons in the PKM space and a proper numerical integration in the case of a nontrivial PKM distribution. Equation (4) leads to the following expression for the (inverse) modulus

$$K^{-1} = \lim_{\Delta\sigma\to 0}\left(\frac{\Delta\varepsilon}{\Delta\sigma}\right) = \frac{1}{K_C(\sigma)} + \hat{\gamma}\,\frac{df_c}{d\sigma}(-\sigma). \tag{5}$$

The calculation of the modulus requires the integration of the PKM density at P $(-\sigma)$ of only those hysterons that change state. For increasing pressure this corresponds to the integration at $P_c=P$ over a vertical line interval, whereas for decreasing pressure the integration occurs at $P_0=P$ over a horizontal line interval. The hysteresis in the modulus-stress relation enters through the second term in Eq. (5), which is changing discontinuously at each reversal point in stress. It is exactly this relation that is used to update the modulus in the model for the nonlinear dynamic wave propagation described in the following.

B. Structure of the finite integration scheme for nonlinear wave propagation in two-dimensional geometry

For the second upscaling task from the mesoscopic (or element level) to the macroscopic (or sample) level, we use the formalism of EFIT, which was originally developed by Fellinger *et al.*[16] This technique is based on a discretization of the incremental equation of state (Hooke's law in linear elastic systems) and the Cauchy equation of motion. In general 3D problems, the EFIT procedure uses the integral form rather than the differential form of the basic equations and performs an integration over control volumes or integration cells. The diagonal stress component integration cells coincide with the material cells defined in the multiscale concept. Velocities are calculated at the sides of the material cells, the off-diagonal stress components at the corners. A detailed description of the implementation and numerical properties of EFIT in multidimensional wave propagation problems can be found in Fellinger *et al.*[16]

The implementation of the 1D nonlinear EFIT model has been described previously in Van Den Abeele *et al.*[11] In order to extend this model to a 2D version for wave propagation, we start with the Cauchy equation of motion and the stress rate equation, which can be expressed in terms of the two particle velocity components (v_x and v_y) and the three stress components (T_{xx}, T_{yy}, and T_{xy}) in the following way:

$$\dot{v}_x = \frac{1}{\rho}\left(\frac{\partial T_{xx}}{\partial x} + \frac{\partial T_{xy}}{\partial y}\right) + F_x, \quad \dot{v}_y = \frac{1}{\rho}\left(\frac{\partial T_{xy}}{\partial x} + \frac{\partial T_{yy}}{\partial y}\right) + F_y, \tag{6a}$$

$$\dot{T}_{xx} = K_1\dot{\varepsilon}_{xx} + K_2\dot{\varepsilon}_{yy} = K_1\frac{\partial v_x}{\partial x} + K_2\frac{\partial v_y}{\partial y}, \tag{6b}$$

$$\dot{T}_{yy} = K_1\dot{\varepsilon}_{yy} + K_2\dot{\varepsilon}_{xx} = K_1\frac{\partial v_y}{\partial y} + K_2\frac{\partial v_x}{\partial x}, \tag{6c}$$

$$\dot{T}_{xy} = \mu(\dot{\varepsilon}_{xy} + \dot{\varepsilon}_{yx}) = \mu\left(\frac{\partial v_y}{\partial x} + \frac{\partial v_x}{\partial y}\right). \tag{6d}$$

Here ρ is the material density, K_1, K_2, and μ are representations for the elastic moduli (for isotropic solids $K_1=\lambda+2\mu$ and $K_2=\lambda$ with λ and μ the Lamé constants) and F_x and F_y are (optional) volume force densities defining the external action.

Following the suggested discretization of Fellinger *et al.*[16] and Schubert *et al.*[17] for the case of inhomogeneous media, we divide the solid medium in material cells of size Δx by Δy in such a way that the diagonal components of the stress (T_{xx} and T_{yy}) are calculated in the center of a cell, T_{xy} is calculated at each of the four corners, and the particle velocity components (v_x and v_y) are determined and updated at the parallel sides of the cell in x and y direction. Figure 2 illustrates a set of four neighboring fundamental material cells used for the 2D EFIT implementation. In this paper we will only consider square grids ($\Delta x=\Delta y=\Delta$). In this particular case the EFIT equations, ignoring for the moment the external forces, take on the following form:[16]

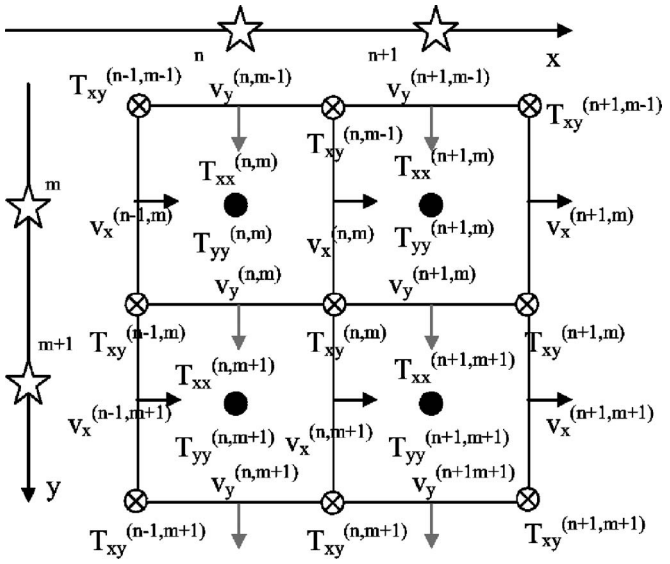


FIG. 2. A set of four elementary mesoscopic cells used in the 2D EFIT model.

$$\dot{v}_x^{(n,m)}(t) = \frac{1}{\Delta}\frac{2}{(\rho^{(n,m)}+\rho^{(n+1,m)})}(T_{xx}^{(n+1,m)}(t) - T_{xx}^{(n,m)}(t) + T_{xy}^{(n,m)}(t) - T_{xy}^{(n,m-1)}(t)), \tag{7a}$$

$$\dot{v}_y^{(n,m)}(t) = \frac{1}{\Delta}\frac{2}{(\rho^{(n,m)}+\rho^{(n,m+1)})}(T_{yy}^{(n,m+1)}(t) - T_{yy}^{(n,m)}(t) + T_{xy}^{(n,m)}(t) - T_{xy}^{(n-1,m)}(t)), \tag{7b}$$

$$\dot{T}_{xx}^{(n,m)}(t) = \frac{1}{\Delta}K_1^{(n,m)}[v_x^{(n,m)}(t) - v_x^{(n-1,m)}(t)] + \frac{1}{\Delta}K_2^{(n,m)}[v_y^{(n,m)}(t) - v_y^{(n,m-1)}(t)], \tag{7c}$$

$$\dot{T}_{yy}^{(n,m)}(t) = \frac{1}{\Delta}K_1^{(n,m)}[v_y^{(n,m)}(t) - v_y^{(n,m-1)}(t)] + \frac{1}{\Delta}K_2^{(n,m)}[v_x^{(n,m)}(t) - v_x^{(n-1,m)}(t)], \tag{7d}$$

$$\dot{T}_{xy}^{(n,m)}(t) = \frac{1}{\Delta}\frac{4}{\left(\frac{1}{\mu^{(n,m)}} + \frac{1}{\mu^{(n+1,m)}} + \frac{1}{\mu^{(n,m+1)}} + \frac{1}{\mu^{(n+1,m+1)}}\right)} \times[v_x^{(n,m+1)}(t) - v_x^{(n,m)}(t) + v_y^{(n+1,m)}(t) - v_y^{(n,m)}(t)]. \tag{7e}$$

The update of the velocities and stresses is performed using central differences for the time derivatives. This procedure results in a leap frog scheme:[11,18]

$$v_i^{(n,m)}(t) = v_i^{(n,m)}(t-\Delta t) + \Delta t\dot{v}_i^{(n,m)}\left(t-\frac{\Delta t}{2}\right) \quad \text{for} \quad i=x,y, \tag{8a}$$

$$T_{ij}^{(n,m)}\left(t+\frac{\Delta t}{2}\right)=T_{ij}^{(n,m)}\left(t-\frac{\Delta t}{2}\right)+\Delta t\dot{T}_{ij}^{(n,m)}(t) \quad \text{for} \quad i,j=x,y. \tag{8b}$$

The numerical scheme defined by Eqs. (7) and (8) is straightforward when applied to linear elastic solids. In the nonlinear case the formulation of the basic equations remains valid, but when microdamage is considered within the material cells, the moduli K_1, K_2, and μ which appear in Eq. (7) may depend on the actual values of the stresses acting on a material cell and on their time history. As a result, the modulus-stress relations cannot be expressed anymore in a closed analytical form. Moreover, it is necessary to update the values of K_1, K_2, and μ constantly at each time step because of the amplitude dependence of the material constants.

In the 2D case, we need to update three stress components at each moment in time. A first attempt to address this problem could be the extension of the scalar Preisach model to a vectorial Preisach model, as is done in the field of magnetism. Helbig showed that this methodology is unable to correctly account for the Poisson effect in solid media,[23] and therefore the vectorial approach was rejected for use in elasticity. Instead Helbig[23,24] suggested to make use of the representation of the elastic tensor by its eigensystem, consisting of eigenstiffnesses and eigenstrains.

In a 2D isotropic solid medium, the stress-strain relation can be rewritten as follows:

$$T_p=C_{pq}\varepsilon_q \quad (p,q=1,2,3), \tag{9}$$

where the elasticity tensor is given by

$$\tilde{C}_{pq}=\begin{pmatrix}\lambda+2\mu & \lambda & 0\\ \lambda & \lambda+2\mu & 0\\ 0 & 0 & 2\mu\end{pmatrix}, \tag{10}$$

and the stress and strain vectors $\boldsymbol{T}$ and $\boldsymbol{\varepsilon}$ are, respectively defined as

$$\boldsymbol{T}=(T_{xx},T_{yy},T_{xy})^T, \tag{11a}$$

$$\boldsymbol{\varepsilon}=(\boldsymbol{\varepsilon}_{xx},\boldsymbol{\varepsilon}_{yy},\boldsymbol{\varepsilon}_{xy})^T. \tag{11b}$$

In Eqs. (11a) and (11b) $(\cdots)^T$ denotes the transpose of the considered line matrix. Making use of these notations, it can be shown that the three eigenvectors of the elasticity tensor correspond to three eigenstress/eigenstrain vectors $\hat{\varepsilon}^{(k)}$.[23,24] These eigenvectors are particular strains for which the corresponding eigenstress vector is perfectly collinear with the eigenstrain vector. The mutual orthogonality of the eigenstresses implies that when the medium is stressed along each of the eigenstresses, there is no elastic energy involved along the other eigenstresses. Because of this property, it is possible to associate a scalar PKM model as described before to each of the three eigenstress directions. Each PKM space modifies the associated eigenstiffness. An extension to a 3D solid media is straightforward and yields six associated PKM spaces to the six eigenstress directions.

For an isotropic 2D solid, the eigenvalues are $2\lambda+2\mu$, 2μ, and 2μ, with the following three corresponding eigenstrains:

$$\tilde{\varepsilon}^{(1)}=\frac{1}{\sqrt{2}}\begin{pmatrix}1\\1\\0\end{pmatrix}, \quad \tilde{\varepsilon}^{(2)}=\frac{1}{\sqrt{2}}\begin{pmatrix}1\\-1\\0\end{pmatrix}, \quad \tilde{\varepsilon}^{(3)}=\begin{pmatrix}0\\0\\1\end{pmatrix}. \tag{12}$$

The projection of the general stress-strain state given by Eq. (11) along these three eigenvectors defines the volumetric, deviatoric, and shear stresses and strains:

$$T_V=\frac{T_{xx}+T_{yy}}{\sqrt{2}}, \quad T_D=\frac{T_{xx}-T_{yy}}{\sqrt{2}}, \quad T_S=T_{xy}, \tag{13a}$$

$$\varepsilon_V=\frac{\varepsilon_{xx}+\varepsilon_{yy}}{\sqrt{2}}, \quad \varepsilon_D=\frac{\varepsilon_{xx}-\varepsilon_{yy}}{\sqrt{2}}, \quad \varepsilon_S=\varepsilon_{xy}. \tag{13b}$$

In terms of these quantities, the rate equations (6b)–(6d) can be rewritten as

$$\dot{T}_V=K_V\dot{\varepsilon}_V, \quad \dot{T}_D=K_D\dot{\varepsilon}_D, \quad \dot{T}_S=K_S\dot{\varepsilon}_S, \tag{14a}$$

with

$$K_V=K_1+K_2, \quad K_D=K_1-K_2, \quad K_S=2\mu. \tag{14b}$$

The new moduli K_V, K_D, and K_S are the eigenstiffnesses corresponding to the eigenstresses defined in Eq. (12). As a first approximation, we assume that the moduli K_V, K_D, and K_S only depend on the actual value and the history of the corresponding eigenstress component which appears in the scalar equations, i.e., $K_V(T_V,T_{V,\text{his}})$, $K_D(T_D,T_{D,\text{his}})$, and $K_S(T_S,T_{S,\text{his}})$. In that case, the three rate equations can be treated as scalar equations with uncoupled PKM spaces. A similar assumption was made by Boudjema *et al.*[33] with regard to the analysis of quasistatic measurements of axial and radial strain responses to complicated mean stress and shear stress protocols. Furthermore, Helbig[23] showed that, up to a certain strain level, the eigenstress/eigenstrain system resulting from available measurement data does not substantially change, and we therefore may assume that the output strain response of each stress projection is still collinear with the initial isotropic eigenstress/eigenstrain vector $\tilde{\varepsilon}^{(k)}$. The critical strain level is generally very low and its value strongly depends on the type of material.[23] In any case, the uncoupling of the PKM spaces remains an assumption that must be confirmed by results of careful triaxial experiments.

Once the eigenstress components (T_V,T_D,T_S) have been updated, it is easy to recalculate the Cartesian stress components (T_{xx},T_{yy},T_{xy}), and one can proceed with the common leap frog scheme for the calculation of the dynamic nonlinear and hysteretic response of the material.

When dealing with problems of nonlinear wave propagation, it is impossible to express the stability condition for a finite integration technique like EFIT in a closed analytical form. If the nonlinear contribution to the modulus in Eq. (5) is not too strong however, as is the case in most applications of nonlinear acoustics in solid media, we can safely use the stability condition for *linear* EFIT as an approximation for the stability properties of the numerical nonlinear scheme.

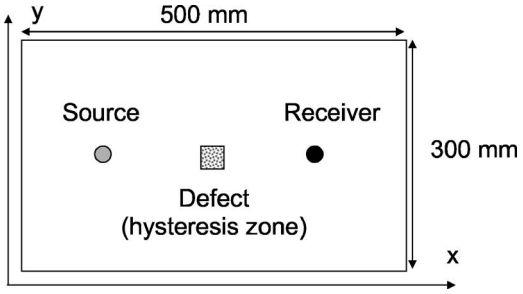


FIG. 3. Geometry of the 2D in-plane wave propagation simulations.

Moreover, hysteretic nonlinearity is known to reduce the average modulus/velocity with amplitude. For the square grids ($\Delta x=\Delta y=\Delta$) considered here, it is consequently safe to assume a time step Δt that satisfies the following inequality:

$$\Delta t \leqslant \frac{\Delta}{v_l\sqrt{2}}, \tag{15}$$

with v_l the longitudinal velocity of the material.

In the following sections, the results of two types of numerical simulations based on the nonlinear hysteretic multiscale model described earlier will be presented. The first group of simulations deals with an in-plane investigation of localized damage in a plate using pulsed ultrasound. The second problem concerns the propagation of Rayleigh waves through a medium with localized surface deterioration represented by hysteretic stress-strain relations.

III. NUMERICAL SIMULATIONS

A. 2D in-plane wave propagation

1. Configuration and simulation parameters

As a first illustration of the numerical model, we consider a steel plate with a density of 7820 kg/m^3, an overall longitudinal velocity of 6198 m/s, and a shear velocity of 3122 m/s. The dimensions of the plate are 500×300 mm (Fig. 3). We assume a zone with uniformly distributed damage over a square region of 30×30 mm, centered in the middle of the plate ($x_d=250$ mm, $y_d=150$ mm). The microdamaged zone is modeled by imposing a hysteretic modulus-stress relation with uniform (P_o, P_c) distribution of hysterons and a typical hysteretic strain parameter $\gamma=10^{16}$. With this value of γ, the total contribution of all hysterons to the strain increase when changing the pressure from -1 to 1 MPa would be of the order of 0.2 millistrain. Everywhere outside this zone, the medium is assumed to be linearly elastic. The above-mentioned choice for the PKM distribution of hysterons and the value of the hysteretic strain parameter should only be considered illustrative at this moment. The (P_o, P_c) distribution of hysterons and γ can in principle be determined by inverting experimental results of quasistatic stress-strain tests.[25,33,34] Once this information is available, it can be used as input for numerical simulations of dynamic wave propagation. Recent progress in the inversion of *quasistatic* stress-strain tests has been encouraging,[34] and we therefore expect that a comparison of *dynamic* wave propagation simulations with experiments will become feasible in the near feature, at least for simulations of bulk wave propagation and in cases where the nonlinear properties of microdamage in bulk media can be extrapolated to localized damage. As a simplification of the model, we consider for the moment only the case of volumetric stress-strain hysteresis. This means that the PKM spaces for deviatoric and shear eigenstresses are empty.

For the simulations reported in this section, no *linear* attenuation mechanism is taken into account, even though this can be readily implemented using the scheme outlined in Robertson *et al.*[35] Stress-free boundary conditions are applied at the borders of the computational domain.

The sample is excited by an explosive source which is tapered in space and time, and applied over a region of 30×30 mm, centered at $x_s=125$ mm and $y_s=150$ mm. The implemented force magnitude per unit mass as a function of space and time is given by

$$F = A\exp\left(-\frac{1}{2}\left[\frac{x-x_s}{w_x}\right]^2\right)\exp\left(-\frac{1}{2}\left[\frac{y-y_s}{w_y}\right]^2\right)\times\sin(2\pi ft)\exp\left(-\frac{1}{p}\left[\frac{t}{w}\right]^p\right)\times\left(1-\exp\left(-\left[\frac{t}{sw}\right]^2\right)\right), \tag{16}$$

in which $w_x=w_y=3$ mm, $f=250$ kHz, $p=2$, $w=2/f$, and $sw=3/f$. The horizontal and vertical components to be used in Eq. (6a) then correspond to

$$F_x(x,y) = F\frac{x-x_s}{\sqrt{(x-x_s)^2+(y-y_s)^2}},$$

$$F_y(x,y) = F\frac{y-y_s}{\sqrt{(x-x_s)^2+(y-y_s)^2}},$$

respectively.

The above-described configuration is discretized on a grid with 2500 cells in the horizontal direction and 1500 cells in the vertical direction. The sample therefore contains $2500\times1500=3.75$ million grid cells and the edge length of one grid cell is 2×10^{-4} m. This grid resolution corresponds to nearly ten grid cells per wavelength for the seventh harmonic of the shear waves. As will be discussed later on, we applied convergence testing and confirmed that a higher resolution of the grid does not result in significantly different amplitudes of the harmonics up to the fifth order. The simulations reported in this section have been performed with a time step of 2×10^{-8} s, in agreement with the constraint imposed by Eq. (15).

2. Simulation results

A typical output of the numerical model is illustrated in Fig. 4. This plot displays the T_{xx} stress component at a receiver located at $x_r=375$ mm and $y_r=150$ mm over a time scale of 250 μs for a source amplitude of $A=10^4$ N/kg in the absence of hysteresis (full line) and for a source amplitude of $A=5\times10^4$ N/kg (dotted line) in the presence of hysteresis. The received signal for $A=5\times10^4$ N/kg has been divided by a factor 5 to facilitate comparison. The influence of the hys-

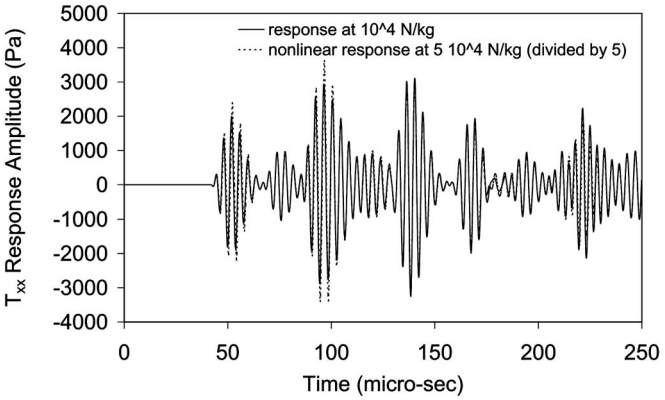


FIG. 4. Time signals of the T_{xx} stress component at the receiver for $A=10^4$ N/kg without hysteresis (full line) and for $A=5\times10^4$ N/kg in the presence of hysteresis (dotted line); The signal at $A=5\times10^4$ N/kg has been divided by a factor 5 to facilitate comparison.

teretic nonlinearity in the damage zone on the time response is very small for $A=10^4$ N/kg, and is therefore not shown separately. In addition, the difference is even hardly discernible when comparing the nonlinear responses at $A=5\times10^4$ N/kg and $A=10^4$ N/kg. The nonlinear signatures can only be disclosed easily by analyzing the signals in the frequency domain, either in their entire form, partially or by using moving windows.

In Fig. 5 we analyze the harmonic content of the responses by windowing the recorded time signals with a Gaussian tapering function with a maximum at $t=50$ μs and a width of 10 μs. This isolates the part of the signals which corresponds to the first arrival of the longitudinal wave at the receiver after traversing the defect region. The Fourier transforms for Gaussian windowed signals at $A=10^4$ and $A=5\times10^4$ N/kg are shown in Figs. 5(a) and 5(b), respectively. On each plot, the FFT of the linear simulation is shown by the solid line, while the dashed lines represent the FFT of the windowed responses from simulations in the presence of hysteresis performed with the grid resolution described earlier. The generation of harmonics of the fundamental frequency in the presence of hysteresis is clearly evident in both Figs. 5(a) and 5(b). Note also that the pure hysteretic nonlinearity assumed here creates only odd harmonics in the time signals, as expected.[6,9–11,26,27]

A convergence check of the model was performed by doubling the spatial resolution of the grid in both directions, i.e., increasing the number of cells in the computational domain by a factor 4. Accordingly, the time step was reduced by a factor 2 in order to keep satisfying the stability condition. The FFTs of the resulting time signals are indicated with circles in the plots of Figs. 5(a) and 5(b). The influence of doubling the resolution of the numerical model on the spectral content of the time signals is obviously very small. We are therefore confident that the numerical convergence with the code is achieved, even in the presence of a localized zone of hysteresis.

3. Nonlinear dependence relations

When comparing the amplitude of the third harmonic in Figs. 5(a) and 5(b), we notice an increase of only about a factor 10, although one would expect a 25-fold increase in agreement with the first-order quadratic dependence of nonclassical nonlinearity on the amplitude of the fundamental harmonic as described, for example, in Van Den Abeele *et al.*[10] and the references cited in that paper.

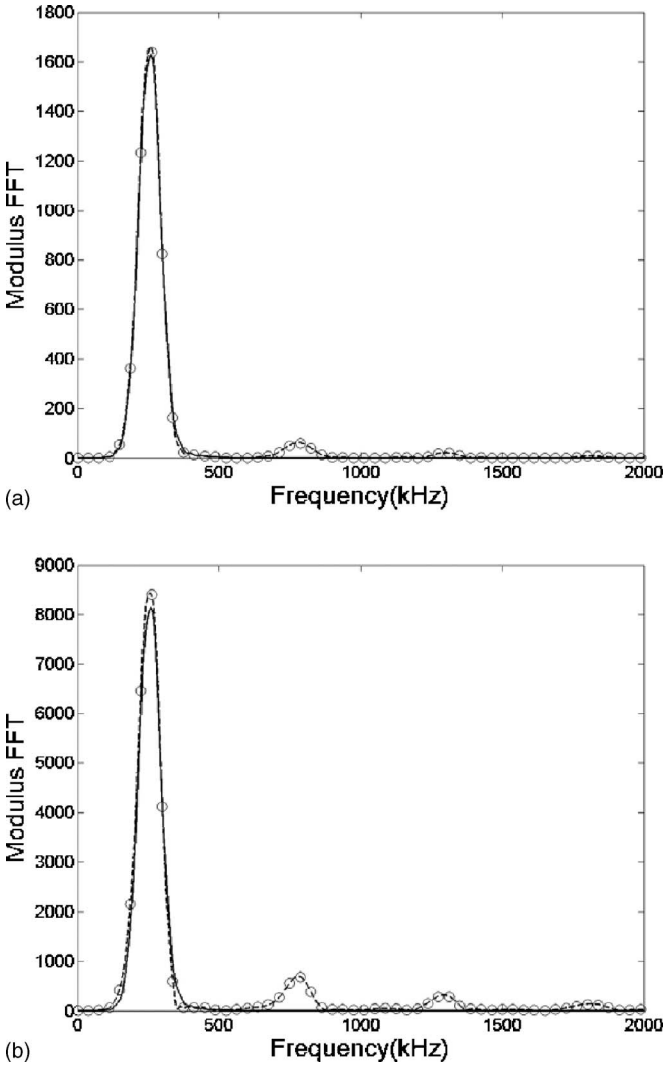


FIG. 5. (a) FFT of the T_{xx} stress signals from simulations including hysteresis at $A=10^4$ N/kg (dashed line) and from a linear simulation with $A=10^4$ N/kg (solid line). The circles indicate the FFT of the stress signal obtained from a simulation including hysteresis and with increased resolution (for details see the text). (b) Same as in plot (a), but for $A=5\times10^4$ N/kg.

To investigate this issue in more detail, we performed a number of code tests. For these tests, we considered an effective 1D wave propagation problem, in which the sample has the same dimensions and material parameters as before, but the damage zone is extended to cover the entire vertical height of the sample, as illustrated in Fig. 6. The center of the damage zone is shifted to ($x=100$ mm, $y=150$ mm). Again only volumetric hysteresis with a uniform PM space is considered for the microdamaged zone. We inject a plane longitudinal continuous 250 kHz sinusoidal wave from the left side of the sample by including a horizontal force component in the equations of motion [Eq. 6(a)]. At the upper and lower borders of the sample, plane wave boundary conditions are implemented as described by Fellinger *et al.*[16] A

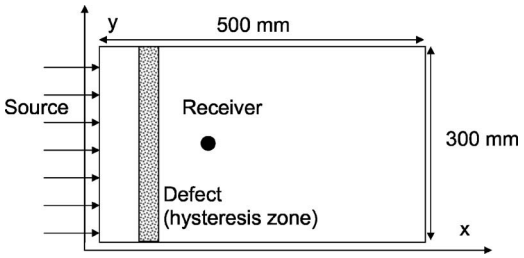


FIG. 6. Geometry of the 2D wave propagation experiment with plane waves excited at the left border of the sample.

receiver is located behind the microdamaged zone at (x = 150 mm, y = 150 mm). Figure 7 summarizes the results of this virtual experiment. The squares in Fig. 7(a) show the amplitude of the third harmonic versus the fundamental of the T_{xx} stress signals arriving at the receiver, for source amplitudes going from 200 to 2000 N/kg in steps of 200 N/kg and for a hysteresis parameter γ set to 10^{-16}. The duration of the time signals is 120 μs and the FFT of these signals is performed on the second half of this time frame. The solid line shows a quadratic fit to the numerical results. An almost perfect agreement with the numerical data is obtained. We checked that the same low amplitude behavior holds for the pulsed wave experiments illustrated in Figs. 4 and 5.

Figure 7(b) illustrates what happens when the amplitude range of the excitation is increased further from 2000 to 14 000 N/kg in steps of 2000 N/kg. The solid line shows the power law fit of the data shown in Fig. 7(a). When the source amplitude of the waves grows, the first-order quadratic dependence observed in Fig. 7(a) clearly starts to break down and turns into a quasilinear dependence instead. This quasilinear behavior occurs for the same range of wave amplitudes as illustrated in the pulse wave experiments discussed in Figs. 4 and 5. The quadratic dependence obviously does not hold for too high amplitudes.

Figures 7(c) and 7(d) show a similar behavior when the

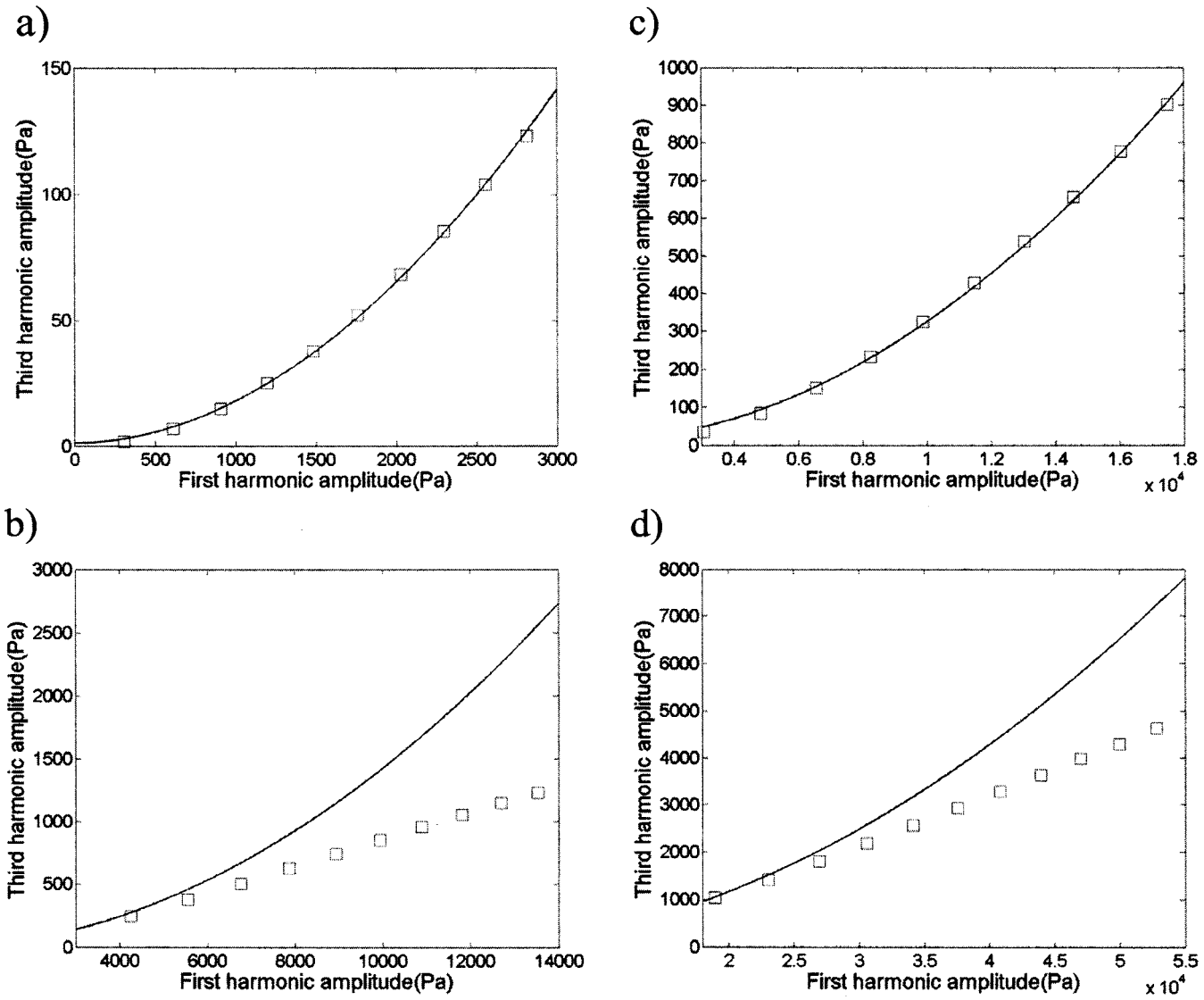


FIG. 7. Amplitude of the third harmonic of the stress signals as function of the amplitude of the fundamental for $\gamma = 10^{-16}$ and for (a) excitation amplitudes going from 200 to 2000 N/kg and (b) excitation amplitudes going from 2000 to 14 000 N/kg. The solid lines in (a) and (b) indicate a quadratic fit to the data shown in (a). Plots (c) and (d) show the same dependence for $\gamma = 2 \times 10^{-17}$ and excitation amplitudes ranging from 2000 to 14 000 N/kg and 14 000 to 40 000 N/kg, respectively. The solid lines in these plots indicate a quadratic fit to the data shown in (c).

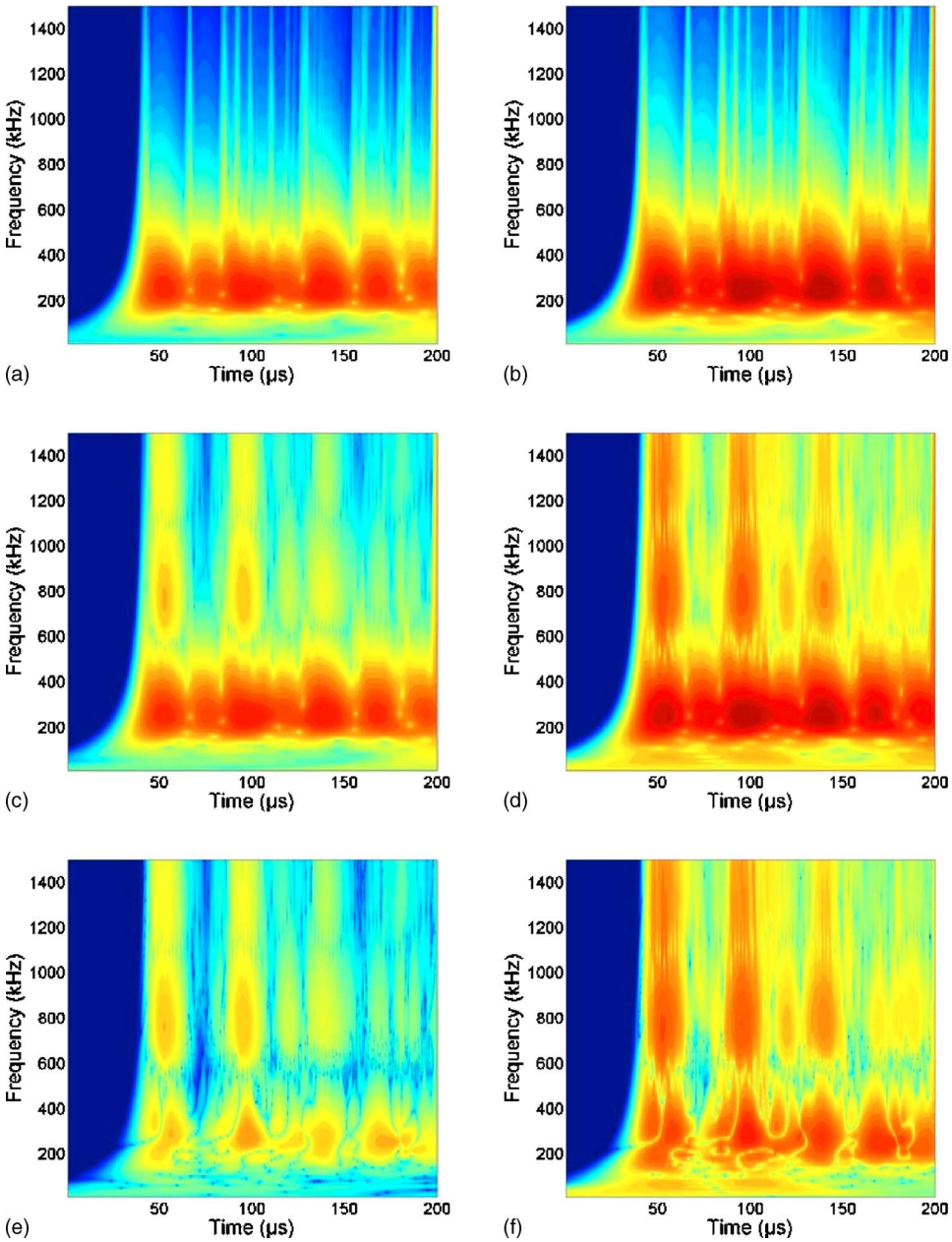


FIG. 8. (Color online) Continuous wavelet transform of the stress signals at $A=10^4$ N/kg (left) and $A=5\times10^4$ N/kg (right) for simulations without volumetric hysteresis [(a) and (b)] and including volumetric hysteresis [(c) and (d)]. (e) and (f) the absolute value of the difference between the wavelet maps at $A=10^4$ N/kg [(c)-(a)] and $A=5\times10^4$ N/kg [(d)-(b)]. All plots are on the same logarithmic scale from 0 to 9.

hysteresis parameter is lowered to 2×10^{-17}. The plots visualize the third harmonic dependence for excitation amplitudes ranging from 2000 to 14 000 N/kg in steps of 2000 N/kg and from 14 000 to 40 000 N/kg in steps of 3600 N/kg, respectively. The effect of the reduced hysteresis parameter is to postpone the transition from the quadratic to the quasilinear regime to higher amplitudes. Since the theoretical expectation of a quadratic dependence is based on first-order approximations in analytical models, one should therefore be cautious with interpreting experimental results that might fall out of its validity range.

4. Alternative analysis of the nonlinear signatures

In order to obtain a more detailed, time-resolved picture of the influence of nonlinearity on wave propagation, it is enlightening to apply wavelet analysis.[36,37] Wavelet analysis decomposes a function into a sum of shifted and scaled copies of a suitably chosen wavelet kernel. The wavelet theory covers the limitations in Fourier analysis by windowing the signal into variable sized regions, using long time intervals for information on low frequencies and shorter intervals for high frequencies. The result of applying the continuous wavelet transform on a given time signal is a wavelet map in which the amplitude of the wavelet coefficients is plotted as a function of time and the pseudofrequency of the wavelets.[36,37]

We calculated the continuous wavelet transform of the stress signals at $A=10^4$ N/kg and $A=5\times10^4$ N/kg shown in Fig. 4 using the complex Morlet wavelet.[37] The results of this calculation are illustrated in Fig. 8. Figures 8(a) and 8(b) display the wavelet maps for $A=10^4$ N/kg and $A=5\times10^4$ N/kg resulting from "linear" simulations without hysteresis in the defect region. The corresponding wavelet maps for simulations including the hysteretic stress-strain relation are shown in Figs. 8(c) and 8(d), respectively. Figures 8(e) and 8(f) show the absolute value of the difference between

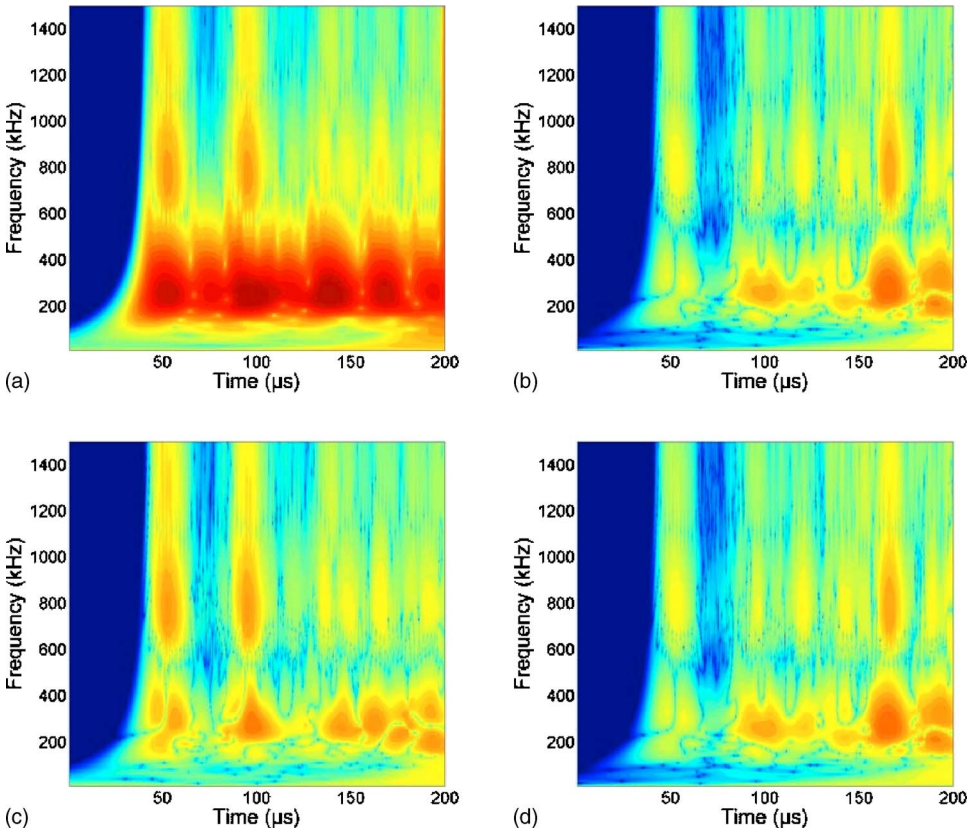


FIG. 9. (Color online) Continuous wavelet transform of the stress signals at $A=5\times10^4$ N/kg for simulations including deviatoric hysteresis (a) and shear hysteresis (b). (c), (d) The absolute value of the difference between the wavelet maps (a) and (b) and the wavelet transform of the signal without hysteresis [Fig. 5(b)]. All plots are on the same logarithmic scale from 0 to 9.

the wavelet maps for the linear and nonlinear simulations at $A=10^4$ N/kg and $A=5\times10^4$ N/kg, respectively. In order to facilitate comparison, all plots are shown on the same logarithmic scale from 0 to 9.

The objective of Fig. 8 is to illustrate the changes in the received time signals which are brought about by the introduction of hysteresis. When comparing Figs. 8(e) and 8(f), we can clearly observe the presence of the third and fifth harmonics of the fundamental at 750 and 1250 kHz in the first two to three pulse arrivals. In the second half of the time window, one observes a decrease in the amplitude of the harmonics. This reduction is a consequence of the arrival of shear waves at the receiver which are generated through mode conversion as a result of wave reflection at the boundaries of the computational domain. Shear waves are less affected by the nonlinearity because we did not include shear hysteresis in the simulations shown in Fig. 8. Apart from the generation of the harmonics, we also note a change in the amplitude at the fundamental frequency on the wavelet maps. This is consistent with the conservation of energy and the fact that energy is being transferred from the fundamental frequency to the harmonics as a result of the nonlinearity in the damage zone.

5. Influence of the eigenstress specific hysteresis

Figure 9 finally illustrates the influence of the different types of hysteresis. In Figs. 9(a) and 9(b) we show the wavelet transform of the T_{xx} signals at $A=5\times10^4$ N/kg including only deviatoric hysteresis and shear hysteresis, respectively. Figures 9(c) and 9(d) represent the absolute value of the difference with the wavelet map of the linear simulation at $A=5\times10^4$ N/kg. The results of the simulation involving only deviatoric hysteresis [Figs. 9(a) and 9(c)] are remarkably similar in structure as the ones in Figs. 8(c)–8(e) and 8(d)–8(f). We only notice a decreased level of the generated harmonics. In contrast, the simulation with only shear hysteresis [Figs. 9(a) and 9(c)] shows that the effect of the shear hysteresis is primarily observed on the fifth wave package arriving at the receiver position, while the other pulses which mainly involve longitudinal waves are considerably less affected. The above-mentioned results on the influence of the various types of hysteresis can be compared to Fig. 5 in the recent paper of Delsanto *et al.*,[22] in which the influence of the type of hysteresis on the nonlinear response of a plane compressional wave traversing a localized defect zone is investigated with the LISA model. Here the different types of hysteresis are introduced by choosing either the volumetric, deviatoric, or shear stress as the driving parameter governing the hysteretic nonlinear response of the interstice elements. Since the wave train arriving at the receiver is mainly of compressional nature (as for the first arriving pulse in our simulations), a similar behavior is indeed observed: The highest nonlinearity occurs with the volumetric stress as the driving parameter, whereas the deviatoric and shear driving parameters induce a correspondingly smaller response.

B. Interaction of Rayleigh waves with a surfacial microdamaged zone

The second example concerns the interaction of Rayleigh waves propagating in a linear medium with a small microdamaged zone with hysteretic nonlinearity located near the surface of the sample.

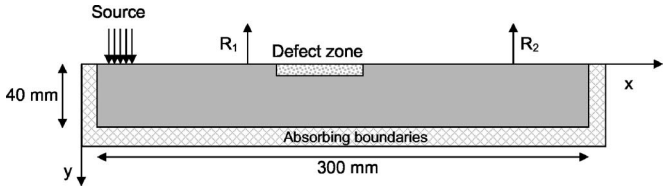


FIG. 10. Geometry of the 2D Rayleigh wave propagation simulations.

Several experimental studies on progressively damaged materials have shown that microdamage can result in a reduction of the linear moduli on the one hand, and/or an increase of the hysteretic nonlinearity (which dominates the classical nonlinearity from anharmonicity) on the other hand.[2,5,38,39] We will investigate these effects in our model simulations by considering a selection of cases with gradually varying linear material characteristics and nonlinear hysteretic parameters in the "microdamaged" zone near the surface. At this point, we must emphasize that any nonlinearity present in our simulations is assumed to be of the hysteretic type and restricted to a localized microdamaged zone. We are therefore not concerned here with the propagation of Rayleigh waves through a medium including *global classical* nonlinearity. The problem of Rayleigh wave propagation through a classically nonlinear medium, including the generation of shock waves, has been studied extensively by Hamilton and co-workers[40–42] and has been verified experimentally by, for example, Lomonosov *et al.*[43] The present approach is very similar to the recent work of Kawashima *et al.*,[44] who simulated the nonlinear acoustic response of Rayleigh waves propagating through minute surface cracks with a bilinear stress-strain relation using a finite element code.

1. Configuration and simulation parameters

The geometry of the model simulations is illustrated in Fig. 10. We consider a solid medium measuring 300 $\times 40$ mm with a density of 1000 kg/m^3, an overall longitudinal velocity of 2000 m/s, and a shear velocity of 1155 m/s. Stress-free boundary conditions apply at the top side of the medium. In order to avoid reflections from the boundaries of the numerical grid, the solid body is surrounded by three absorbing perfectly matching layers (PML)[45] (left, right, and bottom). The sample contains two receivers R_1 and R_2 located at $x=140$ and at $x=250$ mm, respectively. A surfacial zone, centered at $x=180$ mm, is considered to be microdamaged. Everywhere else the solid body (SB) is assumed linearly elastic with the following constants: $\lambda_{SB}=1.332$ GPa, $\mu_{SB}=1.334$ GPa, and $\gamma_{SB}=0$. The medium is discretized by a rectangular grid with 3000 cells in the horizontal direction and 400 cells in the vertical direction. The size of the additional PML layers is 240 cells in either direction. The time step used in the simulations is 2×10^{-8} s.

The microdamaged zone is modeled as follows. In order to avoid as much as possible the occurrence of artificial reflections from discontinuous transitions in the linear material parameters between layers, we assume generalized Gaussian distributions for the shear modulus and the hysteresis parameter, such that

$$\mu(x,y) = \mu_{SB}\left(1 - \mu_R \exp\left(-\frac{1}{p}\left[\frac{x-x_s}{w_x}\right]^p - \frac{1}{2}\left[\frac{y-y_s}{w_y}\right]^2\right)\right), \tag{17a}$$

$$\gamma(x,y) = \gamma_{max} \exp\left(-\frac{1}{p}\left[\frac{x-x_s}{w_x}\right]^p - \frac{1}{2}\left[\frac{y-y_s}{w_y}\right]^2\right), \tag{17b}$$

centered in the x direction at $x_s=180$ mm with width equal to $w_x=12$ mm and $p=8$, and centered in the y direction at the surface ($y_s=0$) with width equal to w_y. The maximum reduction of the modulus in the damaged zone (at the surface) is 20% ($\mu_R=0.2$). The maximum strength of the hysteretic nonlinearity (also at the surface) is set to $\gamma_{max}=2\times10^{-17}$. As in the previous simulations, the stress-strain relation of the material cells in the damaged zone is updated using a uniform PKM distribution and only volumetric hysteresis is considered.

When exciting Rayleigh waves in model simulations, it is beneficial to avoid the generation of too many additional waves that may complicate the picture. Furthermore, the waves must be sampled at a sufficient distance from the source to avoid near field effects. We have implemented a source that maximizes the energy transferred into the Rayleigh waves by the following procedure. We consider a source region that extends over a distance of 40 mm from $x_l=2.5$ to $x_r=42.5$ mm, and include an excitation signal, which favors the generation of Rayleigh waves in the sample, as follows:

$$F_y = A \sin\left(2\pi f\left[t - \left(\frac{x-x_l}{c_R}\right)\right]\right)\exp\left(-\frac{1}{p_t}\left[\frac{t-t_c}{w_t}\right]^{p_t}\right) \times\left(1 - \exp\left(-\left[\frac{t}{sw}\right]^2\right)\right), \tag{18}$$

in which c_R denotes the (linear) Rayleigh wave velocity ($c_R=1067$ m/s for the chosen material parameters). This expression is introduced as a force term for the v_y component in Eq. (6a) at all points on the free surface within the excitation region. No source term is considered in the horizontal position. By changing the values of (t_c, p_t, w_t, sw) one can easily describe any type of signal between a pulsed or continuous signal. The phase dependence of the sinusoidal term in Eq. (18) ensures optimal excitation of Rayleigh waves by keeping the time delay of the excitation in phase with the propagation of the Rayleigh wave. Figure 11 verifies that the wave fields produced in this way indeed correspond to genuine Rayleigh waves. For a fully intact sample without nonlinearity or a modulus gradient, Fig. 11 illustrates vertical slices of the x and y components of the particle velocity at $x=18$ cm for two source frequencies of 100 and 300 kHz, respectively. The two frequencies shown are the limits of the range of frequencies that will be used later on in the model simulations. We clearly observe that the simulated wave field has indeed developed into Rayleigh waves without adverse interference effects.

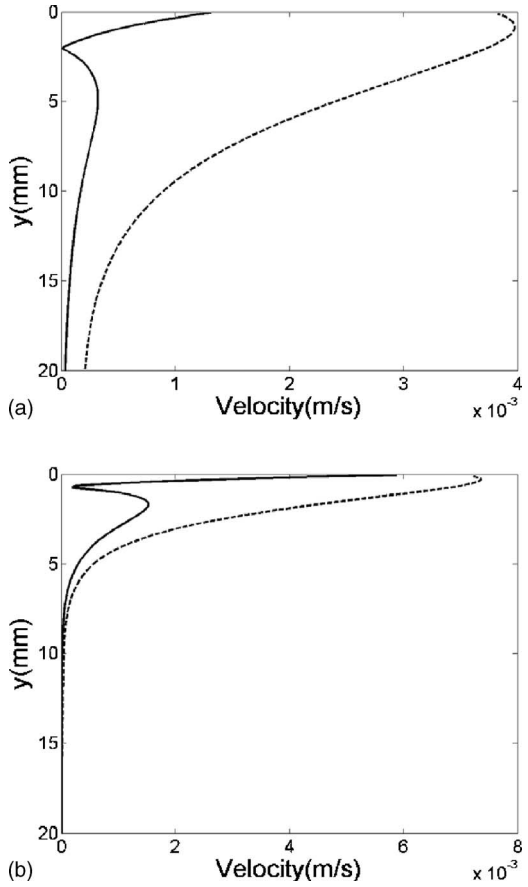


FIG. 11. Vertical slice of the x (solid line) and y (dashed line) components of the material velocity at $x=180$ mm for source frequencies of 100 kHz (a) and 300 kHz (b), respectively.

2. Simulation results: Effect of hysteresis on the Rayleigh wave velocity

In our analysis of the effects of damage on the Rayleigh wave propagation, we first focus on the effect of hysteresis and/or modulus reduction on the phase velocity of the Rayleigh waves along the surface. In the simulations, we generated Rayleigh waves at different frequencies (f) by considering pulsed excitation signals with the following choice for the parameters in Eq. (18): $t_c=0.0$, $p_t=2$, $w_t=\sqrt{2}/f$, $sw=3/f$, and $A=5\times10^4$ N/kg. We then computed the time of flight between receivers R_1 and R_2 by correlating the received time signals at these two positions. Finally, a "mean" or "spatially averaged" Rayleigh wave velocity was obtained from the time of flight and the known distance between receivers R_1 and R_2. This procedure was repeated for several cases with and without modulus reduction and with and without hysteresis. From the width of the correlation peaks, we generally estimate that the wave speeds obtained in this way are accurate to within 1 to 2 m/s. We calculated the frequency dependence of the Rayleigh wave speed in the range 100–300 kHz, assuming that $w_y=2.5$ mm for both the shear modulus and the hysteresis distributions in Eq. (17).

In Fig. 12, we compare the frequency dependence of the wave speed for simulations with and without hysteresis, and for three values of the modulus gradient parameter μ_R [Eq. (17)]: $\mu_R=0$, $\mu_R=10\%$ and 20%. In NDT and seismology,[46] it is well known that the wave speed of Rayleigh surface

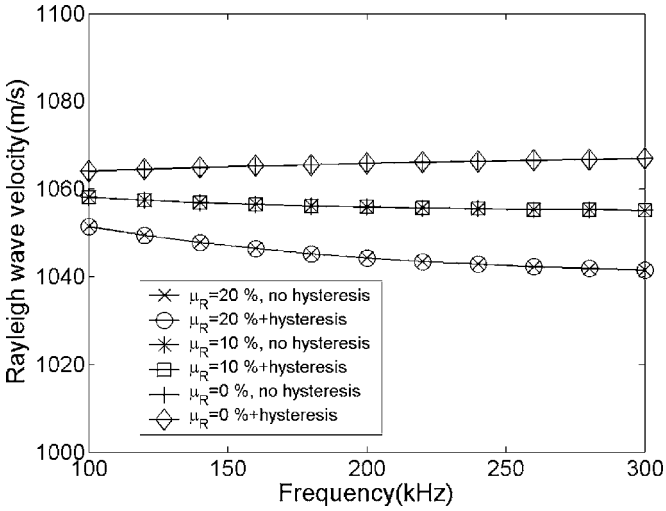


FIG. 12. Comparison between the frequency dependence of the Rayleigh wave velocity in the range 100–300 kHz with and without hysteresis for simulations with $\mu_R=20\%$, $\mu_R=10\%$, and without modulus reduction. The vertical width of the damaged zone is $w_y=2.5$ mm for both the hysteresis and shear modulus distributions.

waves is highly dependent on the stiffness of the near-surface layers. In particular, the presence of a modulus gradient introduces dispersion of the Rayleigh wave speeds. Looking at the curves without the introduction of hysteresis in Fig. 12, we indeed observe a slower mean Rayleigh wave speed when the maximum shear modulus reduction increases from 0% to 10% and finally to 20%. This effect can be readily attributed to the reduction of the shear wave speed in the region containing the defect zone. The results also show the expected dispersion of the Rayleigh wave speeds with frequency, since the wave speeds at 300 kHz show a larger spreading than the wave speeds at, for example, 100 kHz. We conclude that these results can be well understood in the context of linear depth profiling techniques.

In addition to the expected effect of a modulus gradient, Fig. 12 also shows the dispersion curves resulting from simulations including modulus reduction *and* layered hysteresis. To within the error of the wave speed determination in our calculations, we observe that the presence of the microscopic hysteretic nonlinearity in the microdamaged zone has little or no "measurable" *extra* effect on the wave speed of the propagating Rayleigh waves. This latter observation can also be readily explained by recalling that the typical values of hysteretic nonlinearity considered here produce a second-order effect on the modulus reduction which is typically below 0.1%.[1–5,38,39,47] Although we are currently not aware of any experiments that could confirm these findings, we may conjecture that the detection of hysteretic nonlinearity induced by microdamage will not be feasible by means of depth profiling techniques based on measuring the speed of Rayleigh waves. We will therefore focus our attention on a potential *nonlinearity based* depth profiling technique, in which the frequency dependence of the harmonic content of the signals is used instead.

3. Simulation results: Effect of hysteresis on the harmonic content of Rayleigh waves

In the second set of simulations, we study the harmonic content of the Rayleigh waves by analyzing the level of the

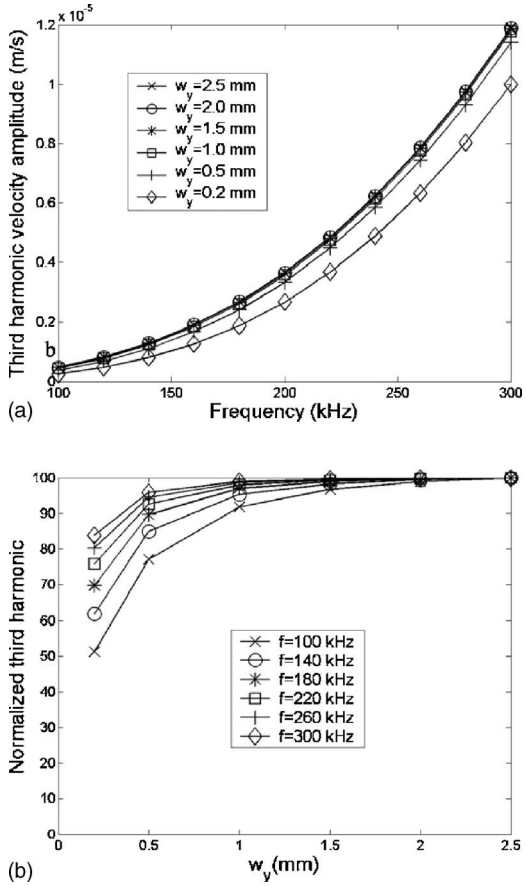


FIG. 13. (a) Frequency dependence in the range 100–300 kHz of the third-harmonic particle velocity component $v_{y,3f}$ in the Rayleigh wave propagation for different values of the depth w_y (w_y=0.2, 0.5, 1.0, 1.5, 2.0, and 2.5 mm) in the absence of modulus reduction; (b) Normalized dependence of the third-harmonic particle velocity component $v_{y,3f}$ in the Rayleigh wave propagation on w_y for frequencies in the range 100–300 kHz.

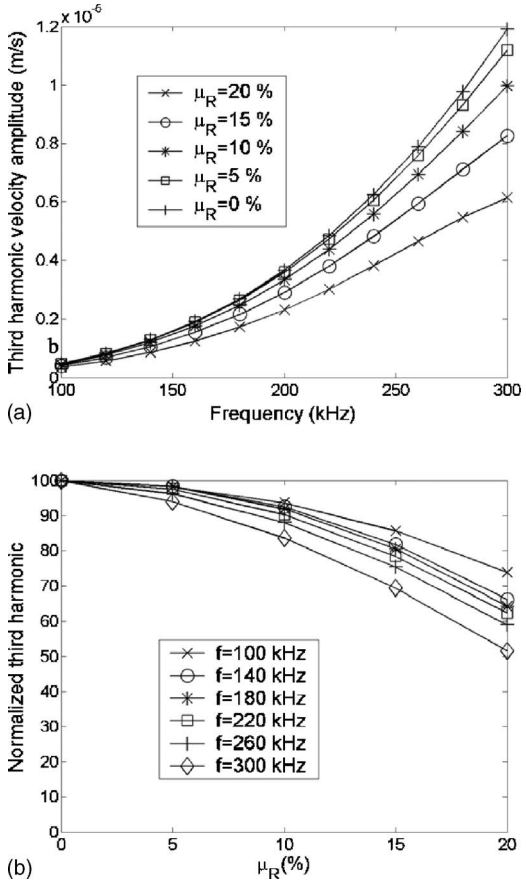


FIG. 14. (a) Frequency dependence in the range 100–300 kHz of the third-harmonic particle velocity component $v_{y,3f}$ in the Rayleigh wave propagation for different levels of the shear modulus reduction (no modulus reduction, μ_R=5%, 10%, 15%, and 20%). The vertical width of the microdamaged zone is w_y=2.5 mm for both the hysteresis and the shear modulus distributions; (b) Normalized dependence of the third-harmonic particle velocity component $v_{y,3f}$ in the Rayleigh wave propagation on μ_R for frequencies in the range 100–300 kHz.

third harmonic spectral component of the Rayleigh waves' out of plane particle velocity v_y at the position of the second receiver. For the simulations reported in this section, we used continuous excitation signals with t_c=0.0, p_t=1, w_t=5.0, $sw=10^{-5}$, and again $A=5\times10^4$ N/kg. With regard to the frequency dependence of the harmonics, it is well known that in a 1D bar of (nonclassical) hysteretic material, the leading order analytical approximations predict a dependence of the third-harmonic displacement amplitude u_{3f} which is generally quadratic in the driving frequency.[10] Taking into account the extra frequency dependence for velocity components ($v=2\pi fu$), we may expect a cubic frequency dependence for the third-harmonic particle velocity $v_{y,3f}$ at a fixed position and for a fixed excitation level.

First, we consider a microdamaged zone without modulus reduction, but with a layering of hysteresis decreasing from the surface on downward in the material. In Fig. 13, we compare the simulation results of $v_{y,3f}$ at receiver position R_2 for different depths of the microdamaged region by assuming different values for the width w_y (w_y=0.2, 0.5, 1.0, 1.5, 2.0, and 2.5 mm). Figure 13(a) shows the frequency dependence of the third-harmonic velocity amplitudes for the range of w_y values mentioned earlier. The results first of all show that we indeed recover the predicted cubic relationship with frequency. As can be expected, we observe that the third-harmonic component diminishes with decreasing depth of the microdamaged zone because the total amount of nonlinearity in the system decreases. The reduction of the strength of the harmonics as function of the width is, however, strongly frequency dependent. This becomes evident by looking at Fig. 13(b), where the harmonic component is plotted versus the width of the microdamaged zone for different frequencies after normalization with respect to the value at w_y=2.5 mm for each frequency. Compared to the value at w_y=2.5 mm, the third harmonic for w_y=0.2 mm reduces by about 50% at f=100 kHz, but by less than 20% at f=300 kHz. The amplitude of the third harmonic of the vertical component of the Rayleigh wave velocity field is therefore highly sensitive to the microscopic nonlinearity of the near-surface layers and provides a means to measure the depth of the microdamage zone, especially in those cases where the early stage microdamage does not engender a reduction of the linear elasticity characteristics.

Let us now consider the combined effect of a modulus reduction on top of the hysteretic nonlinearity. Figure 14 illustrates the frequency dependence of $v_{y,3f}$ at R_2 for a zone with layered hysteretic nonlinearity and different levels of modulus reduction. The width of the microdamaged zone is

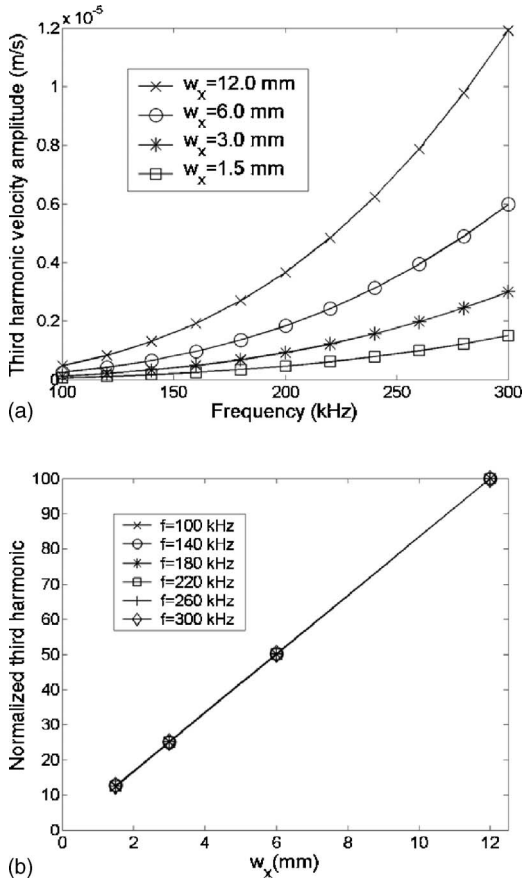


FIG. 15. (a) Frequency dependence in the range 100–300 kHz of the third-harmonic particle velocity component $v_{y,3f}$ in the Rayleigh wave propagation for different values of the horizontal extent w_x (w_x=1.5, 3.0, 6.0, and 12 mm) without modulus reduction. The vertical width of the microdamaged zone is w_y=2.5 mm; (b) normalized dependence of the third-harmonic particle velocity component $v_{y,3f}$ in the Rayleigh wave propagation on w_x for frequencies in the range 100–300 kHz.

w_y.=2.5 mm for both the modulus and the hysteresis gradients. As the strength of the harmonic generation in the presence of (any type of) nonlinearity depends on the stiffness of the material, a reduction of the modulus in the microdamaged zone leads to a reduction in the observation of the third harmonic of the particle velocity v_y as well. This can be readily seen in Fig. 14(a). In addition, we notice that the cubic frequency dependence is no longer valid for large modulus reductions. Figure 14(b) shows the third-harmonic velocity amplitudes as a function of the maximum modulus reduction μ_R for different frequencies, after normalization with respect to the value of $v_{y,3f}$ for μ_R=0 at each frequency. The spreading of the harmonic amplitudes increases at higher values of μ_R. This effect is quite analogous to the increasing dispersion of the Rayleigh wave velocities for increasing values of μ_R (cf. Fig. 12).

Finally, we illustrate in Fig. 15 the influence of the horizontal width of the microdamaged zone on the amplitude of the harmonics, and on their frequency dependence. In these simulations, a layered hysteresis is considered with the value of w_y fixed at 2.5 mm, and no modulus gradient is assumed. As can be expected, we observe a strong reduction of the strength of the harmonic generation with decreasing horizontal width [Fig. 15(a)]. When the data are normalized with respect to the value at w_y=2.5 mm for each frequency, we obtain Fig. 15(b), in which no dispersive behavior is observed, and the (expected) linear accumulation of harmonic energy with distance is put into evidence.

The numerical results obtained in this section suggest that there is a strong potential for the development of new depth profiling techniques based on the nonlinear response of Rayleigh waves to microdamage located near the surface of a sample. Apart from the urgent need of experimental confirmation, future theoretical studies should provide a more in-depth look at the possibility of solving inverse problems based on the dispersion curves of the harmonics and should include the effect of global and local classical nonlinearity on surface wave propagation in the models as well. This is part of our ongoing research.

IV. CONCLUSIONS

We have illustrated a multiscale approach to model wave propagation in materials containing hysteretic nonlinearity. At the microlevel, we have simulated microcracks by trigger like elements with a two state stress-strain relation (hysterons). On the mesoscopic level, a scalar Preisach model has been used to compute the stress-strain relations of a representative volume in terms of its eigenstresses and eigenstrains. Finally, a staggered grid formulation based on the EFIT method has been implemented to predict the macroscopic response to an arbitrary excitation signal.

Using the multiscale model, we investigated the influence of a microdamaged zone on the propagation characteristics of bulk waves and Rayleigh waves in 2D wave propagation.

From simulations of the in-plane propagation of pulses in plates, we conclude that the nonlinear effects only induce appreciable changes in the time signals at high amplitudes, but although two signals may not differ visually very much, comparison of time signals in terms of spectral content or wavelet analysis is generally very clarifying.

Numerical simulations of the interaction of a Rayleigh wave with a microcracked zone located close to the surface of a solid have shown that the effect of the nonlinear response of the Rayleigh waves is not particularly critical on the wave velocities. On the other hand, our results encourage the use of new depth profiling methods based on the analysis of the generated harmonic content as a function of frequency.

ACKNOWLEDGMENTS

This research was supported by the Foundation for Scientific Research, Flanders, Belgium (Grant Nos. G.0206.02, G.0257.02, G.0554.06, and G.0433.07), and by the European Union (STREP AST-CT-2003-502927 AERONEWS). The authors acknowledge the support of the High Performance Computing Center (VIC) at the K.U.Leuven.

[1]V. E. Nazarov, L. A. Ostrovskii, I. A. Soustova, and A. M. Sutin, "Anomalous acoustic nonlinearity in metals," Sov. Phys. Acoust. **34**, 284–289 (1988).

[2]P. B. Nagy and L. Adler, "Acoustic nonlinearity in plastics," in *Review of Progress in Quantitative Nondestructive Evaluation*, edited by D. O. Thompson and D. E. Chimenti (Plenum, New York, 1992), Vol. **I IB**, pp.

2025–2032 (1992).
[3]P. A. Johnson, B. Zinszner, and P. N. J. Rasolofosaon, "Resonance and elastic nonlinear phenomena in rock," J. Geophys. Res. **101**, 11553–11564 (1996).
[4]K. Van Den Abeele, P. A. Johnson, and A. M. Sutin, "Nonlinear elastic wave spectroscopy (NEWS) techniques to discern material damage. I. Nonlinear wave modulation spectroscopy (NWMS)," Res. Nondestruct. Eval. **12**, 17–30 (2000).
[5]K. Van Den Abeele, J. Carmeliet, J. A. TenCate, and P. A. Johnson, "Nonlinear elastic wave spectroscopy (NEWS) techniques to discern material damage. II. Single mode nonlinear resonant acoustic spectroscopy (SIMONRAS)," Res. Nondestruct. Eval. **12**, 31–42 (2000).
[6]V. E. Nazarov, A. V. Radostin, L. A. Ostrovskii, and I. A. Soustova, "Wave processes in media with hysteretic nonlinearity. I," Acoust. Phys. **49**, 344–353 (2003).
[7]P. P. Delsanto and M. Scalerandi, "Modeling nonclassical nonlinearity, conditioning, and slow dynamics effects in mesoscopic elastic materials," Phys. Rev. B **68**, 064107 (2003).
[8]M. Scalerandi, P. P. Delsanto, V. Agostini, K. Van Den Abeele, and P. A. Johnson, "Local interaction simulation approach to modelling nonclassical, nonlinear elastic behavior in solids," J. Acoust. Soc. Am. **113**, 3049–3059 (2003).
[9]V. Gusev and V. A. Aleshin, "Strain wave evolution equation for nonlinear propagation in materials with mesoscopic mechanical elements," J. Acoust. Soc. Am. **112**, 2666–2679 (2002).
[10]K. Van Den Abeele, P. A. Johnson, R. A. Guyer, and K. R. McCall, "On the quasi-analytic treatment of hysteretic nonlinear response in elastic wave propagation," J. Acoust. Soc. Am. **101**, 1885–1898 (1997).
[11]K. Van Den Abeele, F. Schubert, V. Aleshin, F. Windels, and J. Carmeliet, "Resonant bar simulations in media with localized damage," Ultrasonics **42**, 1017–1024 (2004).
[12]M. Hirsekorn, A. S. Gliozzi, M. Nobili, and K. Van Den Abeele, "A 2D spring model for the simulation of nonlinear hysteretic elasticity," in *The Universality of Nonclassical Nonlinearity with Applications to NDE and Ultrasonics*, edited by P. P. Delsanto (Springer, 2006), Chap. 18, pp. 287–308.
[13]F. Preisach, "Über die magnetische Nachwirkung," ("About the magnetic after-effect") Z. Phys. **94**, 277–302 (1935).
[14]M. A. Krasnosel'skii and A. V. Pokrovskii, *Systems with Hysteresis* (Springer, Berlin, 1989).
[15]I. D. Mayergoyz, "Hysteresis models from the mathematical and control-theory points of view," J. Appl. Phys. **57**, 3803–3805 (1985).
[16]P. Fellinger, R. Marklein, K. J. Langenberg, and S. Klaholz, "Numerical modeling of elastic-wave propagation and scattering with EFIT - Elastodynamic finite integration technique," Wave Motion **21**, 47–66 (1995).
[17]F. Schubert and B. Koehler, "Numerical modeling of elastic wave propagation in random particulate composites," in *Nondestructive Characterization of Materials VIII* (Plenum, New York, 1998), pp. 567–574.
[18]F. Schubert, B. Koehler, and A. Peiffer, "Time domain modeling of axisymmetric wave propagation in isotropic elastic media with CEFIT - cylindrical elastodynamic finite integration technique," J. Comput. Acoust. **9**, 1127–1146 (2001).
[19]F. Schubert and B. Koehler, "Three-dimensional time domain modeling of ultrasonic wave propagation in concrete in explicit consideration of aggregates and porosity," J. Comput. Acoust. **9**, 1543–1560 (2001).
[20]O. Bou Matar, S. Dos Santos, S. Callé, T. Goursolle, S. Vanaverbeke, and K. Van Den Abeele "*Simulations of nonlinearity based time reversal imaging of damaged materials*," on the CD-Rom: 9th European Conference on NDT, Berlin September 25–29, 2006 (ISBN 3–931381–86–2, Deutsche Gesellschaft für Zerstörungsfreie Prüfung e.V., Germany, Proceedings BB 103–CD).
[21]A. S. Gliozzi, M. Nobili, and M. Scalerandi, "Modelling localized nonlinear damage and analysis of its influence on resonance frequencies," J. Phys. D **39**, 3895–3903 (2006).
[22]P. P. Delsanto, A. S. Gliozzi, M. Hirsekorn, and M. Nobili, "A 2D spring model for the simulation of ultrasonic wave propagation in nonlinear hysteretic media," Ultrasonics **44**, 279–286 (2006).
[23]K. Helbig, "A formalism for the consistent description of non-linear elasticity of anisotropic media," in *Proceedings of the Eighth International Workshop on Seismic Anisotropy*, edited by P. N. J. Rasolofosaon, Rev. de l'IFP., 1998, **53**, pp. 693–708.
[24]K. Helbig and P. N. J. Rasolofosaon, "A theoretical paradigm for describing hysteresis and nonlinear elasticity in arbitrary anisotropic rocks," in *Anisotropy 2000: Fractures, Converted Waves and Case Studies* (Society of Exploration Geophysics, Tulsa, OK, 2001), pp. 383–398.
[25]R. A. Guyer, K. R. McCall, G. N. Boitnott, L. B. Hilbert, Jr., and T. J. Plona, "Quantitative implementation of Preisach-Mayergoyz space to find static and dynamic elastic moduli in rock," J. Geophys. Res., [Solid Earth] **102**, 5281–5293 (1997).
[26]K. R. McCall and R. A. Guyer, "Equation of state and wave-propagation in hysteretic nonlinear elastic-materials," J. Geophys. Res., [Solid Earth] **99**, 23887–23897 (1994).
[27]R. A. Guyer, K. R. McCall, and G. N. Boitnott, "Hysteresis, discrete memory, and nonlinear-wave propagation in rock - A new paradigm," Phys. Rev. Lett. **74**, 3491–3494 (1995).
[28]L. D. Landau and E. M. Lifshitz, *Theory of Elasticity* (Pergamon, London, 1959).
[29]C. Pecorari, "Adhesion and nonlinear scattering by rough surfaces in contact: Beyond the phenomenology of the Preisach-Mayergoyz framework," J. Acoust. Soc. Am. **116**, 1938–1947 (2004).
[30]V. A. Aleshin and K. Van Den Abeele, "Micro-potential model for stress-strain hysteresis of micro-cracked materials," J. Mech. Phys. Solids **53**, 795–824 (2005).
[31]V. A. Aleshin and K. Van Den Abeele, "Microcontact based theory for acoustics in microdamaged materials," J. Mech. Phys. Solids **55**, 366–390 (2007).
[32]V. A. Aleshin and K. Van Den Abeele, "Friction in unconforming grain contacts as a mechanism for tensorial stress-strain hysteresis," J. Mech. Phys. Solids **55**, 765–787 (2007).
[33]M. Boudjema, I. B. Santos, K. R. McCall, R. A. Guyer, and G. N. Boitnott, "Linear and nonlinear modulus surfaces in stress space, from stress-strain measurements on Berea sandstone," Nonlinear Processes Geophys. **10**, 589–597 (2003).
[34]V. Aleshin, W. Desadeleer, M. Wevers, and K. Van Den Abeele, "Characterization of hysteretic stress-strain behavior using the integrated Preisach density," accepted for publication in J. Nonlinear Mechanics (2007).
[35]J. O. A. Robertsson, J. O. Blanch, and W. W. Symes, "Viscoelastic finite-difference modeling," Geophysics **59**, 1444–1456 (1994).
[36]M. Siqueira, C. Gatts, R. R. da Silva, and J. Rebello, "The use of ultrasonic guided waves and wavelets analysis in pipe inspection," Ultrasonics **41**, 785–797 (2004).
[37]M. Misiti, Y. Misiti, G. Oppenheim, and J. M. Poggi, *Wavelet Toolbox for use with Matlab* (The MathWorks Inc., Natick, MA, 1996).
[38]K. Van Den Abeele and J. De Visscher, "Damage assessment in reinforced concrete using spectral and temporal nonlinear vibration techniques," Cem. Concr. Res. **30**, 1453–1464 (2000).
[39]K. Van Den Abeele, K. Van De Velde, and J. Carmeliet, "Inferring the degradation of pultruded composites from dynamic nonlinear resonance measurements," Polym. Compos. **22**, 555–567 (2001).
[40]M. F. Hamilton, Yu. A. Il'insky, and E. A. Zabolotskaya, "Evolution equations for nonlinear Rayleigh waves," J. Acoust. Soc. Am. **97**, 891–897 (1995).
[41]E. Yu. Knight, M. F. Hamilton, Yu. A. Il'insky, and E. A. Zabolotskaya, "General theory for the spectral evolution of nonlinear Rayleigh waves," J. Acoust. Soc. Am. **102**, 1402–1417 (1997).
[42]D. J. Shull, M. F. Hamilton, Yu. A. Il'insky, and E. A. Zabolotskaya, "Harmonic generation in plane and cylindrical nonlinear Rayleigh waves," J. Acoust. Soc. Am. **94**, 418–427 (1993).
[43]A. Lomonosov, V. G. Mikhalevich, P. Hess, E. Yu. Knight, M. F. Hamilton, and E. A. Zabolotskaya, "Laser-generated nonlinear Rayleigh waves with shocks," J. Acoust. Soc. Am. **105**, 2093–2096 (1999).
[44]K. Kawashima, R. Omote, I. Ito, H. Fujita, and T. Shima, "Nonlinear acoustic response through minute surface cracks: FEM simulation and experimentation," Ultrasonics **40**, 611–615 (2002).
[45]F. Collino and C. Tsogka, "Application of the perfectly matched absorbing layer model to the linear elastodynamic problem in anisotropic heterogeneous media," Geophysics **66**, 294–307 (2001).
[46]K. Aki and P. G. Richards, *Quantitative Seismology* (University Science Books, Sausolito, CA, 2002).
[47]P. A. Johnson and A. M. Sutin, "Slow dynamics and anomalous nonlinear fast dynamics in diverse solids," J. Acoust. Soc. Am. **117**, 124–130 (2005).

Multi-mode nonlinear resonance ultrasound spectroscopy for defect imaging: An analytical approach for the one-dimensional case

Koen Van Den Abeele[a)]
Interdisciplinary Research Center, Katholieke Universiteit Leuven Campus Kortrijk, E. Sabbelaan 53, +B-8500 Kortrijk, Belgium

(Received 19 July 2006; revised 28 March 2007; accepted 8 April 2007)

A nonlinear version of the resonance ultrasound spectroscopy (RUS) theory is presented as an extension of the RUS formalism to the treatment of microdamage characterized by nonlinear constitutive equations. General analytical equations are derived for the one-dimensional case, describing the excitation amplitude dependent shift in the resonance frequency and the generation of harmonics resulting from the interaction between bar modes due to the presence of either localized or volumetrically distributed nonlinearity. Solutions are obtained for classical cubic nonlinearity, as well as for the more interesting case of hysteresis nonlinearity. The analytical results are in excellent quantitative agreement with numerical calculations from a multiscale model. Finally, the analytical formulas are exploited to infer critical information about damage position, degree of nonlinearity, and width of the damage zone either from the shifts in resonance frequency occurring at different excitation modes, or from the shift and the harmonics predicted at a single mode. Unlike other techniques, the multi-mode-nonlinear RUS method does not require a spatial scan to locate the defect, as it lets different excitation modes, with different vibration patterns, probe the structure. Two general methods are suggested for inverting experimental data. © *2007 Acoustical Society of America.* [DOI: 10.1121/1.2735807]

PACS number(s): 43.25.Gf, 43.25.Dc, 43.25.Ba [MFH] Pages: 73–90

I. INTRODUCTION

"Can we 'hear' the shape of a drum?" is one of the most famous questions in physics that was posed by the mathematician Mark Kac (1966). The answer is yes. If perfect, the drum will have a family of vibration modes, also known as resonances, and, upon analyzing these modes, we are able to describe the physical parameters of the drum.

Now, suppose we make a hole in the drumhead using a hole punch. "Can we then 'hear' the location of the hole?" The answer is again yes. The hole is a localized perturbation in the drumhead and causes a shift in some of the resonance frequencies depending on the corresponding motion of the drumhead at the location of the hole. If the hole is located in a place where a particular mode has a node, that mode will not show a frequency shift. If it is situated in a place at an antinode, that mode will show a frequency shift relative to the perfect drumhead. The information about the location of the hole is thus encoded in the frequency shifts of the spectral modes that are affected by the local perturbation. A non-exhaustive overview of the applications of this concept to nondestructive testing can be found in Migliori and Sarrao (1997). In addition to resonance measurements, several tomographic inversion routines have been proposed to invert the location of the hole based on the information contained in (linear) wave propagation (e.g., Leonard *et al.*, 2002; Michaels and Michaels, 2006).

Suppose we replace the hole in the drumhead with a very small defect like a tear or a crack, which is only visible with the aid of a magnifying glass. "Can we hear the location of the microdefect?" This is exactly the question that we would like to answer in this paper, at least for the simple case of a localized defect in a one-dimensional (1D) bar. The methodology to do this is based on an extension of the formalism of (linear) resonance ultrasound spectroscopy (RUS) to nonlinear resonance ultrasound spectroscopy (NRUS). Even though we limit the treatment to longitudinal waves in a bar, it can be straightforwardly extended to flexural modes, and possibly to higher dimensions. A related technique based on nonlinear wave propagation has been proposed by Kazakov *et al.* (2002) evaluating the modulation of pulsed waves upon cyclic loading of a sample.

Resonance ultrasound spectroscopy (RUS) (Visscher *et al.*, 1991; Maynard, 1996; Migliori and Sarrao, 1997) is a bench-top measurement technique used to determine the full linear elastic tensor of a sample from the combined information contained in its resonance frequencies, its geometry, and its density. RUS is very accurate when it is applied to samples having a well defined geometry and homogeneously distributed linear elastic constants. Nevertheless, it has been successfully applied to determine the elastic constants of anisotropic media, to study thermoelectric materials, rocks, etc. (Keppens *et al.*, 1998; Ulrich *et al.*, 2002; Ogi *et al.*, 2002; Nakamura *et al.*, 2004; Ichitsubo *et al.*, 2002, Teklu *et al.*, 2004).

Microcracks caused by incipient damage will affect the resonance spectrum of a sample only very slightly. In many

[a)]Electronic mail: koen.vandenabeele@kuleuven-kortrijk.be

cases the effect is masked by the resolution of the frequency spectrum. Therefore, a standard RUS analysis may not be able to determine the presence of damage at early stages. However, this does not mean that we are not able to hear damage at all. We simply need to operate in a slightly different manner. It is known that damage produces a nonlinear relation between stress and strain and that the nonlinearity can be put into evidence by analyzing the response of the system at increasing excitation amplitudes. The more damage, the larger is the level of nonlinearity, and the sooner it can be picked up in the analysis. Several studies have shown that the sensitivity of the variation of the nonlinearity with increasing damage is far better than what can be obtained from the evolution of the linear material parameters (Nagy, 1998; Van Den Abeele *et al.*, 2000a, b).

The above considerations underline the need for a nonlinear version of RUS: nonlinear RUS or NRUS. Rather than only limiting the analysis to finding the location of resonances in the spectrum and comparing them to the spectrum of an intact sample, NRUS investigates and analyzes the amplitude dependence of certain resonance frequencies and uses this information to quantify the location and degrees of nonlinearity.

Several experimental techniques have been developed that exploit the principle of nonlinearity. Some are based on the nonlinear analysis of resonance modes, others are using the amplitude dependent interaction of two-component signals. Examples are single mode nonlinear resonant ultrasound spectroscopy (Nazarov *et al.*, 1988; Nazarov and Sutin 1989; Zimenkov and Nazarov, 1995; Johnson *et al.*, 1996; Van Den Abeele *et al.*, 2000b, 2001; Van Den Abeele and De Visscher, 2000), nonlinear wave modulation spectroscopy (Antonets *et al.*, 1986; Sutin and Donskoy, 1998; Van Den Abeele *et al.*, 2000a), nonlinear wave propagation (Morris *et al.*, 1979; Cantrell and Yost, 1994; Krohn *et al.*, 2002), nonlinear time of flight spectroscopy (Kazakov *et al.*, 2002), and slow dynamics (TenCate and Shankland, 1996, TenCate *et al.*, 2000, Johnson and Sutin, 2005), etc.

In this paper we focus on the theoretical support of single mode nonlinear RUS and present a means to use this technique for the characterization and localization of damage by extending it to the analysis of multiple modes. A purely numerical treatment of this problem based on the local interaction simulation approach (LISA) can be found in Gliozzi *et al.* (2006). The development we propose here is based on an analytical formulation, validated by comparison with a multiscale model, and works in a two-way direction.

First, on the level of the direct problem, we study and predict the nonlinear signatures and analytical dependencies of the resonances from the given nonlinear elastic constants inside the sample. Over the last years, many researchers have developed numerical models to predict these effects using numerical methods such as the multiscale model or the local interaction simulation approach (Van Den Abeele *et al.*, 2004; Gliozzi *et al.*, 2006; Delsanto *et al.*, 2006). However, in order to preserve the computational simplicity of linear RUS where the resonances can be directly determined as matrix eigenvalues (cf. Visscher *et al.*, 1991), we wanted to avoid the use of numerical models and deliberately choose to return to analytical formulations. So from the viewpoint of computational speed and physical insight, a nonlinear variant of the analytical theory behind RUS is highly desirable. Of course this can only be realized under certain limiting conditions.

Secondly, a simple solution of the direct problem always makes it easier to solve the inverse problem. From this perspective, it is clear that an analytical version of the NRUS model for the direct problem will be far more advantageous than its numerical counterpart in terms of inverse characterization of nonlinearity because of the calculation speed and the transparency of the formulas.

In this paper, we limit ourselves to the derivation of analytical NRUS formulas for the case of a one-dimensional bar with distributed damage features. The manuscript is built up as follows. In Sec. II, we recall the semi-analytical version of the nonlinear wave equation for nonlinear and nonunique equations of state and comment on the general solution procedure in terms of normal modes. In Sec. III, we treat the particular case in which the damage can be represented by a classical nonlinear perturbation in the local stress strain relation, and we derive the solutions for the resonant frequency shift, and the harmonic amplitudes as function of the nonlinear characteristics of the defect and its position. In Sec. IV, we repeat this for a nonclassical representation of damage using a hysteretic stress-strain perturbation. In the fifth section, we address the inverse problem of damage characterization and location using NRUS. Our study shows that the different vibration patterns of different modes probe different parts of the structure, giving rise to mode-dependent nonlinear signatures, which can be used to solve the inverse problem. This concept has many advantages: The use of information from different modes eliminates the need for a laborious scanning apparatus as is used in traditional (linear and nonlinear) damage localization techniques (Kazakov *et al.*, 2002; Krohn *et al.*, 2002; Stoessel *et al.*, 2002; Ballad *et al.*, 2004). The detector and excitation source can therefore remain fixed (as in linear RUS), while the modes themselves do the scanning job.

II. GENERAL EQUATION AND SOLUTION

A. General nonlinear 1D equation

In this section, we consider the wave equation for a one-dimensional bar in the presence of damage, which is represented by a nonlinear and nonunique stress-strain equation. As a commonly used approximation, classical nonlinearity and hysteresis in the local stress-strain relation can be accounted for by assuming the following stress-strain relation (Landau and Lifshitz, 1969; McCall and Guyer, 1994, 1996; Van Den Abeele *et al.*, 1997; Guyer *et al.*, 1998):

$$\sigma = K(1 + \beta\varepsilon + \delta\varepsilon^2 + \cdots)\varepsilon + K\frac{\alpha}{2}[\mathrm{sign}(\partial_t\varepsilon)((\Delta\varepsilon)^2 - \varepsilon^2) - 2(\Delta\varepsilon)\varepsilon], \quad (1)$$

where K is the linear stiffness constant (modulus), $\varepsilon = \partial_x u$ is the strain, and $\Delta\varepsilon$ is the strain amplitude. The parameters β and δ are combinations of third and fourth order elastic con-

stants representing the atomic anharmonicity or acoustoelasticity. The last term in Eq. (1) accounts for the nonuniqueness of the stress-strain relation and can be deduced as the leading approximation of the Preisach representation with a uniform density distribution of a statistical ensemble of bistable microscopic units (hysterons) (McCall and Guyer, 1994, 1996; Van Den Abeele *et al.*, 1997; Guyer *et al.*, 1998). The parameter α is the strength of the hysteresis and quantifies the opening of the stress-strain loop. The loop area is $4K\alpha(\Delta\varepsilon)^3/3$, and thus relates α also to energy loss. It is important to realize that Eq. (1) is the representation of the stress-strain relation at a certain position and its parameters K, β, δ, and α may be position dependent functions.

We will focus on the description of "forced" wave resonances in a one-dimensional bar of length L, with stress-free boundaries. The excitation is supplied by a sinusoidal force with amplitude $\tilde{F}$ and circular frequency Ω at one end of the bar. The response is measured at the other end. Attenuation is accounted for by considering a damping term proportional to the velocity, containing a frequency independent quality factor. However, other models are also possible. With the above representation of the stress-strain relation, the 1D nonlinear equation for the displacement $u(x,t)$ as a function of space and time coordinates, including attenuation and external sinusoidal excitation, then becomes

$$\rho\partial_{tt}^2 u = \partial_x \Bigg\{ K\partial_x u(1 + \beta\partial_x u + \delta(\partial_x u)^2 + \cdots) + K\frac{\alpha}{2}[\mathrm{sign}(\partial_t\partial_x u)((\Delta\partial_x u)^2 - (\partial_x u)^2) - 2(\Delta\partial_x u)\partial_x u] \Bigg\} - \rho\frac{\Omega}{Q}\partial_t u + \tilde{F}\cos(\Omega t)\,\delta_{x,O}. \quad (2)$$

Here, ρ denotes the density, and $\delta_{x,0}$ is the Kronecker symbol indicating that the force is located at at $x=0$.

B. General solution

For the purpose of investigating the effect of locally distributed microdamage induced nonlinearity on the resonance properties, we will assume that the linear material properties (ρ, K, and Q) are homogeneous over the length of the one-dimensional bar, and that only the nonlinear parameters have a dependence on the space coordinate x, i.e., $\beta(x)$, $\delta(x)$, and $\alpha(x)$.

We seek the solution of Eq. (2) by decomposing the physical field u into a sum of products, separating the variables x and t (Pohit *et al.*, 1999):

$$u(x,t) = \sum_i \psi_i(x) z_i(t), \quad (3)$$

where $\{\psi_i(x)\}$, $i=-\infty,\ldots,+\infty$, is a set of chosen—and hence known—*spatial* functions satisfying the boundary conditions, i.e., $\psi_n(x)=\cos(n\pi x/L)$ for free boundaries, whereas $\{z_i(t)\}$, $i=-\infty,\ldots,+\infty$, is a set of *temporal* functions, which represent the new unknowns. We can choose the shape functions $\{\psi_i(x)\}$ such that they have the property of being mutually orthogonal, both for the displacement and the strain, in which case they represent the "normal modes,"

$$\int_0^L \psi_i(x)\psi_j(x)dx = 0 \quad \text{and}$$

$$\int_0^L \frac{\partial\psi_i}{\partial x}(x)\frac{\partial\psi_j}{\partial x}(x)dx = 0 \quad \text{for } i \neq j. \quad (4)$$

To find the general solution to the problem, we now substitute the proposed normal mode solution given by Eq. (3) into Eq. (2), multiply both sides of the equation by $\psi_n(x)$, and integrate the result over the spatial coordinate x from 0 to L. Due to the hysteretic contribution, the resulting equation is a fairly complicated equation for the unknown spatial functions $z_n(t)$ in the normal mode expansion. We can easily simplify this equation using two assumptions that are necessary to continue the algebraic calculations: (1) the excitation is performed at a frequency Ω in the close neighborhood of the resonance frequency of the mth mode (as in the single mode NRUS experiments), and (2) to express the strain amplitude $\Delta\partial_x u$ in Eq. (2), we use the approximation $\Delta\partial_x u \approx m(\pi/L)A_m$. In other words, we assume that only the linear strain solution of the problem, which is $-m(\pi/L)A_m\sin(m\pi x/L)\cos(\Omega t+\phi_m)$ as we will show later, is the dominant contribution in the strain field amplitude, and that perturbations on the strain amplitude value can be neglected since they only involve higher order terms in the strain amplitude itself (which is of the order 10^{-6}). The validity of the latter assumption and all other suppositions, such as the approximate form of the stress-strain relation in Eq. (1), is checked later in this paper by a comparison with a multiscale numerical method, which uses a nonanalytical stress-strain relation based on the full Preisach space formalism, and involves no assumption concerning the dominance of certain contributions.

After some tedious calculations we arrive at the following set of coupled differential relations:

$$\partial_{tt}^2 z_n + \frac{\Omega}{Q}\partial_t z_n + \omega_n^2 z_n = F\cos(\Omega t) - \sum_{j,k} B_{njk} z_j z_k - \sum_{j,k,l} D_{njkl} z_j z_k z_l - \sum_m H_{nm}[\mathrm{sign}(\partial_t z_m)(A_m^2 - z_m^2) - 2A_m z_m], \quad (5)$$

where $F=(2\tilde{F})/(\rho L)$, and ω_n are the frequencies of the linear (low amplitude) resonances (Visscher *et al.*, 1991), i.e., $\omega_n = n(\pi c/L)$ with c the linear bar velocity.

The advantage of using normal mode shape functions is that the mode-coupling occurs solely by way of the nonlinear interaction. There is no coupling at the linear level. The coupling constants appearing in the second, third, and fourth summations on the right hand side of Eq. (5) correspond to the components of the higher order tensors B, D, and H, and can be expressed as integrals of the strain field multiplications over x:

$$B_{njk} = \frac{2K}{\rho L}\int_0^L dx\beta(x)\partial_x\psi_n\partial_x\psi_j\partial_x\psi_k, \quad (6a)$$

$$D_{njkl} = \frac{2K}{\rho L}\int_0^L dx\,\delta(x)\partial_x\psi_n\partial_x\psi_j\partial_x\psi_k\partial_x\psi_l, \tag{6b}$$

$$H_{nm} = \frac{K}{\rho L}\int_0^L dx\,\alpha(x)\partial_x\psi_n|\partial_x\psi_m|\partial_x\psi_m. \tag{6c}$$

Thus, as soon as nonlinearity is considered inside the medium, all existing linear modes and their perturbations interact with each other at the nonlinear level, provided they *all* have a nonzero strain level *at those places where the nonlinearity is present*. This will be important for the inverse mapping of the damage location using different excitation modes.

In a single mode NRUS experiment, one analyzes the dependence of a resonance curve in the neighborhood of a chosen frequency, for instance around ω_m for mode m, at increasing forcing amplitude. In order to obtain a resonance curve at a constant forcing amplitude F, the sinusoidal source frequency Ω applied to the sample is stepped up discretely over a small range encompassing the chosen frequency. At each discrete frequency during the scan, the steady state amplitude response $|z_n(\Omega;F)|$ obtained after $5Q$ periods is monitored, for $n=m$, $2m$, $3m$, $4m$, $5m$, etc.

Before focusing on the nonlinear features resulting from a single mode NRUS simulation, we first recall the solution of the linear problem.

C. Linear solution

If everything is linear, and we investigate the response of the bar at source frequencies Ω in the close neighborhood of ω_m, the only mode that is receptive to this excitation is exactly the mode m. All other modes are not activated since the driving frequency is too far away from their corresponding resonance frequencies. The solution of Eq. (5) in the absence of nonlinearity ($\beta\equiv 0$, $\delta\equiv 0$, $\alpha\equiv 0$) is of the form $A_m\cos(\Omega t+\phi_m)$. Basic differential calculus tells us that the linear response amplitude A_m and phase ϕ_m satisfy following equations:

$$A_m = |z_m(\Omega;F)| = F/\sqrt{(\Omega^2-\omega_m^2)^2+(\Omega^2/Q)^2}, \tag{7a}$$

$$\tan\phi_m = -\Omega^2/(Q(\Omega^2-\omega_m^2)). \tag{7b}$$

Equation (7b) clearly shows that the phase lag for an excitation in the neighborhood of any mode is independent of the mode number, which implies that the attenuation depends linearly on the frequency. The true resonance frequency corresponds to the maximum amplitude response (solution of $dA/d\Omega=0$) and is given by

$$\Omega_{res} = \omega_m/\sqrt{1+1/Q^2}. \tag{7c}$$

This is the linear response. However, when nonlinearity is present in the system, one can expect that some signatures of the resonance spectrum will be modified. In general, the position of the resonance peaks will become amplitude dependent, harmonics start to appear, attenuation may become nonlinear, etc. The observation of such nonlinear behavior betrays the presence of the nonlinear zone inside the sample. The deduction of analytical relations for the expected shifts and the harmonics can help in the inverse procedure of damage localization from nonlinear observations. This is exactly the subject of the rest of the paper. We consider two cases: classical cubic nonlinearity and nonclassical hysteretic nonlinearity.

III. SOLUTION FOR CLASSICAL CUBIC NONLINEARITY

For the purpose of illustrating the calculation procedure, we now only consider the nonlinear term containing the cubic "δ" nonlinearity in Eqs. (1), (2), and (5). We will investigate the resulting effect of a δ nonlinearity on the resonance frequency and on the generation of harmonics.

A. Shift of the resonance frequency for sinusoidal excitation near mode *m*

Equation (5) for $\beta(x)\equiv\alpha(x)\equiv 0$ states that mode n can be influenced due to the cubic nonlinear interaction of the modes j, k, and l (summation over j, k, and l). If we concentrate on the excitation of the bar around the frequency ω_m and investigate the dominant nonlinear effect affecting mode m due to this excitation, we can safely say that the leading perturbation on the behavior of this mode will only be due to the combination of modes j, k, and l with $j=k=l=m$. (All other responses are zero in the linear case.) Therefore the amplitude dependence of mode m is basically governed by the following equation:

$$\partial_{tt}z_m + \frac{\Omega}{Q}\partial_t z_m + \omega_m^2 z_m = F\cos(\Omega t) - D_{mmmm}z_m^3, \tag{8a}$$

where:

$$D_{mmmm} = \frac{2}{\rho L}\int_0^L dx K\delta(x)(\partial_x\psi_m)^4 = \frac{2K}{\rho L}\left(m\frac{\pi}{L}\right)^4\int_0^L dx\,\delta(x)\sin^4\left(m\frac{\pi}{L}x\right). \tag{8b}$$

One can solve Eq. (8a) using rigorous perturbation methods or comprehensive multiple timescale methods (Landau and Lifshitz, 1969; Nayfeh, 1973; Pohit *et al.*, 1999). However, in this case, it suffices to simply substitute $z_m = A_m\cos(\Omega t+\phi_m)$ into Eq. (8a) and to equate the corresponding terms in $\cos(\Omega t)$ and $\sin(\Omega t)$ [neglecting the force term in $\cos(3\Omega t)$ arising as a secondary part of the nonlinear contribution]. As a result, we find that the amplitude A_m and the phase ϕ_m should satisfy the following equations:

$$F = A_m\left[-\left(\Omega^2-\omega_m^2-\frac{3}{4}D_{mmmm}A_m^2\right)\cos\phi_m - \frac{\Omega^2}{Q}\sin\phi_m\right],$$

$$0 = A_m\left[\left(\Omega^2-\omega_m^2-\frac{3}{4}D_{mmmm}A_m^2\right)\sin\phi_m - \frac{\Omega^2}{Q}\cos\phi_m\right], \tag{9}$$

which yields

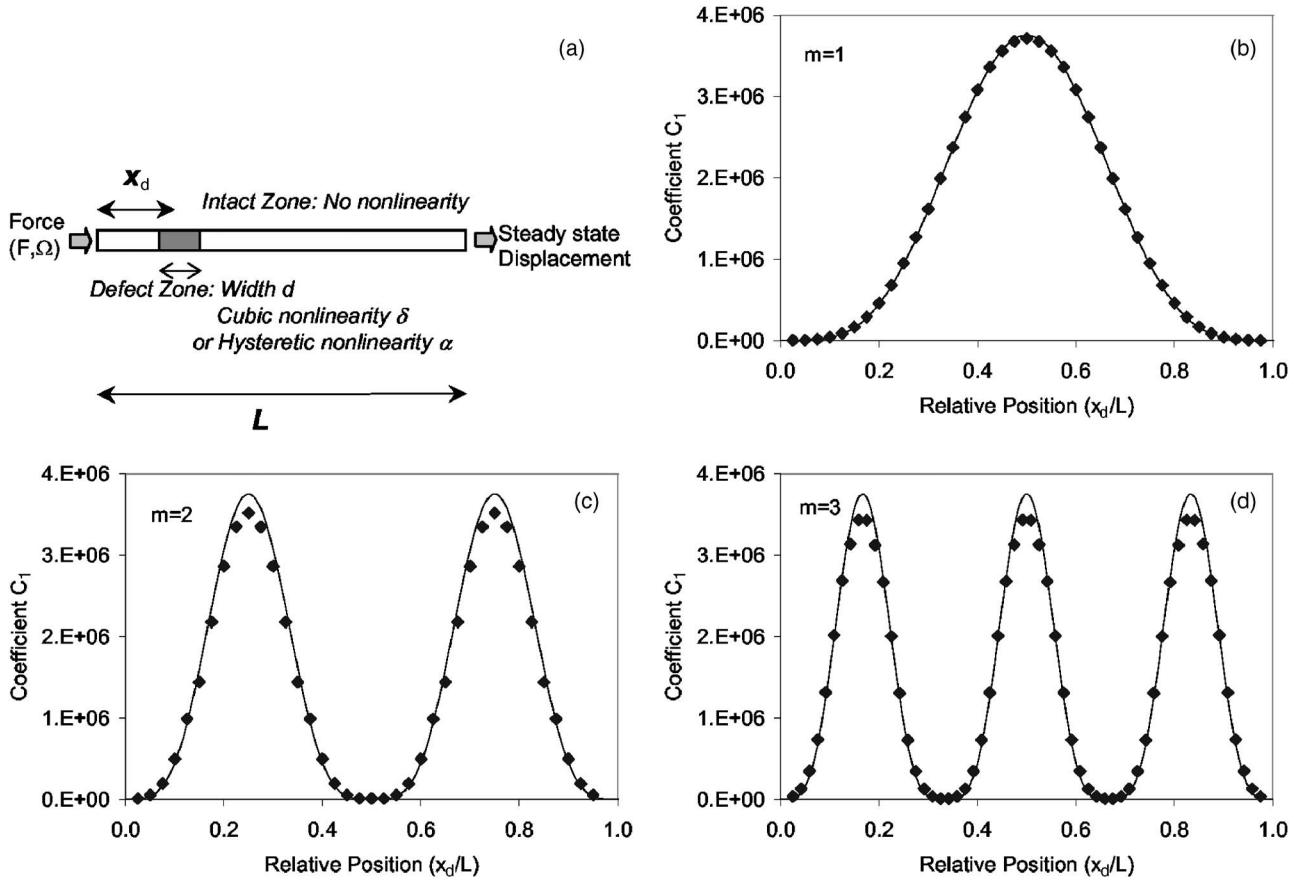


FIG. 1. Investigation of the nonlinear signature $C_{1,m}$ (proportionality coefficient in the strain amplitude dependence of the resonance frequency shift) from a cubic nonlinear defect as a function of its position x_d in a resonant bar. (a) Schematic geometry for the resonant bar simulations. (b) Signature $C_{1,1}$ for excitation near the fundamental mode ($m=1$). (c) Signature $C_{1,2}$ for the second bar mode excitation ($m=2$). (d) Signature $C_{1,3}$ for the third bar mode excitation ($m=3$). Lines represent the analytical formulas; diamonds correspond to the numerical simulations. The width of the defect zone is $d=L/20$.

$$A_m = \frac{F}{\sqrt{\left(\Omega^2 - \omega_m^2 - \frac{3}{4}D_{mmmm}A_m^2\right)^2 + (\Omega^2/Q)^2}}. \tag{10}$$

For a fixed forcing amplitude F, Eq. (10) gives an implicit relation between the response amplitude A_m and the driving frequency Ω. Since the nonlinear contribution to the amplitude change is small, we can consider the term in the denominator containing the square of A_m as being constant. Consequently, the resonance frequency, for which the response amplitude is maximal, is approximately given by

$$\Omega_{res}(A_m) \approx \frac{\omega_m\sqrt{1 + (3/4\omega_n^2)D_{mmmm}A_m^2}}{\sqrt{1 + 1/Q^2}} \approx \Omega_{res}(0)\left(1 + \frac{3}{8\omega_n^2}D_{mmmm}A_m^2\right). \tag{11}$$

Expressed in terms of the maximal strain amplitude $\varepsilon_m = \Delta\varepsilon_{max} = (m\pi/L)\, A_{m,max}$, this becomes

$$\Omega_{res}(\varepsilon_m) \approx \Omega_{res}(0)\left[1 + \frac{3}{4L}\varepsilon_m^2\int_0^L dx\,\delta(x)\sin^4\left(m\frac{\pi}{L}x\right)\right]. \tag{12}$$

The relative frequency shift is quadratic in the strain, which agrees with literature on nonlinear harmonic oscillators with uniform distribution of nonlinearity (Landau and Lifshitz, 1969). It is important to notice, however, that the quadratic dependence on strain is irrespective of the distribution of the cubic nonlinearity: global or localized.

One possible way to express the nonlinearity distribution over the bar is by means of a form function $G(x)$, such that $\delta(x) = \bar{\delta}G(x)$, with $\bar{\delta} = (1/L)\int_0^L \delta(x)dx$ the average value of the nonlinearity over the bar, and $(1/L)\int_0^L G(x)dx = 1$. The advantage of doing so is that the form function is independent of the strength of the nonlinearity and only accounts for the geometric effects. For a localized damage area, centered at $x=x_d$, extending over a width $d \ll L$, and represented by a local constant cubic nonlinearity $\hat{\delta}$ [Fig. 1(a)], we use

$$\delta(x) = \bar{\delta}G_{\text{Local}}(x;x_d,d) \tag{13a}$$

with

$$\bar{\delta} = \hat{\delta}\frac{d}{L} \quad \text{and}$$

$$G_{\text{Local}}(x;x_d,d) = \begin{cases} 0 & \text{if } x \notin \left[x_d - \frac{d}{2}, x_d + \frac{d}{2}\right], \\ \frac{L}{d} & \text{if } x \in \left[x_d - \frac{d}{2}, x_d + \frac{d}{2}\right], \end{cases} \tag{13b}$$

and we obtain the following expression for the relative frequency shift

TABLE I. Parameters used in the simulations.

General parameters	$K=10$ GPa, $\rho=2600$ kg/m^3, $Q=80$, $L=0.25$ m, $d=L/20$	
Single mode NRUS simulation parameters: *analytical* approach	Cubic nonlinearity	Hysteretic nonlinearity
	$\hat{\delta}=-10^8$	$\hat{\alpha}=2000$
Single mode NRUS simulation parameters: *numerical* approach	Cubic nonlinearity	Hysteretic nonlinearity
	$\delta_N=-3\times10^{-12}$ which yields $\hat{\delta}=(\delta_N K^2)/3=-10^8$	$\hat{\gamma}=10^{-3}$ which yields $\hat{\alpha}=\hat{\gamma}K^2/5\times10^{13}=2000$
Discretization parameters for *numerical* simulation: (m=mode number; Ω=circular frequency)	$\Delta x=L/(120m)$	$\Delta t=2\pi/(384m\Omega)$

$$\Omega_{res,\mathrm{Local}}(\varepsilon_m)\approx\Omega_{res}(0)\left[1+\frac{3}{4}\frac{\hat{\delta}d}{L}\sin^4\left(m\frac{\pi}{L}x_d\right)\varepsilon_m^2\right]$$
$$=\Omega_{res}(0)\left[1+\frac{3}{4}\bar{\delta}\sin^4\left(m\frac{\pi}{L}x_d\right)\varepsilon_m^2\right]$$
$$=\Omega_{res}(0)[1+C_{1,m}\varepsilon_m^2]. \quad (13c)$$

For a global constant nonlinearity homogeneously extending all along the bar ($\bar{\delta}=\hat{\delta}$, and $G_{\mathrm{Global}}(x)=1$), the shift becomes

$$\Omega_{res,\mathrm{Global}}(\varepsilon_m)\approx\Omega_{res}(0)\left[1+\tfrac{9}{32}\bar{\delta}\varepsilon_m^2\right]. \quad (14)$$

In most cases $\hat{\delta}$ is considered to be negative (softening), implying that the resonance frequency *reduces* quadratically with amplitude. Note that the above expressions [Eqs. (13c) and (14)] are just special cases to illustrate the effect of localized and global damage on the frequency shift. The proposed formalism is far more widely applicable since the integral in Eq. (12) can be calculated for an arbitrary form function describing the distribution of the nonlinearity. We also observe that the coefficient in front of the square strain dependence for the case of uniformly distributed nonlinearity in Eq. (14) is independent of the mode number, whereas a clear mode dependence occurs in the coefficient $C_{1,m}$ of Eq. (13c). The (sine)^4 dependence on the position x_d of the damage in Eq. (13c) is quite understandable keeping in mind that modes will contribute to the nonlinearity to an amount that is proportional to their strain-field amplitude at the defect position. Since the shift is essentially the result of a threefold nonlinear interaction of "mode m" type resonances thereby affecting the same mode m, we indeed arrive at the fourth order dependence. Another important observation is that the shift depends on the product of the local nonlinearity strength $\hat{\delta}$ and the relative width d/L, causing the average nonlinearity parameter $\bar{\delta}=\hat{\delta}d/L$ to be an "effective" nonlinear strength parameter.

We have compared the analytical prediction of the shift for the case of localized damage [Eq. (13c)] to the results calculated by means of a numerical multiscale finite difference time domain (FDTD) model discussed in one of our previous publications (Van Den Abeele *et al.*, 2004), ignoring hysteretic nonlinearity. Figure 1(b) shows the numerical and analytical predictions of the proportionality coefficient $C_{1,m}$ in Eq. (13c) as function of the damage position, for the fundamental mode ($m=1$) and for a fixed value of d and $\hat{\delta}$. The parameter values used in the analytical simulations are given in Table I, together with the corresponding numerical simulation values. The agreement between the analytical predictions and the numerical simulation results is excellent and validates the analytical NRUS predictions. Similar agreement can be found for higher order modes: $m=2$ in Fig. 1(c) and $m=3$ in Fig. 1(d). This and further comparisons confirm that the perturbation model and the assumptions we have made to find the analytical solution indeed give the leading terms of the numerical solution, including the proportionality coefficients that appear in the dependencies.

In addition, we investigated the influence of the width W of the defect zone for a fixed centered location of the defect ($x_d=L/2$) with a fixed local nonlinearity $\hat{\delta}$ extending from $(L-W)/2$ to $(L+W)/2$. In this case $\bar{\delta}=\hat{\delta}W/L$ and $G_{\mathrm{Centered}}(x;W)=G_{\mathrm{Local}}(x;L/2,W)$, cf. Eq. (13b). The analytical prediction [following the general Eq. (12)] and numerical results (FDTD method) illustrated in Fig. 2(a) again agree very well for $m=1$. Initially, the shift increases linearly with the width [Eq. (13c) if $W=d\ll L$]. However, for larger values, a saturation is obtained due to the particular shape of the resonance mode that is weighing the integral contribution in Eq. (12) (strain field is a sine function). The asymptotic behavior is the value of the proportionality coefficient for a complete volumetrically nonlinear bar with cubic nonlinearity $\bar{\delta}=\hat{\delta}$ [Eq. (14)]. The predictions for higher modes are shown in Fig. 2(b) and their behavior can be readily understood from the particular mode shapes of the second and third resonances.

B. Amplitude of the harmonics for sinusoidal excitation near mode *m*

To find the leading order amplitudes of the second and third harmonics, given a sinusoidal excitation near mode m, we need to solve the following equations:

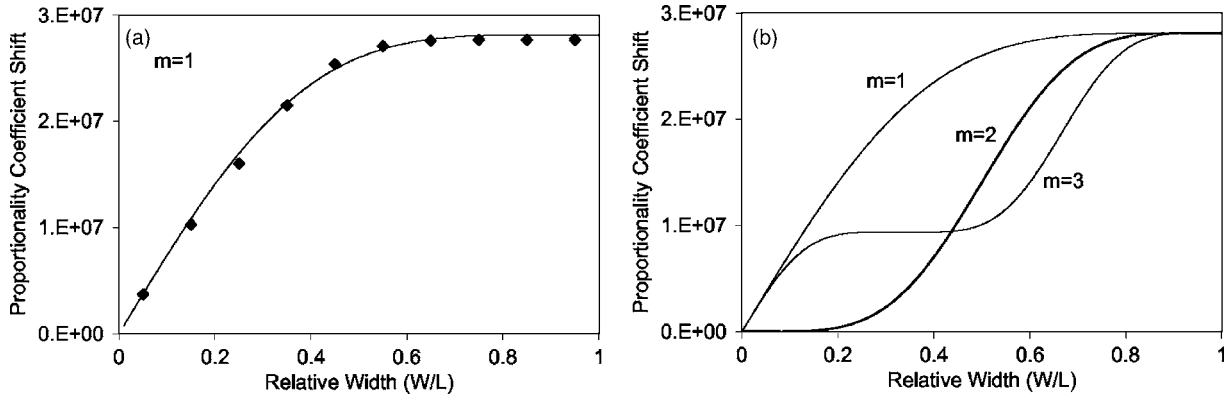


FIG. 2. Investigation of the proportionality coefficient in the strain amplitude dependence of the resonance frequency shift from a cubic nonlinear defect as a function of its width (W). The defect is located in the middle of the bar ($x_d=L/2$). (a) For the fundamental mode excitation ($m=1$). Lines represent the analytical formulas; diamonds correspond to the numerical simulations. (b) Analytical predictions for the first three bar modes.

$$\partial_{tt}^2 z_{2m} + \frac{\Omega}{Q}\partial_t z_{2m} + \omega_{2m}^2 z_{2m} = -D_{2mmmm} z_m^3, \tag{15}$$

$$\partial_{tt}^2 z_{3m} + \frac{\Omega}{Q}\partial_t z_{3m} + \omega_{3m}^2 z_{3m} = -D_{3mmmm} z_m^3. \tag{16}$$

By substituting the solution for z_m into Eqs. (15) and (16), the right hand terms will be composed of source terms containing $\cos(\Omega t)$ and $\cos(3\Omega t)$. Since none of these terms have a frequency near ω_{2m}, there will be no activation of the second harmonic and therefore z_{2m} will be zero at all times. On the other hand, the 3Ω component on the right hand side of Eq. (15) will produce a nonzero amplitude for the third harmonic when Ω is swept around ω_m. Representing z_{3m} by $A_{3m}\cos(3\Omega t+\phi_{3m})$, the amplitude A_{3m} can be easily calculated and reads

$$A_{3m} = \frac{|D_{3mmmm}|}{4} \frac{A_m^3}{\sqrt{((3\Omega)^2-\omega_{3m}^2)^2 + \Omega^2(3\Omega)^2/Q^2}}. \tag{17}$$

The amplitude of the strain field of the third harmonic for an excitation near the resonance frequency ω_m of mode m consequently becomes

$$\varepsilon_{3m} = \left| \frac{3Q}{2L}\int_0^L \delta(x)\sin\left(\frac{3m\pi}{L}x\right)\sin^3\left(\frac{m\pi}{L}x\right)dx \right| \varepsilon_m^3. \tag{18}$$

The integral can be calculated for an arbitrary form function describing the distribution of the cubic nonlinearity over the bar. Irrespective of the distribution, the analytical solution shows that the third harmonic is proportional to the third power in the strain, and that the quality factor plays an important role in the efficiency of the harmonic generation. The third power dependence on the strain is well known and physically comes from a threefold interaction of "mode m" type resonances to form a $3m$ vibration [cf. the diagrammatic schemes in McCall (1994)]. Regarding the efficiency, the fact that higher Q values (lower attenuation) will produce higher harmonics is related to the particular case in which the third harmonic of the mode m coincides with the normal mode $3m$.

For the two special cases of a localized damage zone (Eq. (13b)) and a globally distributed cubic nonlinearity we respectively obtain

$$\varepsilon_{3m,\text{Local}} = \frac{3}{2}Q|\bar{\delta}| \left| \sin\left(\frac{3m\pi}{L}x_d\right)\sin^3\left(\frac{m\pi}{L}x_d\right) \right| \varepsilon_m^3 = C_{3,m}\varepsilon_m^3, \tag{19}$$

$$\varepsilon_{3m,\text{Global}} = \frac{3}{16}Q|\bar{\delta}|\varepsilon_m^3. \tag{20}$$

Note again that the coefficient $C_{3,m}$ for localized nonlinearity depends on the excited mode, whereas no mode dependence appears for the uniform distributed nonlinearity.

For the higher order even harmonics, one can repeat the reasoning applied for the second harmonic. The cubic nonlinearity never produces a source contribution that contains an even multiple of the source frequency near ω_{2km}, for any integer k value. All even harmonics thus remain zero. For the odd harmonics, it is always possible to identify combinations of source terms that lead to a source contribution with frequency $(2k+1)\Omega$. For the fifth harmonic there are three terms that yield such a source term, giving rise to the following equation:

$$\partial_{tt}^2 z_{5m} + \frac{\Omega}{Q}\partial_t z_{5m} + \omega_{5m}^2 z_{5m} = -D_{5m3mmm} z_{3m} z_m^2 - D_{5mm3mm} z_{3m} z_m^2 - D_{5mmm3m} z_{3m} z_m^2. \tag{21}$$

The three terms on the right hand side of the equation physically represent the (only) three possible threefold interactions of a $3\omega_m$ wave with two ω_m waves to form a $5\omega_m$ wave. The $3\omega_m$ wave can be the first, middle, or last wave in this interaction. Without doing any calculation, the physical factors for the amplitude of the $5\omega_m$ wave can be easily predicted as follows: Recalling Eqs. (18)–(20), we know that the $3\omega_m$ wave amplitude comes from a threefold interaction of three ω_m waves yielding an amplitude for the $3\omega_m$ component, which is proportional to ε_m^3, to Q, and to $\bar{\delta}$. To form a $5\omega_m$ wave, we need a second threefold interaction, this

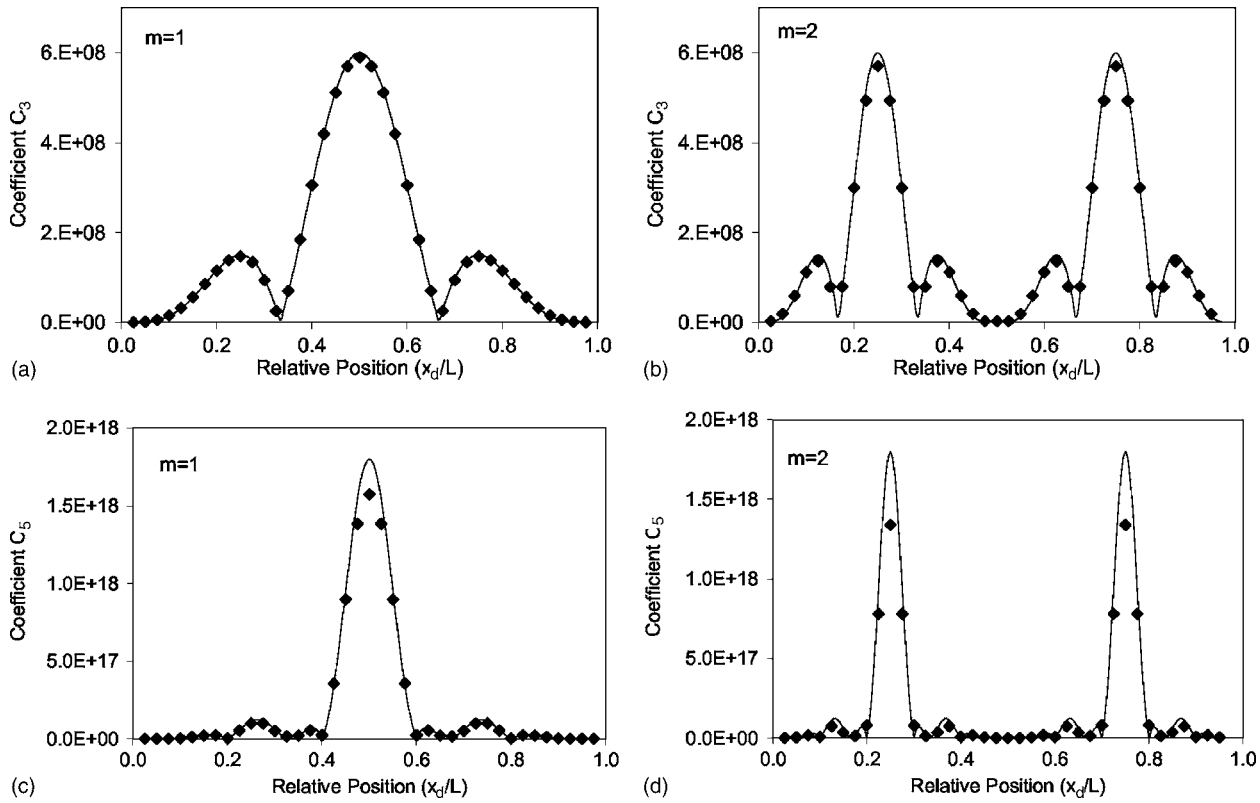


FIG. 3. Investigation of the nonlinear signature $C_{3,m}$ and $C_{5,m}$ (proportionality coefficient in the strain amplitude dependence of the third and fifth harmonics) from a cubic nonlinear defect as a function of its position x_d in a resonant bar. (a) Signature $C_{3,1}$ for fundamental mode excitation ($m=1$). (b) Signature $C_{3,2}$ for the second bar mode excitation ($m=2$). (c) Signature $C_{5,1}$ for the fundamental mode excitation ($m=1$). (d) Signature $C_{5,2}$ for the second bar mode excitation ($m=2$). Lines represent the analytical formulas; diamonds correspond to the numerical simulations. The width of the defect zone is $d=L/20$.

time between two ω_m and one $3\omega_m$ waves, again with account of the efficiency factors Q and $\bar{\delta}$. Therefore, the amplitude of the $5\omega_m$ wave should be proportional to ε_m^5, to Q^2, and to $\bar{\delta}^2$.

In analogy with the third harmonic, we went through the analytical calculations. After some tedious calculus, we indeed obtain a general expression for the strain amplitude of the fifth harmonic for an excitation near the resonance frequency ω_m of mode m, which shows the proportionality to the fifth power of the fundamental strain amplitude, and the quadratic Q and $\bar{\delta}$ dependences, irrespective of the damage distribution:

$$\varepsilon_{5m} = \frac{45Q^2}{4L^2}\left|\int_0^L \delta(x)\sin\left(\frac{3m\pi}{L}x\right)\sin^3\left(\frac{m\pi}{L}x\right)dx\right| \times \left|\int_0^L \delta(x)\sin\left(\frac{5m\pi}{L}x\right)\sin\left(\frac{3m\pi}{L}x\right)\sin^2\left(\frac{m\pi}{L}x\right)dx\right|\varepsilon_m^5. \tag{22}$$

For the two special cases of a localized damage zone and a globally distributed cubic nonlinearity, we respectively obtain

$$\varepsilon_{5m,\mathrm{Local}} = \frac{45}{5}Q^2\bar{\delta}^2 \times \left|\sin\left(\frac{5m\pi}{L}x_d\right)\sin\left(\frac{3m\pi}{L}x_d\right)\sin^2\left(\frac{m\pi}{L}x_d\right)\right|\varepsilon_m^5 = C_{5,m}\varepsilon_m^5, \tag{23}$$

$$\varepsilon_{5m,\mathrm{Global}} = \frac{45}{256}Q^2\bar{\delta}^2\varepsilon_m^5. \tag{24}$$

The particular position dependent behavior of the coefficients $C_{3,m}$ and $C_{5,m}$ in Eqs. (19) and (23) in the case of a localized damage zone with cubic nonlinearity is checked by comparing it to the results of the multiscale model (Van Den Abeele *et al.*, 2004) as was done before for the amplitude dependent frequency shift. Figure 3 illustrates the good qualitative and quantitative agreement between the analytical and numerical results for the proportionality coefficients of the third and fifth harmonics taken at the resonance frequency of the fundamental of mode $m=1$ and mode $m=2$. Of importance for experiments is that a localized defect may show a reasonable resonance frequency shift without a sign of harmonics being generated. Indeed, by comparing the results for the resonance frequency shift in Fig. 1 to the harmonic generation efficiency in Fig. 3, we notice, for instance, no harmonic generation when the defect is located at $L/3$ for $m=1$, or at $L/6$ for $m=2$, whereas the frequency shift is

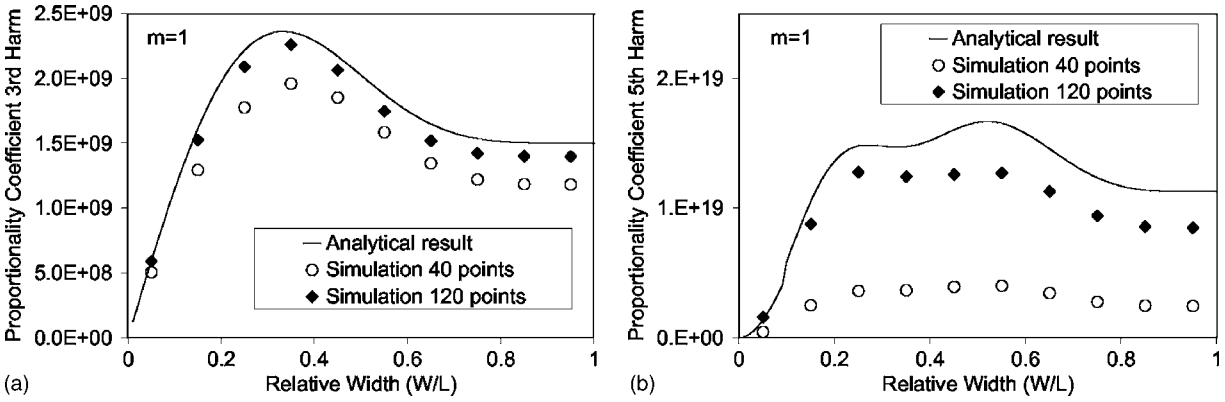


FIG. 4. Investigation of the proportionality coefficient in the strain amplitude dependence of the third and fifth harmonics from a cubic nonlinear defect as a function of its width (W). The defect is located in the middle of the bar ($x_d = L/2$). (a) Third harmonic efficiency for the fundamental mode excitation ($m=1$). (b) Fifth harmonic efficiency for the fundamental mode excitation ($m=1$). Lines represent the analytical formulas; diamonds correspond to the numerical simulations, with different discretizations.

nonzero at these locations. This is entirely due to the influence of the mode shapes on the efficiency of the harmonic generation.

Finally, in Fig. 4, we illustrate the width (W) dependence of the damaged zone on the third and fifth harmonic proportionality coefficients for excitation near mode $m=1$. Here we notice that—although still satisfactory—the agreement between analytical prediction and numerical result becomes somewhat inferior. However, this is entirely due to the fact that the numerical results for higher harmonics and large nonlinearity zones critically depend on the discretization. The difference between 40 and 120 grid points in the multiscale model is clearly visible.

The proportionality coefficients $C_{3,m}$ and $C_{5,m}$ for the chosen value of $\hat{\delta}$ are of the order of 10^8 and 10^{18}, respectively. This means that a strain wave of amplitude 10^{-6} produces third and fifth harmonics that are 80 to 120 dB below the fundamental component. Currently data acquisition hardware can easily manage to receive 16-bit data at low frequencies, which corresponds to 90-dB dynamic range. However, the value of $\hat{\delta}$ corresponds to typical values for uniformly distributed nonlinearity estimated from nonlinear wave propagation experiments on (intact) rocks (Van Den Abeele and Johnson, 1996) and will probably be on the low side when dealing with realistic localized microdamage. (A factor 10 increase in $\hat{\delta}$ results in third and fifth harmonics that are 60 to 80 dB down, respectively.) In addition, it is expected that the dynamic range will improve considerably in the future (up to 21–24 bit), leading to the conclusion that even small harmonics can be detected with confidence. A general caution in this story, however, is that transducers may generate harmonics as well.

IV. NONCLASSICAL NONLINEARITY

In this section, we focus on the influence of the leading term in the nonuniqueness of the nonlinear stress-strain equation. Therefore, we set all classical nonlinearity to be zero $[\beta(x) \equiv \delta(x) \equiv 0]$ and consider only the effect of the hysteresis distribution $\alpha(x)$ in Eq. (1), (2), and (5).

The procedure for finding the amplitudes of the fundamental and of the higher harmonics is to identify the oscillating terms on the right hand side of Eq. (5) with circular frequencies equal to $(n/m)\Omega \approx \omega_h$. Only those terms will contribute to the amplitude of $z_n(t)$. To do this, we are forced to express the sign function into a Fourier series (cf. McCall and Guyer, 1994). Using

$$\begin{aligned}\mathrm{sign}(\partial_t z_m) &= \mathrm{sign}(-\Omega A_m \sin(\Omega t + \phi_m)) \\ &= -\sum_{k=0}^{\infty} \frac{4}{(2k+1)\pi} \sin((2k+1)(\Omega t + \phi_m)),\end{aligned} \tag{25}$$

it becomes clear that only the odd harmonics will be generated with nonzero amplitudes, since there will be no oscillating contributions on the right hand side source terms of Eq. (5) containing even multiples of Ω.

A. Shift of the resonance frequency for sinusoidal excitation near mode *m*

For $n=m$, we basically have to find the solution of

$$\begin{aligned}\partial_{tt}^2 z_m + \frac{\Omega}{Q}\partial_t z_m + \omega_n^2 z_m = F\cos(\Omega t) \\ - H_{mm}\left[\frac{-8}{3\pi}A_m^2 \sin(\Omega t + \phi_m)\right. \\ \left. - 2A_m^2 \cos(\Omega t + \phi_m) + \cdots\right],\end{aligned} \tag{26}$$

where the three dots represent terms containing higher order oscillations. In a similar manner as in the case of cubic nonlinearity, we now obtain that

$$\begin{aligned}F = A_m\left[-(\Omega^2 - \omega_m^2 + 2H_{mm}A_m)\cos\phi_m\right. \\ \left. -\left(\frac{\Omega^2}{Q} + \frac{8}{3\pi}H_{mm}A_m\right)\sin\phi_m\right],\end{aligned}$$

$$0 = A_m\left[(\Omega^2 - \omega_m^2 + 2H_{mm}A_m)\sin\phi_m - \left(\frac{\Omega^2}{Q} + \frac{8}{3\pi}H_{mm}A_m\right)\cos\phi_m\right], \quad (27)$$

which yields

$$A_m = \frac{F}{\sqrt{(\Omega^2 - \omega_m^2 + 2H_{mm}A_m)^2 + (\Omega^2/Q + (8/3\pi)H_{mm}A_m)^2}}. \quad (28)$$

Equation (28) gives again an implicit relation between the response amplitude A_m and the driving frequency Ω, for a fixed forcing amplitude F. Under the assumption of small nonlinear contributions, the resonance frequency can be analytically expressed, as follows:

$$\Omega_{res}(A_m) = \frac{\omega_m\sqrt{(1-2/\omega_m^2)(1+4/3\pi Q)H_{mm}A_m}}{\sqrt{1+1/Q^2}} \approx \Omega_{res}(0)\left(1 - \frac{1}{\omega_m^2}\left(1+\frac{4}{3\pi Q}\right)H_{mm}A_m\right), \quad (29)$$

which can be rewritten in terms of the maximal strain response amplitude $\varepsilon_m = (m\pi/L)\, A_{m,max}$ as

$$\Omega_{res}(\varepsilon_m) \approx \Omega_{res}(0)\left(1 - \left(1+\frac{4}{3\pi Q}\right)\frac{cH_{mm}\varepsilon_m}{\omega_m^3}\right) = \Omega_{res}(0)\left[1 - \left(1+\frac{4}{3\pi Q}\right)\frac{\varepsilon_m}{L}\int_0^L dx\alpha(x) \times \left|\sin^3\left(m\frac{\pi}{L}x\right)\right|\right]. \quad (30)$$

The relative frequency shift for hysteretic nonlinearity changes linearly with the strain, and again this agrees with literature on lumped nonlinear harmonic oscillators (Guyer *et al.*, 1998). This is a well-known result, which is experimentally validated for certain materials and also numerically verified, at least qualitatively (Johnson *et al.*, 1996; Van Den Abeele *et al.*, 2000a, b; Gusev 2000; Capogrosso-Sansone and Guyer, 2002; Scalerandi *et al.*, 2003; Van Den Abeele *et al.*, 2004). However, the analytical expressions deduced here show clearly that the linear dependence on strain is irrespective of the status of the hysteretic nonlinearity: global or localized.

The spatial distributions of the hysteretic nonlinearity over the bar can again be expressed by use of form functions $\alpha(x) = \bar{\alpha}G(x)$. For a localized damage, centered at $x = x_d$, extending from $[x_d - d/2, x_d + d/2]$ (with $d \ll L$) and represented by a local constant hysteretic nonlinearity $\hat{\alpha}$ [cf. Eq. (13b): $G(x) = G_{Local}(x)$, $\bar{\alpha} = \hat{\alpha}d/L$, and Fig. 1(a)], we obtain

$$\Omega_{res,\mathrm{Local}}(\varepsilon_m) \approx \Omega_{res}(0)\left(1 - \frac{\hat{\alpha}d}{L}\left(1+\frac{4}{3\pi Q}\right) \times \left|\sin^3\left(m\frac{\pi}{L}x_d\right)\right|\varepsilon_m\right) \approx \Omega_{res}(0)\left(1 - \bar{\alpha}\left(1+\frac{4}{3\pi Q}\right)\left|\sin^3\left(m\frac{\pi}{L}x_d\right)\right|\varepsilon_m\right) = \Omega_{res}(0)[1 - X_{1,m}\varepsilon_m]. \quad (31)$$

For a global constant hysteretic nonlinearity extending all along the bar $[G_{\mathrm{Global}}(x) = 1,\ \bar{\alpha} = \hat{\alpha}]$, the shift becomes

$$\Omega_{res,\mathrm{Global}}(\varepsilon_m) \approx \Omega_{res}(0)\left[1 - \frac{4\bar{\alpha}}{3\pi}\left(1+\frac{4}{3\pi Q}\right)\varepsilon_m\right]. \quad (32)$$

Since $\hat{\alpha}$ represents the hysteretic strength (and is by consequence a positive parameter), the change in resonance frequency will always be directed towards lower values. The most general expression for the shift contains a weighted integral over an arbitrary distribution of the nonlinearity. Again, we can interpret the (sine)^3 dependence on the position x_d of the damage in Eqs. (30) and (31) keeping in mind that modes will contribute to the nonlinearity to an amount that is proportional to their strain-field amplitude at the defect position. This time, however, the shift is essentially the result of a twofold nonlinear interaction of mode m, which affects the same mode m ["hysteresis corresponds to a quadratic nonlinearity," see McCall and Guyer (1994, 1996) and Gusev (2000)]. The absolute value operation is the consequence of the nonuniqueness of the stress-strain relation. Further, we notice in Eq. (31) that the shift for a localized damage feature depends on the "effective" or average nonlinear strength parameter $\bar{\alpha} = \hat{\alpha}d/L$, and that the proportionality coefficient in the strain dependent term for a uniform distribution is independent of the mode number, as in the case of classical nonlinearity.

Besides quantifying the relative frequency shift at resonance, Eq. (28) reveals another interesting relation connected with the contribution of hysteresis to the attenuation, confirming results put forward by the lumped element treatment (Guyer *et al.*, 1998) and by recently developed numerical approaches (Capogrosso-Sansone and Guyer, 2002). If the (linear) quality factor is not too small ($Q > 10$), the hysteretic nonlinearity will be responsible for an amplitude dependent change in the loss factor, which is mainly caused by a change in the Q factor satisfying

$$\frac{1}{Q_{NL}} = \frac{1}{Q_L} + \frac{8}{3\pi\omega_m^2}H_{mm}A_m = \frac{1}{Q_L} + \frac{8}{3\pi}\frac{cH_{mm}\varepsilon_m}{\omega_m^3}. \quad (33)$$

The amplitude dependent Q factor is typical for hysteretic systems only. For instance, Eq. (10) for classical cubic nonlinearity does not show any amplitude dependence in the second term of the denominator, and therefore no nonlinear damping is accounted for.

Recalling that $\varepsilon_m \approx m\pi FQ/(L\omega_m^2)$ at resonance, we can compare the relative decrease in strain amplitude at resonance with the relative decrease of the resonance frequency obtained in Eq. (30). The analytical treatment then yields the following constant relation

$$\frac{[\Omega_{res,0}-\Omega_{res}(A)]/\Omega_{res,0}}{[\varepsilon_{res,0}-\varepsilon_{res}(A)]/\varepsilon_{res,0}}=\frac{\Delta\Omega_{res}(A)/\Omega_{res,0}}{Q_{NL}(1/Q_{NL}-1/Q_L)}$$
$$\approx\frac{3\pi}{8Q}\left(1+\frac{4}{3\pi Q}\right)\approx\frac{3\pi}{8Q}=\frac{1.178}{Q}, \tag{34a}$$

or equivalently

$$\frac{(1/Q_{NL}-1/Q_L)}{\Delta\Omega_{res}(A)/\Omega_{res,0}}\approx 0.849. \tag{34b}$$

Previous discussions on this topic (using a lumped element treatments or a numerical approach) only qualitatively put forward the "order Q" relation for the ratio of the relative amplitude decrease due to hysteretic nonlinearity and the relative frequency shift. Here, the proportionality is calculated using a consequent analytical treatment, and the quantitative value will be verified in the next paragraph by comparison with the multiscale model. As for the experimental verification, Johnson and Sutin (2005) measured the ratio in Eq. (34b) for several materials (metals, rocks, and cracked glass) and found experimental values that range between 0.28 and 1.10. This is quite good in agreement with the analytical predictions.

We again take advantage of the multiscale model (Van Den Abeele *et al.*, 2004) to validate the above analytical formulas in the limiting case of uniform modulus, uniform quality factor, and localized hysteretic nonlinearity. A flexible treatment of hysteretic nonlinearity in the stress-strain relation for rocks and microcracked materials has been put forward by McCall and Guyer (1994, 1996) by means of the Preisach formalism. Since then, this formalism has become the standard for the development of numerical wave propagation and wave resonance simulation techniques in the case of hysteretic media (Scalerandi *et al.*, 2003; Gusev, 2000; Capogrosso-Sansone and Guyer, 2002; Aleshin *et al.*, 2004; Van Den Abeele *et al.*, 2004; Gliozzi *et al.*, 2006). We refer to the above mentioned references for more details.

For the analytical-numerical comparison and the validation of our approximations and assumptions introduced in this paper, we follow the multiscale model of Van Den Abeele *et al.* (2004). It is important to notice that the wave propagation simulation is treated rigorously in the multiscale model and that the model does not introduce leading order assumptions as was done in the analytical case. The Preisach-Mayergoysz (PM) spaces used to model the hysteretic nonlinearity in the numerical simulations consist of uniformly distributed units over a range of −5 to 5 MPa. The other parameter values used in the analytical and numerical simulations are listed in Table I. Figure 5(a) shows the numerical and analytical predictions of the proportionality coefficient $X_{1,m}$ in Eq. (31) for the fundamental mode ($m=1$) and for a fixed value of d and $\hat{\alpha}$. A similar good agreement can be found for higher order modes: $m=2$ in Fig. 5(b).

Figure 5(c) illustrates the ratio of the absolute Q^{-1} reduction to the relative frequency reduction obtained with the numerical model. The average value is indeed close to the analytically predicted value of 0.849, except in the center of the bar, where higher order effects may play a role, and at the ends of the bar where the effect of the nonlinearity is highly tempered by the mode shape.

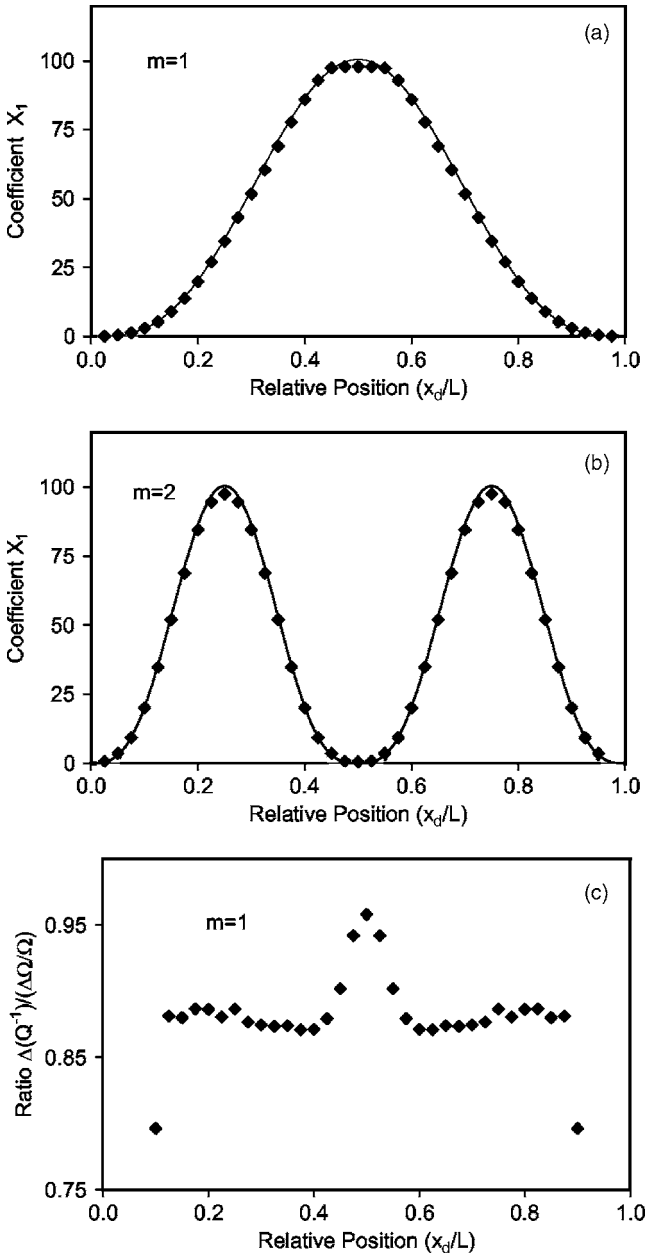


FIG. 5. Investigation of the nonlinear signature $X_{1,m}$ (proportionality coefficient in the strain amplitude dependence of the resonance frequency shift) from a hysteretic nonlinear defect ($d=L/20$) as a function of its position in a resonant bar. (a) Signature $X_{1,1}$ for excitation around the fundamental mode ($m=1$). (b) Signature $X_{1,2}$ for the second bar mode excitation. (c) Ratio of the absolute increase in Q^{-1} to the relative frequency shift for the fundamental mode ($m=1$) derived from the numerical simulations.

In addition, we investigated the influence of the width W of the defect zone for a fixed centered location of the defect with a fixed local nonlinearity value $\hat{\alpha}$ [$\bar{\alpha}=\hat{\alpha}W/L$ and $G_{\text{Centered}}(x;W)=G_{\text{Local}}(x;L/2,W)$]. The analytical prediction and the numerical results of the strain proportionality coefficient in Eq. (30) are illustrated in Fig. 6, and both results agree very well for $m=1$. Initially, the shift increases linearly with the width [Eq. (31) if $W=d\ll L$]. However, for larger values, a saturation is obtained due to the particular shape of

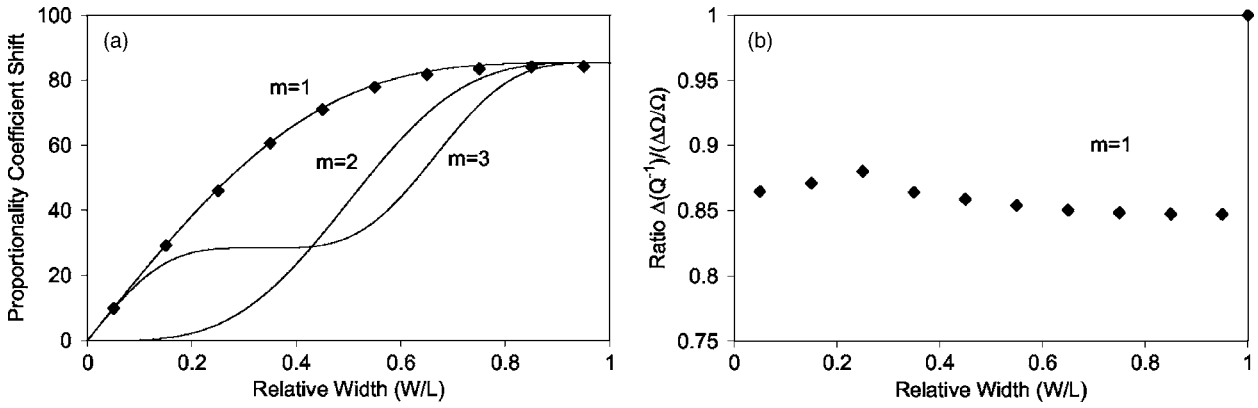


FIG. 6. Investigation of the proportionality coefficient in the strain amplitude dependence of the resonance frequency shift from a hysteretic nonlinear defect as a function of its width (W). The defect is located in the middle of the bar ($x_d = L/2$). (a) Analytical predictions for the first three bar modes, and numerical results for the fundamental mode excitation ($m=1$). Lines represent the analytical formulas; diamonds correspond to the numerical simulations. (b) Ratio of the absolute increase in Q^{-1} to the relative frequency shift for the fundamental mode ($m=1$) as function of the width of the defect.

the resonance mode that is weighing the contribution (strain field is a sine function). The asymptotic behavior is the value of the proportionality coefficient for a complete volumetrically nonlinear bar with hysteretic nonlinearity $\bar{\alpha}=\hat{\alpha}$ [Eq. (32)]. The predictions for higher modes are also shown in Fig. 6(a) and can be readily understood from the particular mode shapes of the second and third resonances.

While performing the simulations as a function of the width of the defect zone, we always observed the predicted constant value of 0.849 for the ratio of the absolute attenuation increase to the relative frequency shift [Fig. 6(b)]. This indeed supports the conclusion that this ratio is independent of the extent of the defect.

B. Amplitude of the harmonics for sinusoidal excitation near mode *m*

We already stated that no even harmonics will be generated for a given sinusoidal excitation near mode m, since there will be no oscillating contributions on the right hand side source terms of Eq. (5) with even multiples of Ω when $\beta(x)\equiv 0$. We can find the leading order amplitude of the third harmonics by solving the following equation:

$$\partial_{tt}^2 z_{3m} + \frac{\Omega}{Q}\partial_t z_{3m} + \omega_{3m}^2 z_{3m} = -H_{3mm}\frac{8}{15\pi}A_m^2 \sin(3(\Omega t + \phi_m)). \tag{35}$$

Representing z_{3m} by $A_{3m}\cos(3\Omega t+\phi_{3m})$, the amplitude A_{3m} can be easily calculated and reads

$$A_{3m} = \frac{8|H_{3mm}|}{15\pi}\frac{A_m^2}{\sqrt{((3\Omega)^2-\omega_{3m}^2)^2+\Omega^2(3\Omega)^2/Q^2}}. \tag{36a}$$

Hence, the third harmonic strain amplitude for an excitation near the resonance frequency ω_m of mode m yields

$$\varepsilon_{3m} = \frac{8Q}{5\pi L}\left|\int_0^L \alpha(x)\sin\left(\frac{3m\pi}{L}x\right)\left|\sin\left(\frac{m\pi}{L}x\right)\right| \times \sin\left(\frac{m\pi}{L}x\right)dx\right|\varepsilon_m^2. \tag{36b}$$

Similarly the fifth strain amplitude can be calculated and yields

$$\varepsilon_{5m} = \frac{8Q}{21\pi L}\left|\int_0^L \alpha(x)\sin\left(\frac{5m\pi}{L}x\right)\left|\sin\left(\frac{m\pi}{L}x\right)\right| \times \sin\left(\frac{m\pi}{L}x\right)dx\right|\varepsilon_m^2. \tag{37}$$

Both the third and fifth harmonics (and by extension all odd harmonics) are proportional to the second power of the fundamental strain amplitude, irrespective of the damage distribution. They all are generated by a second order interaction between two fundamental components, which come into play through the sign-function contribution and physically originate from the nonuniqueness of the stress-strain equation (or, equivalently, the discontinuity of the modulus at stress reversal points). The physical factors that show up in the leading expressions are thus identical for all odd harmonics: quadratic strain dependence ε_m^2 and efficiencies Q and $\bar{\alpha}$. The quadratic strain appearance in the case of hysteretic nonlinearity also makes it easier to observe the higher order harmonics in real experiments with state-of-the-art hardware (in contrast, for classical cubic nonlinearities, a fifth order strain dependence was noticed for the fifth harmonic, requiring a quite large dynamic range).

From Eqs. (36a) and (37), one can easily obtain the expressions for the two special cases of a localized damage zone and a globally distributed hysteretic nonlinearity. We obtain respectively

$$\varepsilon_{3m} = \frac{8Q}{5\pi}\frac{\hat{\alpha}d}{L}\left|\sin\left(\frac{3m\pi}{L}x_d\right)\right|\sin^2\left(\frac{m\pi}{L}x_d\right)\varepsilon_m^2 = X_{3,m}\varepsilon_m^2, \tag{38}$$

$$\varepsilon_{5m} = \frac{8Q}{21\pi}\frac{\hat{\alpha}d}{L}\left|\sin\left(\frac{5m\pi}{L}x_d\right)\right|\sin^2\left(\frac{m\pi}{L}x_d\right)\varepsilon_m^2 = X_{5,m}\varepsilon_m^2, \tag{39}$$

for localized hysteresis nonlinearity, and

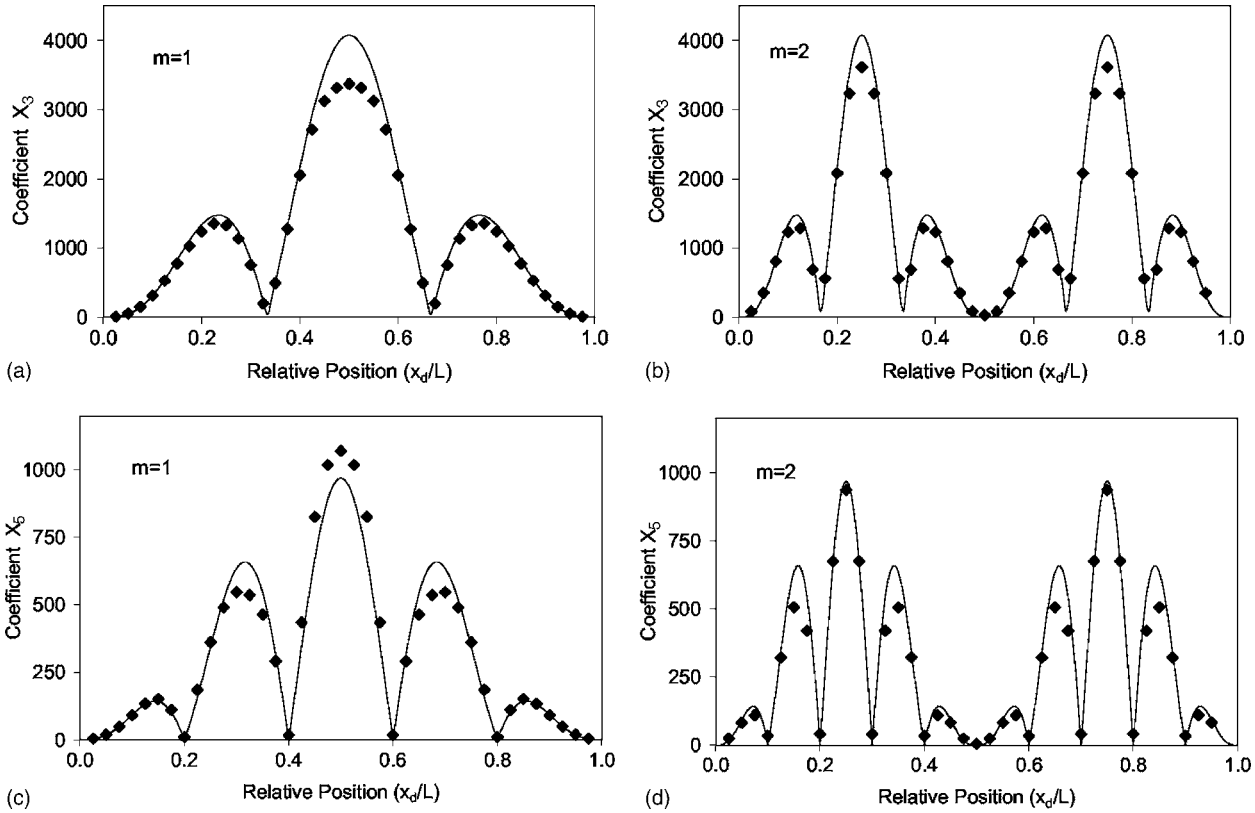


FIG. 7. Investigation of the nonlinear signature $X_{3,m}$ and $X_{5,m}$ (proportionality coefficient in the strain amplitude dependence of the third and fifth harmonics) from a hysteretic nonlinear defect as a function of its position in a resonant bar. (a) Signature $X_{3,1}$ for the fundamental mode excitation ($m=1$). (b) Signature $X_{3,2}$ for the second bar mode excitation ($m=2$). (c) Signature $X_{5,1}$ for the fundamental mode excitation ($m=1$). (d) Signature $X_{5,2}$ for the second bar mode excitation ($m=2$). Lines represent the analytical formulas; diamonds correspond to the numerical simulations. The defect zone is $d=L/20$.

$$\varepsilon_{3m} = \frac{32Q}{75\pi^2}\bar{\alpha}\varepsilon_m^2, \tag{40}$$

$$\varepsilon_{5m} = \frac{32Q}{2205\pi^2}\bar{\alpha}\varepsilon_m^2, \tag{41}$$

for global hysteresis nonlinearity.

Again, we have checked the particular position dependent behavior of the coefficients $X_{3,m}$ and $X_{5,m}$ in Eqs. (38) and (39) in the case of a localized damage zone with hysteretic nonlinearity by comparing it to the results of the multiscale model (Van Den Abeele *et al.*, 2004). Figure 7 illustrates the good quantitative agreement between the analytical and numerical results for the proportionality coefficients of the third and fifth harmonics taken at the resonance frequency of the fundamental of mode $m=1$ and mode $m=2$.

Finally, in Fig. 8, we illustrate the width (W) dependence of the damaged zone on the third and fifth harmonic proportionality coefficients [Eqs. (38) and (39)] for excitation near mode $m=1$. For the third harmonic the results are satisfactory. The agreement for the fifth harmonic may improve when using a finer discretization in the numerical model.

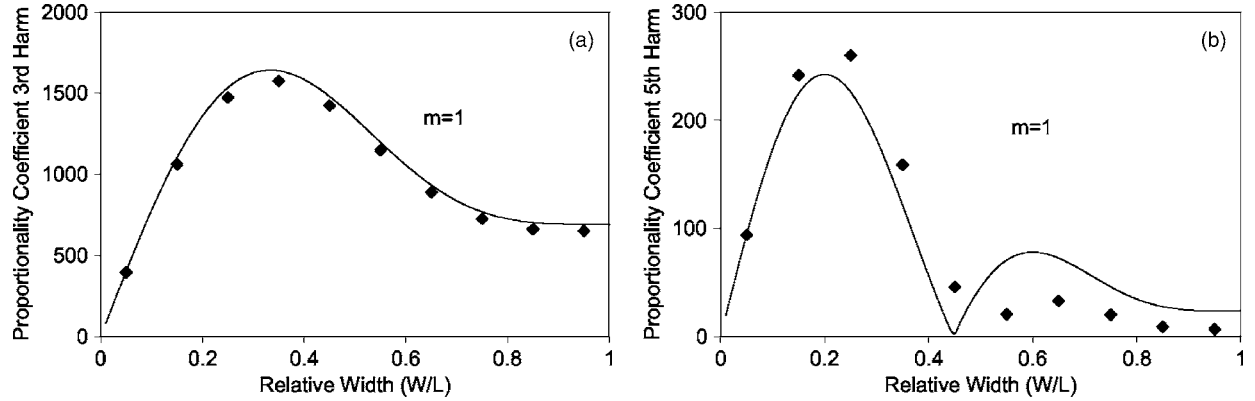


FIG. 8. Investigation of the proportionality coefficient in the strain amplitude dependence of the third and fifth harmonics from a hysteretic nonlinear defect as a function of its width (W). The defect is located in the middle of the bar ($x_d=L/2$). (a) Third harmonic efficiency for the fundamental mode excitation ($m=1$). (b) Fifth harmonic efficiency for the fundamental mode excitation ($m=1$). Lines represent the analytical formulas; diamonds correspond to the numerical simulations.

V. INVERSE MODELING

As the NRUS predictions of frequency shifts and harmonic content are analytical in nature, they offer easy opportunities to solve the inverse problem of defect characterization and localization. Traditionally, scanning techniques are used to localize damage in a material. Zones with defects are spotted by analyzing the changing properties appearing in the reflected or transmitted signals. In most cases the scanning requires a mechanical operation and a coupling with the medium. A consistent coupling is indeed highly critical when nonlinear signatures are sought.

The degree to which nonlinearity can be observed primarily depends on the value of the strain at the location of the nonlinearity. If the strain values are low, the nonlinearity will not be activated. By using multiple mode information and combining harmonic and frequency shift observations, it is possible to interpret those pieces of information in a consistent manner and to invert it for the damage characteristics and location. The fact that global resonances can be used for the inversion is an important advantage over traditional techniques. Unlike other techniques, the NRUS method does not require a spatial scan to locate the defect, as it lets different excitation modes, with different vibration patterns, probe the structure.

In the next paragraphs we give some possibilities of how to combine several of the above deduced formulas to infer the location of the defect, its degree of nonlinearity, and its width, based on the obtained nonlinear signatures of the resonances. We make again a distinction between a classical nonlinear defect for which the resonance frequency shift is quadratic, and a hysteresis dominated defect for which the resonance frequency reduction is basically linear.

A. Classical cubic nonlinearity

If the medium contains a localized defect that can be represented by a zone with classical nonlinearity, we can infer the location and its effective nonlinear strength (average over the bar) by simply considering the lowest two resonance modes. In both cases the resonance frequency shift is given by Eq. (13c) for $m=1$ and $m=2$. After some algebraic calculations, we obtain an expression for the relative location of the defect:

$$\frac{x_d}{L}=\frac{1}{\pi}A\cos\left[\frac{1}{2}\sqrt[4]{\frac{C_{1,2}}{C_{1,1}}}\right]=\frac{1}{\pi}A\cos\left[\frac{1}{2}\sqrt[4]{\frac{\varepsilon_1^2\Delta\omega_2/\omega_2}{\varepsilon_2^2\Delta\omega_1/\omega_1}}\right]. \tag{42}$$

It is clear, however, that the inversion procedure by Eq. (42) is nonunique as defects located symmetrically with respect to the center of the bar cannot be distinguished from each other.

For what concerns the degree of the defect, we find that

$$\bar{\delta}=\hat{\delta}\frac{d}{L}=\frac{64C_{1,1}^2}{3(4\sqrt{C_{1,1}}-\sqrt{C_{1,2}})^2}=\frac{64(\Delta\omega_1/\omega_1/\varepsilon_1^2)^2}{3(4\sqrt{\Delta\omega_1/\omega_1/\varepsilon_1^2}-\sqrt{\Delta\omega_2/\omega_2/\varepsilon_2^2})^2}. \tag{43}$$

This expression confirms that the relative width of the defect and its strength cannot be uncoupled for localized defects, and hence $\bar{\delta}=\hat{\delta}d/L$ must indeed be considered as the effective damage parameter characterizing the defect zone.

The error on the determination of the defect location and of the effective strength of the defect, induced by possible measurement errors on the experimental values of the resonance frequencies and the strain values (using the single mode NRUS technique applied to the first two modes) and reflecting in errors on the proportionality coefficients $C_{1,1}$ and $C_{1,2}$, can be calculated to first approximation in the following manner:

$$\frac{\Delta x_d}{L}=\frac{1}{4\pi}\frac{\sqrt[4]{C_{1,2}/C_{1,1}}}{\sqrt{4-\sqrt{C_{1,2}/C_{1,1}}}}\left(\frac{\Delta C_{1,1}}{C_{1,1}}-\frac{\Delta C_{1,2}}{C_{1,2}}\right), \tag{44a}$$

$$\Delta\bar{\delta}=\frac{128C_{1,1}^2(2\sqrt{C_{1,1}}-\sqrt{C_{1,2}})}{3(4\sqrt{C_{1,1}}-\sqrt{C_{1,2}})^3}\frac{\Delta C_{1,1}}{C_{1,1}}+\frac{64C_{1,1}^2\sqrt{C_{1,2}}}{3(4\sqrt{C_{1,1}}-\sqrt{C_{1,2}})^3}\frac{\Delta C_{1,2}}{C_{1,2}}. \tag{44b}$$

Here, Δx_d is the absolute error on the defect location, $\Delta\bar{\delta}$ is the absolute error on the defect strength, and $\Delta C_{1,m}$ are the absolute errors on the proportionality coefficients for the resonance frequency shifts obtained for mode m, with $m=1$ or 2. One can easily verify that a relative error of 10% on the value of $C_{1,1}$ and $C_{1,2}$ yields an absolute maximum error of about 0.1 on the determination of the relative defect position x_d/L when the defect is located near the edges. For defects away from the edges the uncertainty decreases significantly to below 0.02. Analogously, the relative error on the strength is for defects away from the edges limited to 10%.

A parallel inverse calculation can be obtained from the knowledge of the frequency shift and the third harmonic. For instance, combining Eqs. (13c) and (19) yield

$$\frac{x_d}{L}=\frac{1}{\pi}A\sin\left[\sqrt{\frac{3}{4}-\frac{C_{3,1}}{8QC_{1,1}}}\right]=\frac{1}{\pi}A\sin\left[\sqrt{\frac{3}{4}-\frac{\varepsilon_3}{8\varepsilon_1Q\Delta\omega_1/\omega_1}}\right]; \tag{45a}$$

$$\bar{\delta}=\hat{\delta}\frac{d}{L}=\frac{4C_{1,1}}{3\left(\frac{3}{4}-C_{3,1}/8QC_{1,1}\right)^2}=\frac{4(\Delta\omega_1/\omega_1/\varepsilon_1^2)}{3\left(\frac{3}{4}-\varepsilon_3/8\varepsilon_1Q\Delta\omega_1/\omega_1\right)^2}. \tag{45b}$$

For nonlocalized defects (spread over a finite width comparable to $L/2$, for instance), it is necessary to use the general equations given in Eqs. (12) and (18) for the inversion. An analogous error prediction can be performed as was done in Eq. (44).

B. Nonclassical hysteretic nonlinearity

The above outlined inversion and error estimation on the defect characteristics can also be performed in cases when the defect is represented by a hysteretic nonlinearity. Using Eq. (31) for two modes ($m=1,2$) we get

$$\frac{x_d}{L}=\frac{1}{\pi}A\cos\left[\frac{1}{2}\sqrt[3]{\frac{X_{1,2}}{X_{1,1}}}\right]; \tag{46a}$$

$$\bar{\alpha}=\hat{\alpha}\frac{d}{L}=\frac{X_{1,1}}{\left[1-\frac{1}{4}(X_{1,2}/X_{1,1})^{2/3}\right]^{3/2}(1+4/3\pi Q)}. \tag{46b}$$

On the other hand, if we use the strain dependence of the resonance frequency shift and of the third harmonic at the fundamental mode [Eqs. (31) and (38) for $m=1$], we find that

$$\frac{x_d}{L}=\frac{1}{\pi}A\sin\left[\sqrt{\frac{3}{4}-\frac{5\pi X_{3,1}(1+4/3\pi Q)}{32QX_{1,1}}}\right]; \tag{47a}$$

$$\bar{\alpha}=\hat{\alpha}\frac{d}{L}=\frac{X_{1,1}}{\left[\frac{3}{4}-5\pi X_{3,1}(1+4/3\pi Q)/32QX_{1,1}\right]^{3/2}(1+4/3\pi Q)}. \tag{47b}$$

Again it is necessary to use the more general equations and more sophisticated optimization algorithms for the inversion in the case of nonlocalized defects.

C. Global multi-mode NRUS method

Just like in the case of linear RUS, the determination of the material constants is more robust if one takes into account increasingly more information about higher resonances (Visscher *et al.*, 1991; Maynard, 1996; Migliori and Sarrao, 1997).

From our above analytical calculations, the resonance frequency shift of any particular mode due to a localized damage feature (either classical or nonclassical) contains terms that are linear and quadratic in the mode's maximal strain value at resonance. Thus, in general, we will always find a representation such as

$$\Omega_{res,\mathrm{Local}}(A_m)\approx\Omega_{res}(0)[1-\Upsilon_m\varepsilon_m-\Theta_m\varepsilon_m^2]. \tag{48}$$

The coefficients Υ_m and Θ_m determine the amplitude dependent resonance frequency proportionalities for each resonance mode. They highly depend on the position of the defect, its extent, and its degree of nonlinearity. As mentioned before, the macroscopic observation of nonlinearity is visible only when the strain field has a significant amplitude at the damage location. This is true for both kinds of nonlinearity. The multiplication of each of these coefficients with a field that represents the "strain energy" of the corresponding mode can therefore be a good indicator for a position dependent sensitivity. Based on this idea, we suggest a general procedure for defect localization in a free bar, using longitudinal resonances: first we determine the first and second proportionality coefficients Υ_m and Θ_m for the first M modes; then we evaluate the sensitivity function

$$S_M(x)=\frac{2}{M}\sum_{m=1}^{M}[|\Upsilon_m|+\sqrt{|\Theta_m|}]\sin^2\left(\frac{m\pi}{L}x\right); \tag{49}$$

and finally we identify the maxima in the sensitivity function.

Gliozzi *et al.* (2006) recently suggested this type of function for demonstrating a defect analysis procedure based on numerical data. They admitted that the form of this function was chosen arbitrary and did not explain its origin. Here, the choice is linked to the strain energy fields of the different modes, which are proportional to $\sin^2((m\pi/L)x)$ when dealing with a free bar. Moreover, from a mathematical point of view, the above proposed sensitivity function also makes sense for globally distributed damage. Indeed, Eqs. (14) and (32), for classical and hysteretic nonlinearities, respectively, show that there is no mode dependency in the coefficients Υ_m and Θ_m when the damage is uniformly distributed. Since

$$1=\lim_{m\to\infty}\frac{2}{M}\sum_{m=1}^{M}\sin^2\left(\frac{m\pi}{L}x\right)=\lim_{m\to\infty}\frac{1}{M}\sum_{m=1}^{M}\left[1-\cos\left(2\frac{m\pi}{L}x\right)\right], \tag{50}$$

we necessarily obtain a sensitivity function that is independent of the position. The constant value is an indication of the degree of nonlinearity:

$$\lim_{M\to\infty}S_M(x)=|\Upsilon_m|+\sqrt{|\Theta_m|}=\frac{4}{3\pi}\bar{\alpha}\left(1+\frac{4}{3\pi Q}\right)+3\sqrt{\frac{|\bar{\delta}|}{32}}. \tag{51}$$

As an example of the multi-mode NRUS procedure (i.e., the simultaneous handling of the nonlinear signatures from different modes) we consider a 250-mm-long bar with a localized defect simulated by a hysteresis nonlinearity ($\hat{\alpha}=2000$, $d=L/50$) situated at 72 mm. Figure 9(a) shows the sensitivity coefficient Υ_m for the first 25 longitudinal resonance modes according to analytical expression given in Eq. (32). Figure 9(a) can be viewed as a "nonlinearity impact spectrum." We clearly observe that certain modes, such as mode 3 and 4 and some higher order modes, are practically insensitive to defect. Figure 9(b) illustrates the spatial behavior of the sensitivity function along the bar position, plotted for a progressively increasing maximum number of modes. The darker colors indicate the inferred "defect" zone. It is worth noting that by increasing the number of resonances, the localization of the defect zone (in terms of the width) becomes more and more precise. Of cource, in practice the values of Υ_m should come from available experimental measurements using single mode NRUS experiments at a variety of modes, and in particular attenuation properties may pose limitations on the number of useful resonances.

The above simulations show that the multi-mode NRUS procedure is extremely efficient in the determination of the defect position, except for the symmetry problem. Indeed, the symmetry in $S_M(x)$ reflects the symmetry of the modes with respect to the center of the specimen. This can only be

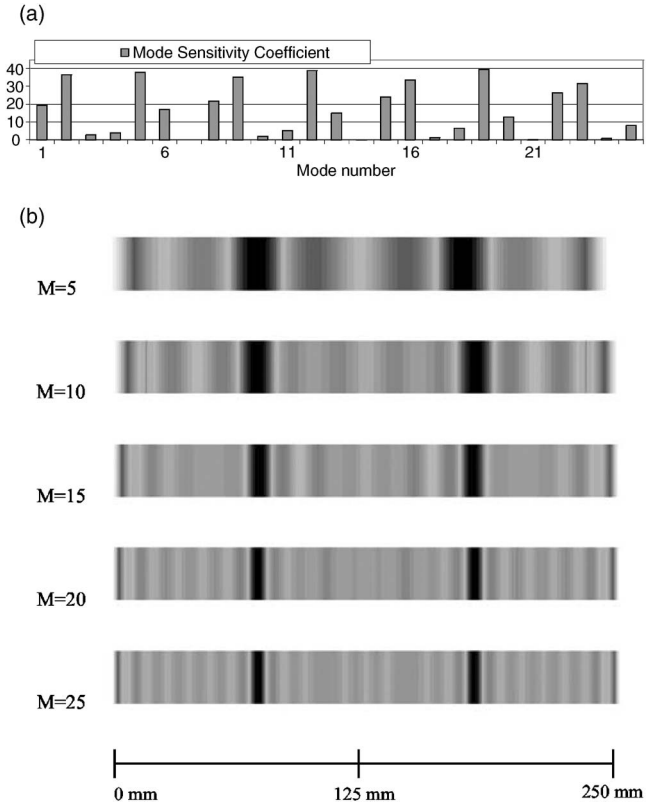


FIG. 9. Multi mode NRUS reconstruction for a localized hysteresis nonlinearity ($\hat{\alpha}=2000$, $d=L/50$) situated at 72 mm in a free bar ($L=250$ mm). (a) Sensitivity coefficient Υ_m as a function of the mode number. (b) Colorplot of the sensitivity functions for an increasing maximum mode taken into account. The darker colors indicate the inferred "defect" zone.

broken by changing the experimental configuration, e.g., by keeping one boundary of the specimen fixed.

In Fig. 10, we illustrate the outcome of multi-mode NRUS for two defects of different types: one corresponding to a hysteresis nonlinearity located at 72 mm ($d=L/50$) and one corresponding to a classical cubic nonlinearity at 45 mm ($d=L/100$). The sensitivity coefficients for the first 25 modes corresponding to the analytical expressions in Eqs. (32) and (14) are plotted in Figs. 10(a) and 10(b). There is a significant difference in the nonlinearity impact spectrum for the first Υ_m and second order $\sqrt{|\Theta_m|}$ coefficients. The sensitivity function for each separate order and for the combination of both orders clearly reveals the defect locations. In general, the procedure also works for a larger number of defects, even though it is obvious that only the largest ones will be visible in the sensitivity function.

By changing the particular mode shape for a longitudinal resonance in a free bar $[\partial_x \psi_n \propto \sin(n(\pi/L)x)]$ in Eq. (49) by its appropriate normal mode representation for clamped or simply supported cases, or for flexural modes, it becomes quite easy to extend this method. The case of two- and three-dimensional objects can possibly be handled by extending the sensitivity function to higher dimensions using the proper "strain energy" fields of the modes in 2D and 3D. We are currently checking this assumption.

It can be easily shown that the induced error on the sensitivity function due to errors in the proportionality coefficients remains limited. Let us consider the case in which Υ_m alone is measured. Each of the Υ_m comes with absolute error, E_m. The weight of this error in the induced absolute error on the sensitivity function is $\Delta_{(m)}S(x) = (2/M)\sin^2(m(\pi/L)x)E_m$, which is smaller than E_m in any case. Therefore, one can assume that the error on the determination of the maximum of the sensitivity function due to errors on the measured coefficients remains small, too. We do not include a thorough or complete error investigation in this paper since at the moment this would lead us too far. The full investigation will, however, be performed when we report on the experimental verification of the multi-mode NRUS method in the near future.

As an alternative for the sensitivity function, one could consider to do the multi-mode NRUS analysis by a least squares method. Suppose that one has measurements of the resonant frequency at increasing strain amplitudes for M modes, yielding M values of Υ_m in Eq. (49). We shall omit here for simplicity the values of $\sqrt{|\Theta_m|}$. As mentioned before, each of the Υ_m coefficients comes with error, E_m. In addition we assume that the damage is concentrated at a single location. As the amplitude of the nonlinearity in this case is not relevant to finding its geometry, we can form an M-component unit vector from the Υ_m, which we call Υ_m^*.

Next, we make a theoretical prediction T_m of Υ_m, based on the general expression for the expected strain dependence of the resonance frequency shift as given in Eq. (30):

$$T_m = \int_0^L G_{\text{Local}}(x;x_d,d)\left|\sin^3\left(m\pi\frac{x}{L}\right)\right| dx \tag{52}$$

with G the form function for a local defect at $x=x_d$ and with width d, modeled as in Eq. (13c).

Now, we form the M-component unit vector from the predicted values T_m, say T_m^*, and we compare the Υ_m^* and T_m^* vectors with a weighting related to the confidence in the experimental values Υ_m,

$$E(x_d,d) = \sum_{m=1}^{M} \frac{(\Upsilon_m^* - T_m^*)^2}{E_m}. \tag{53}$$

$E(x_d,d)$ is a two-dimensional energy landscape. To find the location and the width, we need to search for its minimum. In addition, the sensitivity can be measured as the curvature at the minimum.

VI. CONCLUSIONS

An analytical treatment of the effect of damage on the resonance mode characteristics is not simple, and probably only possible under severe assumptions. Nevertheless, if analytical formulas are available they can be of great use to (1) check the correctness of the simulations performed by numerical models in limiting cases and (2) provide quick methods for the inversion of the damage characteristics.

We have presented a nonlinear version of the resonance ultrasound spectroscopy (RUS) theory by extending the traditional formalism to the treatment of damage in the form of nonlinearity. General equations were developed in the 1-D case, describing the interaction between the modes induced by the nonlinearity. These equations were solved following a

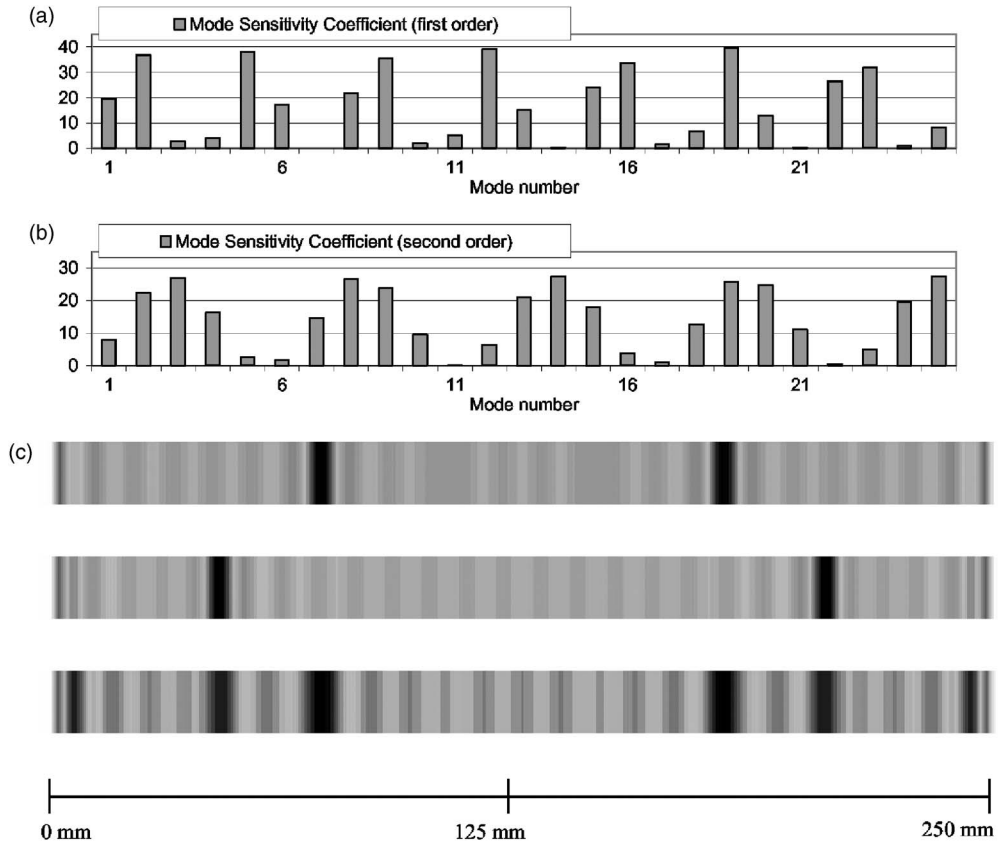


FIG. 10. Multi mode NRUS reconstruction for a free bar (L=250 mm) containing two localized defect spots: a hysteresis nonlinearity ($\hat{\alpha}$=2000, $d=L/50$) situated at 72 mm and a classical cubic nonlinearity ($\hat{\delta}=-10^5$, $d=L/100$) situated at 45 mm. (a) Sensitivity coefficient Υ_m (first order) as a function of the mode number. (b) Sensitivity coefficient $\sqrt{|\Theta_m|}$ (second order) as a function of the mode number. (c) Colorplot of the sensitivity functions for the first 25 modes. From top to bottom: reconstruction based on the first order coefficients only, on the second order coefficients only, and on the combination of first and second order coefficients. The darker colors indicate the inferred "defect" zones.

perturbation approach. We considered two different nonlinear stress-strain signatures that represent the damage: a reversible cubic nonlinearity as well as the more interesting hysteretic nonlinearity.

The solutions provide analytical expressions for the nonlinear shift of the modal resonance frequency and the harmonic generation as a function of the strength of the nonlinearity, the width of the defect, and its position in the 1D system. The physical factors that appear in the analytical formula can be readily explained in terms of the nature and the vibration patterns of the modes that produce the particular nonlinear interaction. The dependence on the strain level is different for both cases of nonlinearity, but always independent on the defect location.

We compared the results to numerical calculations from a multiscale model, showing excellent agreement, provided the discretization is defined adequately. In addition, the constant ratio of the absolute increase in Q^{-1} to the relative frequency shift predicted by the analytical formulas in the case of hysteresis nonlinearity is confirmed by numerical simulations and is of the same order as found in experiments.

Finally, the analytical formulas were exploited to infer critical information about the damage position, the degree of nonlinearity, and the width of the damage zone either from the shifts in resonance frequency occurring at different excitation modes, or from the shift and the harmonics predicted at a single mode. A general defect localization procedure has been put forward based on the introduction of a nonlinearity sensitivity function. This technique can be extended to vibration modes other than longitudinal, and possibly also to higher dimensions. As an alternative, a least squares method for treating experimental data has been suggested. The use of multiple mode information for the inversion of damage characteristics and localization has an important advantage over traditional scanning techniques. Unlike other techniques, the NRUS method does not require a spatial scan to locate the defect, as it lets different excitation modes, with different vibration patterns, probe the structure.

ACKNOWLEDGMENTS

The author gratefully acknowledges the support of the Flemish Fund for Scientific Research (G.0206.02 and G.0257.02), the provisions of the European Science Foundation Program NATEMIS, and the European FP5 and FP6 Grants DIAS (EVK4-CT-2002-00080) and AERONEWS (AST3-CT-2003-502927). The author also appreciated the reviewers' comments and suggestions to improve the reported work.

Aleshin, V., Gusev, V., and Zaitsev, V. Y. (**2004**). "Propagation of acoustics waves of nonsimplex form in a material with hysteretic quadratic nonlinearity: Analysis and numerical simulations," J. Comput. Acoust. **12**(3),

319–354.
Antonets, V. A., Donskoy, D. M., Sutin, A. M. (**1986**). "Nonlinear vibrodiagnostics of flaws in multilayered structures," Mech. Compos. Mater. **15**, 934–937.
Ballad, E., Vezirov, S., Pfleiderer, K., Solodov, I., and Busse, G. (**2004**). "Nonlinear modulation technique for NDE with air-coupled ultrasound," Ultrasonics **42**(1–9), 1031–1036.
Cantrell, J. H., and Yost, W. T. (**1994**). "Acoustic harmonic-generation from fatigue-induced dislocation dipoles," Philos. Mag. A **69**, 315–326.
Capogrosso-Sansone, B., and Guyer, R. A. (**2002**). "Dynamic model of hysteretic elastic systems," Phys. Rev. B **66**, 224101.
Delsanto, P. P., Gliozzi, A. S., Hiresekorn, M., and Nobili, M. (**2006**). "A 2D spreing model for the simulation of ultrasonic wave propagation in nonlinear hysteretic media," Ultrasonics **44**, 279–286.
Gliozzi, A. S., Nobili, M., and Scalerandi, M. (**2006**). "Modelling localized nonlinear damage and analysis of its influence on resonance frequencies," J. Phys. D **39**(17), 3985–3903.
Gusev, V. (**2000**). "Propagation of acoustic pulses in material with hysteretic nonlinearity," J. Acoust. Soc. Am. **107**(6), 3047–3058.
Guyer, R. A., McCall, K. R., and Van Den Abeele, K. (**1998**). "Slow elastic dynamics in a resonant bar of rock," Geophys. Res. Lett. **25**(10), 1585–1588.
Ichitsubo, T., Ogi, H., Hirao, M., Tanaka, K., Osawa, M., Yokokawa, T., Kobayashi, T., and Harada, H. (**2002**). "Elastic constant measurement of Ni-base superalloy with the RUS and mode selective EMAR methods," Ultrasonics **40**, 211–215.
Johnson, P. A., and Sutin, A. M. (**2005**) "Slow dynamics and anomalous nonlinear fast dynamics in diverse solids," J. Acoust. Soc. Am. **117**(1), 124–130.
Johnson, P. A., Zinszner, B., and Rasolofosaon, P. N. J. (**1996**). "Resonance and elastic nonlinear phenomena in rock," J. Geophys. Res. **101**, 11553–11564.
Kac, M. (**1966**). "Can one hear the shape of a drum?" Am. Math. Monthly **73**, 1–23.
Kazakov, V. V., Sutin, A. M., and Johnson, P. A. (**2002**). "Sensitive imaging of an elastic nonlinear wave-scattering source in a solid," Appl. Phys. Lett. **81**(4), 646–648.
Keppens, V., Mandrus, D., Sales, B. C., Chakoumakos, B. C., Dai, P., Coldea, R., Maple, M. B., Gajewski, D. A., Freeman, E. J., and Bennington, S. (**1998**). "Localized vibrational modes in metallic solids," Nature (London) **395**(6705), 876–878.
Krohn, N., Stoessel, R., and Busse, G. (**2002**). "Acoustic non-linearity for defect selective imaging," Ultrasonics **40**(1–8), 633–637.
Landau, L. E., and Lifshitz, E. (**1969**). *Mechanics* (Oxford, Pergamon).
Leonard, K. R., Malyarenko, E. V., and Hinders, M. K. (**2002**). "Ultrasonic lamb wave tomography," Inverse Probl. **18**, pp. 1795–1808.
Maynard, J. (**1996**). "Resonant ultrasound spectroscopy," Phys. Today **49**, 26–31.
McCall, K. R. (**1994**). "Theoretical-study of nonlinear elastic-wave propagation," J. Geophys. Res., [Solid Earth] **99**(B2), 2591–2600.
McCall, K. R., and Guyer, R. A. (**1994**). "Equation of state and wave-propagation in hysteretic nonlinear elastic-materials," J. Geophys. Res., [Solid Earth] **99**(B12), 23887–23897.
McCall, K. R., and Guyer, R. A. (**1996**). "A new theoretical paradigm to describe hysteresis, discrete memory and nonlinear elastic wave propagation in rock," Nonlinear Processes Geophys. **3**, 89–101.
Michaels, J. E., and Michaels, T. E. (**2006**). "Enhanced differential methods for guided wave phased array imaging using spatially distributed piezoelectric transducers," Rev. Prog. Quant. Nondestr. Eval. **25A**, 837–844.
Migliori, A., and Sarrao, J. L. (**1997**) *Resonant Ultrasound Spectroscopy: Applications to Physics, Materials Measurements, and Non-destructive Evaluation* (Wiley, New York).
Morris, W. L., Buck, O., and Inman, R. V. (**1979**). "Acoustic harmonic-generation due to fatigue damage in high-strength aluminum," J. Appl. Phys. **50**(11), 6737–6741.
Nagy, P. B. (**1998**). "Fatigue damage assessment by nonlinear ultrasonic materials characterization," Ultrasonics **36**(1–5), 375–381.
Nakamura, N., Ogi, H., and Hirao, M. (**2004**). "Resonance ultrasound spectroscopy with laser-Doppler interferometry for studying elastic properties of thin films," Ultrasonics **42**(1–9), 491–494.
Nayfeh, A. (**1973**). *Perturbation Methods* (Wiley, New York).
Nazarov, V. E., Ostrovsky, L. A., Soustova, I., and Sutin, A. M. (**1988**). "Nonlinear acoustics of microinhomogeneous media," Phys. Earth Planet. Inter. **50**(1), 65–73.
Nazarov, V. E., and Sutin, A. M. (**1989**). "Theory of parametric sound receiver using a nonlinear layer," Sov. Phys. Acoust. **35**, 510–512.
Ogi, H., Sato, K., Asada, T., and Hirao, M. (**2002**). "Complete mode identification for resonance ultrasound spectroscopy," J. Acoust. Soc. Am. **112**(6), 2553–2557.
Pohit, G., Mallik, A., and Venkatesan, C. (**1999**). "Free out-of-plane vibrations of a rotating beam with non-linear elastomeric constraints," J. Sound Vib. **220**(1), 1–25.
Scalerandi, M., Agostini, V., Delsanto, P. P., Van Den Abeele, K., and Johnson, P. A. (**2003**). "Local interaction simulation approach to modelling nonclassical, nonlinear elastic behavior in solids," J. Acoust. Soc. Am. **113**(6), 3049–3059.
Stoessel, R., Krohn, N., Pfleiderer, K., and Busse, G. (**2002**). "Air-coupled ultrasound inspection of various materials," Ultrasonics **40**(1–8), 159–163.
Sutin, A. M., and Donskoy, D. M. (**1998**). "Vibro-acoustic modulation nondestructive evaluation technique," Proc. SPIE **3397**, 226–237.
Teklu, A., Ledbetter, H., Kim, S., Boatner, L. A., Keppens, V., and McGuire, M. (**2004**). "Single crystal elastic constants of 70Fe-15Ni-15Cr alloy," Metall. Mater. Trans. A **35**, 3149–3153.
TenCate, J. A., and Shankland, T. J. (**1996**). "Slow dynamics in the nonlinear elastic response of Berea sandstone," Geophys. Res. Lett. **23**(21), 3019–3022.
TenCate, J. A., Smith, E., and Guyer, R. A. (**2000**) "Universal slow dynamics in granular solids," Phys. Rev. Lett. **85**(5), 1020–1023.
Ulrich, T., McCall, K. R., and Guyer, R. A. (**2002**). "Determination of elastic moduli of rock samples using resonant ultrasound spectroscopy," J. Acoust. Soc. Am. **111**(4), 1667–1674.
Van Den Abeele, K., and Johnson, P. A. (**1996**). "Elastic pulsed wave propagation in media with second- or higher-order nonlinearity. II. Simulation of experimental measurements on Berea sandstone," J. Acoust. Soc. Am. **99**(6), 3346–3352.
Van Den Abeele, K., Johnson, P. A., Guyer, R. A., and McCall, K. R. (**1997**). "On the quasi-analytic treatment of hysteretic nonlinear response in elastic wave propagation," J. Acoust. Soc. Am. **101**(4), 1885–1898.
Van Den Abeele, K., Johnson, P. A., and Sutin, A. M. (**2000a**). "Nonlinear elastic wave spectroscopy (NEWS) techniques to discern material damage, part I: Nonlinear wave modulation spectroscopy (NWMS)," Res. Nondestruct. Eval. **12**(1), 17–30.
Van Den Abeele, K., Carmeliet, J., TenCate, J. A., and Johnson, P. A. (**2000b**). "Nonlinear Elastic Wave Spectroscopy (NEWS) techniques to discern material damage. Part II: Single Mode Nonlinear Resonant Acoustic Spectroscopy (SIMONRAS)," Res. Nondestruct. Eval. **12**(1), 31–42.
Van Den Abeele, K., and De Visscher, J. (**2000**). "Damage assessment in reinforced concrete using spectral and temporal nonlinear vibration techniques," Cem. Concr. Res. **30**(9), 1453–1464.
Van Den Abeele, K., Van De Velde, K., and Carmeliet, J. (**2001**). "Inferring the degradation of pultruded composites from dynamic nonlinear resonance measurements," Polym. Compos. **22**(4), 555–567.
Van Den Abeele, K., Schubert, F., Aleshin, V., Windels, F., and Carmeliet, J. (**2004**). "Resonant bar simulations in media with localized damage," Ultrasonics **42**, 1017–1024.
Visscher, W., Migliori, A., Bell, T., and Reinert, R. (**1991**). "On the normal modes of free vibration of inhomogeneous and anisotropic elastic objects," J. Acoust. Soc. Am. **90**(4), 2154–2162.
Zimenkov, S., and Nazarov, V. (**1995**). "Propagation of nonlinear acoustic waves in rocks," Izv., Acad. Sci., USSR, Phys. Solid Earth **30**(5), 437–439.

Plane wave source with minimal harmonic distortion for investigating nonlinear acoustic properties

Christopher W. Lloyd, Kirk D. Wallace,[a] Mark R. Holland, and James G. Miller[b]
Department of Physics, Washington University in Saint Louis, Saint Louis, Missouri, 63130

(Received 11 December 2006; revised 11 April 2007; accepted 22 April 2007)

The objective of this investigation is to introduce and validate a practical ultrasound source to be used in the investigation of the nonlinear material properties of liquids and soft tissues studied *in vitro*. Methods based on the progressive distortion of finite amplitude ultrasonic waves in the low megahertz frequency range are most easily implemented under the assumption of plane wave propagation. However, achieving an approximately planar ultrasonic field over substantial propagation distances can be challenging. Furthermore, undesired harmonic distortion of the ultrasonic field prior to insonification of the specified region of interest represents another serious limitation. This paper introduces an approach based on the use of the ultrasonic field emanating from a stainless-steel delay line. Both simulation and direct experimental measurement demonstrate that such a field exhibits relatively planar wave fronts to a good approximation (such that a 3-mm-diam receiver would be exposed to no more than 3 dB of loss across its face) and is free from the significant harmonic distortion that would occur in a conventional water path.
 [DOI: 10.1121/1.2739442]

PACS number(s): 43.25.Zx, 43.25.Ba [MFH] Pages: 91–96

I. INTRODUCTION

In previously published ultrasonic experiments designed to measure the intrinsic nonlinear parameter, β, interpretation of the measured data was frequently simplified by approximating the insonifying ultrasonic field as a plane wave over the region of interaction with the sample under investigation.[1–12] A number of methods to realize this approximation have been utilized and, in general, rely on careful positioning of the sample in a region of the ultrasonic field where the diffraction pattern can be approximated as planar, such as the extreme near field,[1,2] the natural focus,[3–5] or the far field.[6] Estimates of the nonlinear parameter can be determined from finite amplitude measurements of the level of harmonic distortion in the ultrasonic signal after propagation through either a series of thicknesses of a sample[1,2] or a single sample.[3–12]

One method that has been used successfully in our laboratory is to affix a stainless-steel delay line to a transmitting transducer such that the output field from the delay line satisfies a plane wave approximation.[3,4] An advantage with this approach is that the high sound velocity in the stainless steel (approximately four times faster than in water) permits the ultrasonic beam to satisfy the plane wave approximation in a relatively short distance, rather than after traveling a relatively long (possibly prohibitively long) path. Furthermore, this method reduces the generation of undesired harmonic distortion arising from propagation to the sample under investigation, as propagation within the stainless-steel delay line results in only a negligible amount of finite amplitude distortion, eliminating the need for physical attenuators in the propagation path.[6]

In the current investigation, scanned hydrophone pseudo-array measurements and numerical simulations were used to investigate the applicability of a plane wave approximation for an ultrasonic pulse emanating from a stainless-steel delay line and to assess the amount of harmonic distortion generated from propagation in the delay line. Hydrophone measurements of the ultrasonic field in water were conducted in a meridian plane and compared with the results from simulations of the diffracting finite amplitude field associated with a 12.7-mm-diam contact transducer propagating either a 7 or a 4 MHz tone burst through a 46.4 mm steel delay line and 25.0 mm of water.

II. EXPERIMENTAL METHODS

A. Experimental setup

The delay line used in this experiment was machined from an 88.9-mm-diam cylinder of type 304 stainless steel. The faces of the cylinder were machined flat and parallel, and polished smooth, producing a delay line of 46.4 mm in length as measured by Vernier calipers. A 12.7-mm-diam, broadband, 10 MHz center frequency contact transducer (Panametrics V111, Waltham, MA) was bonded to one face of the stainless-steel cylinder using a thin, uniform layer of vacuum grease and held in contact with a plastic (Delrin™) fixture, as illustrated in Fig. 1. This created a watertight seal and enabled the contact transducer-delay line assembly to be submerged into a water tank for hydrophone characterization of the transmitted ultrasonic field.

The first ultrasonic frequency (7 MHz) used in this investigation was chosen such that the ultrasonic beam's diffraction pattern would be entering the natural focus as the

[a]Current address: Washington University School of Medicine, Cardiovascular Division, Cortex Building, Suite 101, 4320 Forest Park Avenue, Saint Louis, Missouri 63108
[b]Author to whom correspondence should be addressed. Electronic mail: james.g.miller@wustl.edu

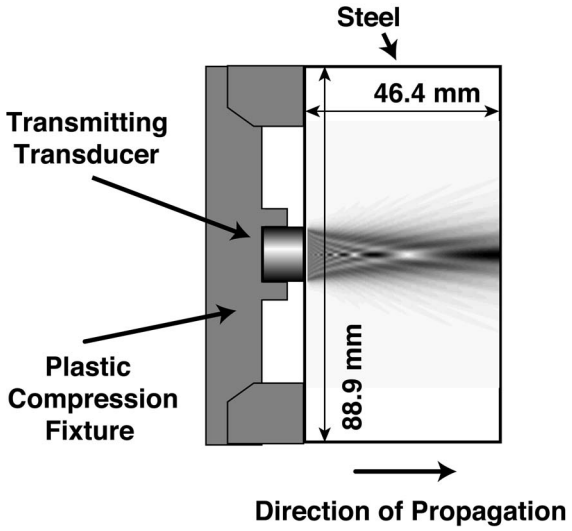


FIG. 1. Diagram of the stainless-steel delay line apparatus. The transducer is a 12.7-mm-diam contact transducer, held in place with a plastic fixture to the face of the steel cylinder. Simulation of the 7 MHz fundamental ($1f$) field component is shown inside the delay line.

pulse exited the delay line apparatus. This frequency was determined using the formula for the near field distance, given by

$$\text{Near Field Distance} = \frac{D^2}{4\lambda}\left[1 - \left(\frac{\lambda}{D}\right)^2\right], \tag{1}$$

where D is the diameter of the source transducer and $\lambda = c_0/f$ is the ultrasonic wavelength.[13] The speed of sound in the stainless steel, c_0, was experimentally determined to be 5.73 mm/μs using a contact transducer in a pulse echo mode to measure the round-trip time of flight inside the stainless-steel delay line. To position the natural focus of the ultrasonic field just after emerging from the stainless-steel delay line, as defined by Eq. (1), a center frequency of 7.0 MHz was chosen for the narrowband transmitted signal. For this study, the transmit transducer was driven with a 4.0 μs duration tone burst gated from a 7.0 MHz continuous-wave signal oscillator (Hewlett-Packard 606B, Palo Alto, CA) and amplified with a 50 dB amplifier (ENI 240L, Rochester, NY). The gating was accomplished with a pulse generator (Hewlett-Packard 8112A) in conjunction with two mixer/modulators in series (Hewlett-Packard 33125).

In addition, a lower frequency, 4 MHz, was chosen to test the applicability of this specific delay line for making measurements within the focal zone to far field transition regime of the ultrasonic field. In this case, a 12.7-mm-diam, broadband, 5 MHz center frequency contact transducer (Panametrics V109) was utilized. The transducer was driven by a 4.0 μs tone burst at 4.0 MHz from a function generator (Hewlett-Packard 8116A) in conjunction with a 50 dB radio frequency amplifier (ENI 240L) and a passive 6.4 MHz low pass filter.

B. Measurements

To characterize the ultrasonic pressure field of the system, the transducer and delay line were positioned in a water tank with the ultrasonic beam aligned along the X-axis of a motion control system (Aerotech Unidex 511, Pittsburgh, PA). The temperature of the water was monitored using a thermistor (Omega OL-710, Stamford, CT) connected to a thermometer (Omega 5831A). The temperature of the water tank was maintained at 23.8 °C±0.1 °C and 21.5 °C±0.5 °C during the measurements of the 7 and 4 MHz signals, respectively. A membrane hydrophone (Sonic Industries, Model 804, now Sonora Medical Systems, Longmont, CO) was scanned in a rectangular, 81×57 pseudo-array using a uniform step size of 0.25 mm (20 mm × 14 mm total grid size) to map out the ultrasonic field in the meridian (X-Y) plane. The hydrophone calibration was performed by the manufacturer using a reference hydrophone that had been calibrated at the National Physical Laboratory in the United Kingdom. The resulting radio frequency signals output from the hydrophone preamplifier were digitized with an 8 bit resolution oscilloscope (Tek 2440, Beaverton, OR) for the 7.0 MHz acquisitions, and a Tektronix 5052B for the 4.0 MHz acquisitions). Data could be acquired with the membrane hydrophone positioned as close as 4.0 mm to the face of the steel cylinder while avoiding reverberations between the steel cylinder and the hydrophone. The received wave forms were signal averaged in the time domain to improve the signal-to-noise ratio for data acquired for both the 7.0 and 4.0 MHz systems. The values of the acoustic pressure at the fundamental and harmonic frequencies were determined by determining the voltage amplitude of each spectral component of the signal averaged wave form and referring to the pressure-voltage calibration data provided by the hydrophone manufacturer.

III. SIMULATIONS

Simulations were carried out using a Burgers' equation enhanced angular spectrum approach.[14–16] This approach permits the modeling of the propagation and the harmonic distortion of a diffracting ultrasonic signal as it propagates through the different propagation media (steel and water). Simulations were performed for the two fundamental frequencies employed experimentally (4 and 7 MHz). Effects of attenuation were taken to be negligible for both the steel and water at these frequencies. The nonlinear parameter β was taken to be 3.5 for both water and stainless steel. (A literature search for similar metals yielded values for β ranging from 2 to 4.5 and thus the nonlinear parameter for stainless steel was taken to be 3.5 for this simulation.)[17]

The predictions from simulation for the normalized particle velocity amplitudes of the 7 MHz fundamental ($1f$) component are depicted in Fig. 2(a). A gray scale image of the magnitude for the meridian plane slice is shown in the top panel and the profile corresponding to the values of the field on the beam axis is shown in the bottom panel. As depicted in Fig. 2(c), the beam is propagating from left to right, first through the stainless-steel delay line and then

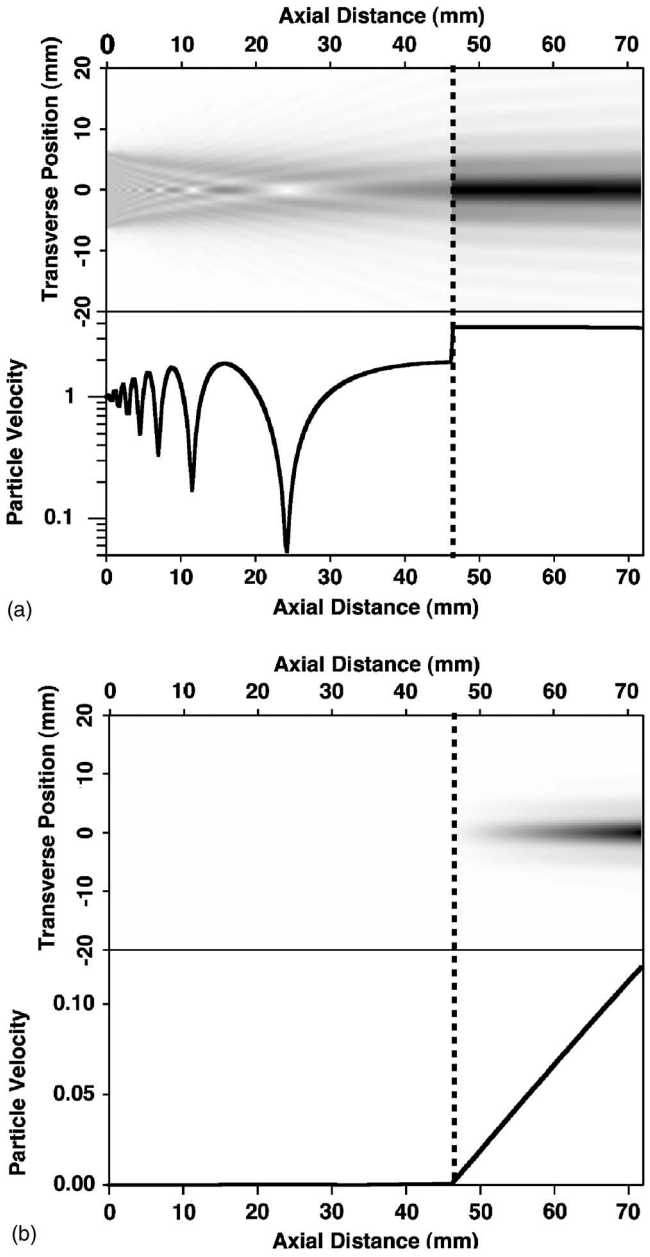


FIG. 2. Simulation of the diffraction pattern through a 304-stainless steel delay line into lossless water medium, for (a) the 7 MHz fundamental component, and (b) the 14 MHz second harmonic. The top portion of each figure component illustrates the magnitude of the meridian plane slice of the field (black is the maximum amplitude and white is zero). The bottom portion of each component shows the amplitude along the propagation axis of the beam. The phase velocity used in the simulation corresponds to measured values of 5.73 mm/μs in steel and 1.48 mm/μs in water. The particle velocity amplitudes are normalized to that of the fundamental component at zero axial distance.

through 25 mm of water. The vertical dashed line indicates the boundary between steel and water. Similarly, the nonlinearly generated 14 MHz second harmonic ($2f$) component of the particle velocity obtained from simulation is displayed in Fig. 2(b). In both cases, the pressure has been normalized to the initial fundamental pressure at the face of the transducer.

Figure 3 presents the corresponding results for the 4 MHz simulation. For both simulations the second harmonic amplitude along the propagation axis is observed to

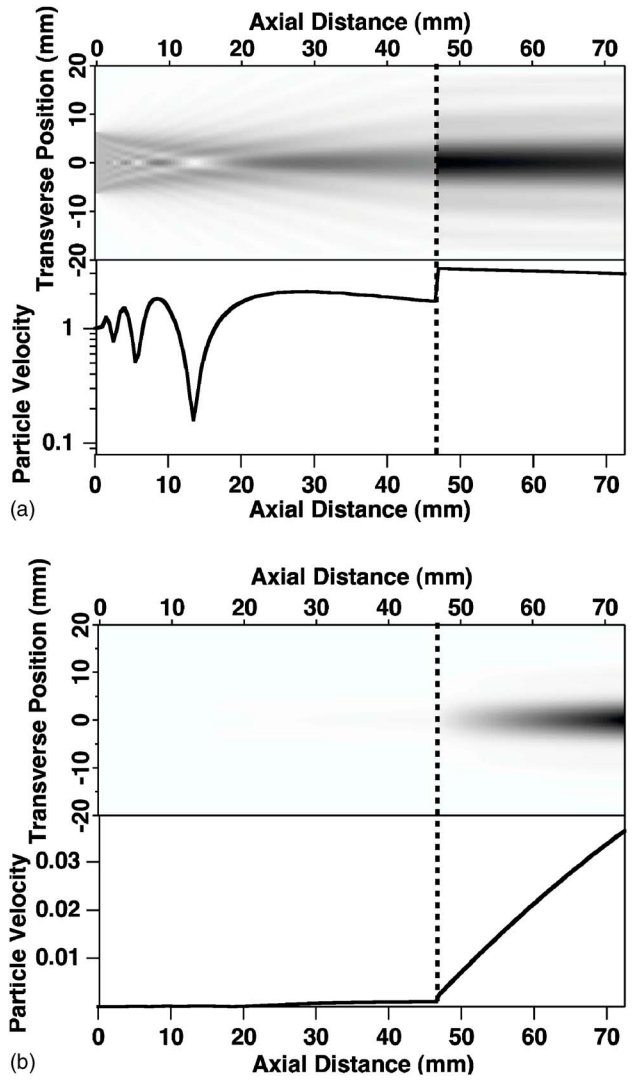


FIG. 3. (Color online) Simulation of the diffraction pattern through a 304-stainless steel delay line into lossless water medium, for (a) the 4 MHz fundamental component, and (b) the 8 MHz second harmonic, in a manner analogous to Fig. 2.

grow linearly as a function of propagation distance in the water region. This is consistent with the second order (weakly nonlinear) approximate form of the Fubini solution.[18]

IV. EXPERIMENTAL MEASUREMENTS AND RESULTS

A. 7 MHz study

Figure 4 shows the measured amplitudes of the 7 MHz fundamental ($1f$) and 14 MHz harmonic ($2f$) components in the meridian plane pseudo-array scan for the identical water-path region shown in the simulations presented in Figs. 2(a) and 2(b). The gray scale is normalized in each image such that black represents the maximum amplitude and white is zero. Because one goal of this work is to produce a valid approximation to a plane wave over an experimentally useful area, the size of a 3-mm-diameter receiving transducer is represented by hash marks on the right hand side of each of the four images in Fig. 4.

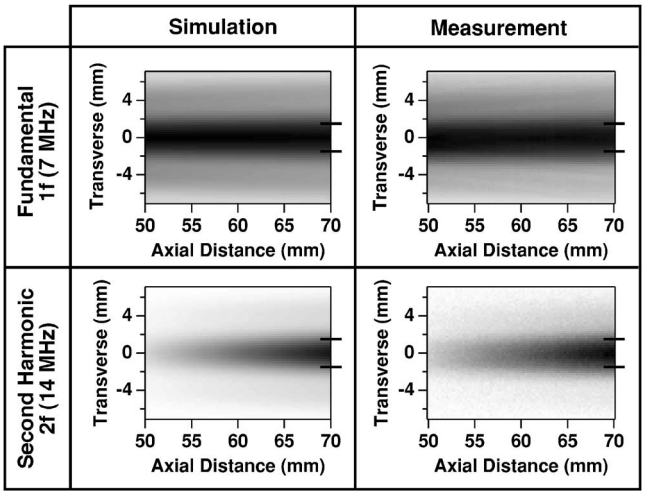


FIG. 4. Comparison of expected beam pattern from simulation (left) with hydrophone measurement of the field (right) in a meridian plane beginning 4 mm beyond the steel-water interface, for a 7 MHz fundamental signal. The separation between the hash marks on the right side of each graph corresponds to the 3 mm diameter of a receiving transducer.

Figure 5(a) illustrates the power along the axis of propagation as a function of distance for each of the four panels in Fig. 4. The curves represent the measured power relative to the fundamental at the left edge of the plot (axial distance of 50 mm). The measured pressure values for the fundamental and second harmonic at an axial distance of 70 mm were 53.2 and 2.2 kPa, respectively. Figure 5(b) illustrates the predicted and measured transverse profiles of the power in the fundamental frequency at an axial distance of 70 mm. Figure 5(c) shows this comparison for the second harmonic portion of the signal.

B. 4 MHz study

Figure 6 shows the measured amplitudes of the 4 MHz fundamental ($1f$) and the 8 MHz second harmonic ($2f$) components of the transmitted ultrasonic beam in the meridian plane. These are displayed along with the simulation results of Figs. 3(a) and 3(b). For each panel, the image is normalized so that black represents the maximum amplitude and white represents zero. Again, hash marks are present on the right side of each panel to indicate the receiving size of a 3-mm-diam transducer.

Figure 7 shows orthogonal one-dimensional slices of the measured and simulated ultrasonic beams detailed in Fig. 6. In Fig. 7(a), the power along the axis of propagation is shown as a function of axial distance for each of the four panels found in Fig. 6. Each curve is normalized to the measured power of the fundamental at the left edge of the plot, which is 50 mm from the source transducer. The values of the pressure for the fundamental and second harmonic were measured at a distance of 70 mm from the transducer, and were 23.6 kPa and 228 Pa, respectively. Figures 7(b) and 7(c) illustrate a comparison of the simulation with the measured results for a transverse profile taken at an axial distance of 70 mm for the fundamental and second harmonic, respectively.

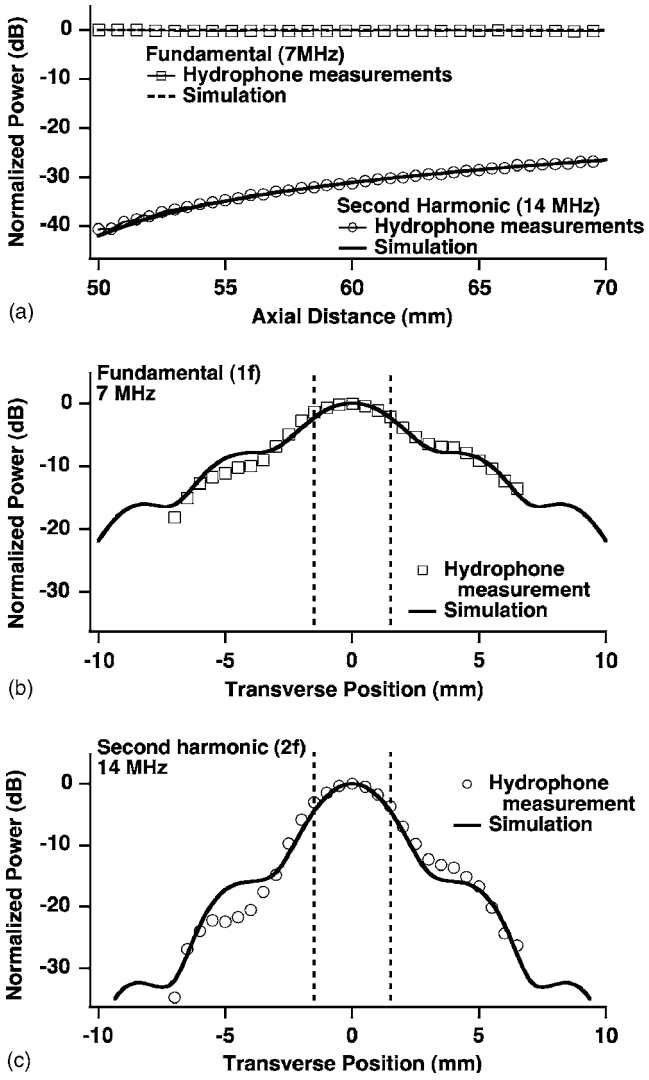


FIG. 5. (a) Comparison of hydrophone measurements and simulation results, for both the fundamental and second harmonic, in the axial direction. (b) The normalized transverse profile for the 7 MHz fundamental, 70 mm away from a 12.7-mm-diam piston source, propagating through 46.4 mm of stainless steel and 23.7 mm of water. (c) Normalized transverse profile of the second harmonic amplitude at the same location as (b).

V. DISCUSSION

Overall, good agreement is observed between the measured and simulated cross sections of the ultrasonic fields. For the 7 MHz case, the measured magnitude and features of the transverse cross sections for the fundamental ($1f$) component appear to be relatively constant over the 20 mm range of axial propagation distances investigated. This agrees with the simulation (Fig. 2), where the amplitude of the 7 MHz fundamental is observed to vary by only 0.13 dB along the axis in the region of the water filled sample chamber. The hydrophone measurements of the on-axis power, for both the fundamental (7 MHz) and second harmonic (14 MHz) components, are in good agreement with predictions from the simulation, as can be seen from Fig. 5(a). Figures 5(b) and 5(c) indicate that the field varies by less than 3 dB for the fundamental and less than 6 dB for the second harmonic across a 3 mm diameter. The relatively good agreement be-

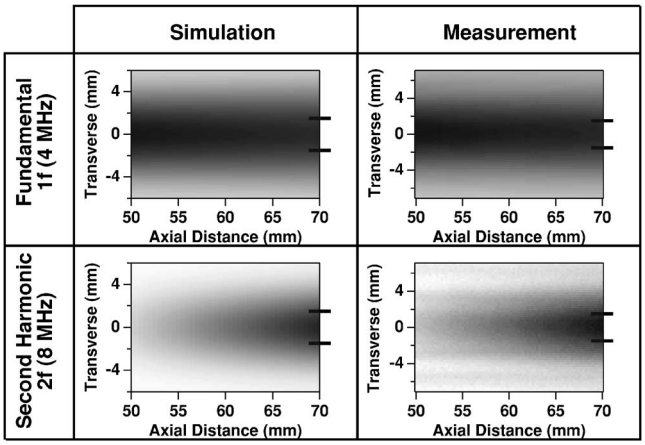


FIG. 6. Comparison of expected beam pattern from simulation (left) with hydrophone measurement of the field (right) in a meridian plane beginning 4 mm beyond the steel-water interface, for a 4 MHz fundamental signal. The separation between the hash marks on the right side of each graph corresponds to the 3 mm diameter of a receiving transducer.

tween the experimental measurements and predictions from the simulation support the validity of a plane wave approximation to the field.

Transmission of a narrowband 4 MHz signal yielded similar results. The fundamental component of the signal remains relatively constant across the measured range of axial propagation. As seen in Fig. 3, the amplitude of the fundamental component decreases by only 0.9 dB over the 20 mm axial range of the figure. Examination of Figs. 7(b) and 7(c) indicates that the ultrasonic power varies by less than 0.75 dB for the fundamental and less than 1.5 dB for the second harmonic across the 3 mm diameter that is marked by the vertical dashed lines. In comparison with the 7 MHz case, where the 3 dB width of the ultrasonic beam was only 3 mm, the 3 dB width for the 4 MHz case has increased to over 6 mm. This indicates a wider region over which a plane wave approximation is valid.

There is, however, a drawback to using the far field region of the ultrasonic beam to generate a wider plane wave region. Due to the nature of the far field, the ultrasonic amplitude of the fundamental signal will not be as large as what is found in the focal zone, so there will be a correspondingly lower rate of harmonic generation. In comparison with the 7 MHz study illustrated in Figs. 2 and 3, the final amplitude after 25 mm of propagation in water is smaller for the 4 MHz case, although in both cases the initial harmonic is negligible. Additionally, the nonlinear generation of the second harmonic is linearly dependent upon the frequency in the weakly nonlinear regime, and thus using a lower frequency reduces the rate of generation of the second harmonic distortion.[11] The decrease in the signal-to-noise ratio for the second harmonic can be seen in Fig. 6. Nevertheless the far field approach does allow for a certain amount of flexibility because an experiment can be adjusted to give a plane wave over the field depth needed for propagation through a sample while maintaining the necessary signal-to-noise ratio needed for measuring the harmonic distortion of a pulse.

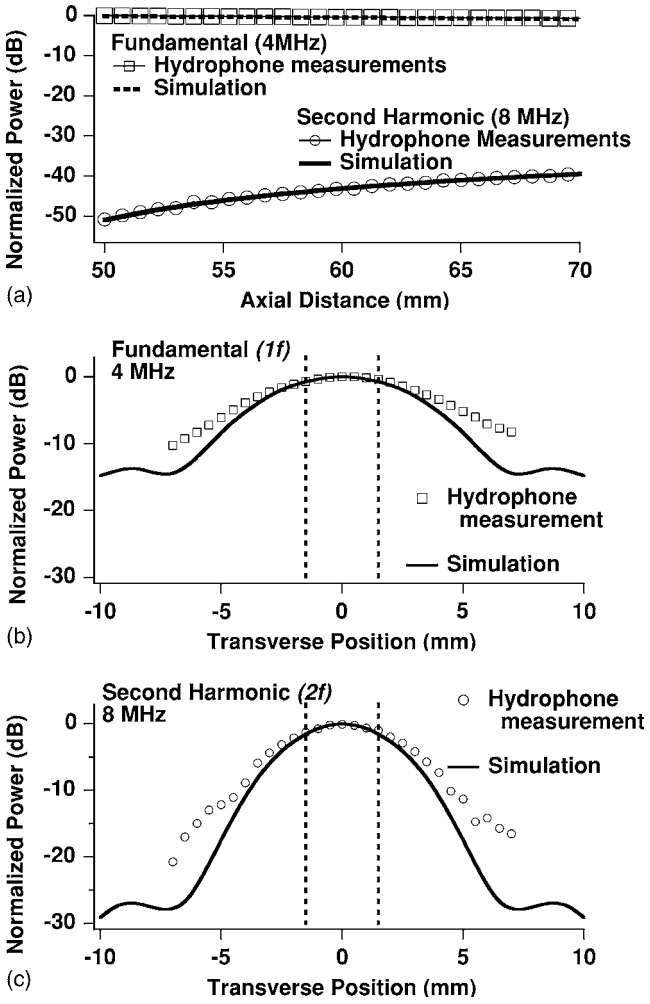


FIG. 7. (a) Comparison of hydrophone measurements and simulation results, for both the fundamental and second harmonic, in the axial direction. (b) The normalized transverse profile for the 4 MHz fundamental, 70 mm away from a 12.7-mm-diam piston source, propagating through 46.4 mm of stainless steel and 23.6 mm of water. (c) Normalized transverse profile of the second harmonic amplitude at the same location as (b).

An additional benefit of the delay line lies in the lack of nonlinear generation within the steel, compared to propagation in water. To understand this, consider the shock formation distance for plane wave propagation

$$\bar{x} = \frac{\rho_0 c_0^3}{\beta \omega_0 p_0}, \tag{2}$$

where ρ_0 is the density, c_0 is the sound speed, β is the coefficient of nonlinearity of the medium, and ω_0 and p_0 are the angular frequency and pressure of the incident pulse, respectively. The density of steel is roughly eight times that of water, and the sound velocity is about four times that of water. The nonlinear parameters of steel and water are approximately the same, as discussed earlier. Thus, calculation of the ratio of shock formation distances for steel and water reveals that the shock formation distance in steel is much larger than that for water. Therefore, the relative amount of nonlinear distortion is significantly less in steel. In addition, propagation to a given portion of the diffraction field in steel is accomplished in one quarter the propagation distance in water; nonlinear generation is reduced further by this reduc-

tion in path length. Thus, the distortion in a typical length delay line is negligible compared to the distortion in the same diffractive region in a water path. This is illustrated in both Figs. 2 and 3, which indicate that the simulations predict that the ultrasonic signal has virtually no amplitude in the second harmonic while the fundamental propagates through the steel. This is also affirmed in the experimental work, in which a very small amount of distortion is measured in the signal when measured as close to the face of the delay line as possible. For the 7 MHz study, the measured amplitude in the second harmonic was 40 dB down from the amplitude in the fundamental signal, and for the 4 MHz study this difference increased to 50 dB.

The use of this delay line apparatus has other practical advantages. One primary benefit is that the absence of transmitted harmonics obviates the need to employ physical means of attenuation to reduce unwanted harmonic signals. The introduction of such materials, such as attenuating media and quarter-wave plates, can cause signal loss in the fundamental frequency.[6,19] Another benefit of the delay line is that it can be used in other types of nonlinearity experiments. If the whole assembly is oriented vertically and a chamber wall for containing fluid samples is attached, measurements of the nonlinearity parameter based on a pullback approach can be performed.[1–4] Such a setup benefits from the plane waves generated as well as the lack of transmitted harmonics, but permits the receiving transducer to be manually translated in the fluid medium of interest.[4]

ACKNOWLEDGMENTS

This work was supported, in part, by NIH Contract Nos. R01 HL-072761 and R37 HL-40302.

[1]W. K. Law, L. A. Frizzell, and F. Dunn, "Ultrasonic determination of the nonlinearity parameter B/A for biological media," J. Acoust. Soc. Am. **69**, 1210–1212 (1981).
[2]W. K. Law, L. A. Frizzell, and F. Dunn, "Determination of the nonlinearity parameter B/A of biological media," Ultrasound Med. Biol. **11**, 307–318 (1985).
[3]K. D. Wallace, M. R. Holland, and J. G. Miller, "Novel approach for overcoming effects of diffraction in measurements of the nonlinear B/A parameter in liquidlike media," J. Acoust. Soc. Am. **116**, 2566 (2004).
[4]K. D. Wallace, C. W. Lloyd, M. R. Holland, and J. G. Miller, "Finite amplitude measurements of the nonlinear parameter B/A for liquid mixtures spanning a range relevant to tissue harmonic mode," Ultrasound Med. Biol. **33**, 620–629 (2007).
[5]L. Germain, R. Jacques, and J. D. Cheeke, "Acoustic microscopy applied to nonlinear characterization of biological media," J. Acoust. Soc. Am. **86**, 1560–1565 (1989).
[6]F. Dong, E. Madsen, M. MacDonald, and J. Zagzebski, "Nonlinearity parameter for tissue-mimicking materials," Ultrasound Med. Biol. **25**, 831–838 (1999).
[7]L. Bjorno, "Characterization of biological media by means of their nonlinearity," Ultrasonics **24**, 254–259 (1986).
[8]X.-F. Gong, F. Ruo, Z. Cheng-ya, and S. Tao, "Ultrasonic investigation of the nonlinearity parameter B/A in biological media," J. Acoust. Soc. Am. **76**, 949–950 (1984).
[9]X.-F. Gong, Z.-M. Zhu, T. Shi, and J.-H. Huang, "Determination of the acoustic nonlinearity parameter in biological media using FAIS and ITD methods," J. Acoust. Soc. Am. **86**, 1–5 (1989).
[10]S. Saito, A. Yamamoto, and K. Nakamura, "B/A measurement for liquid media using an LN transducer with inverted-domain layer," Jpn. J. Appl. Phys., Part 1 **44**, 4431–4435 (2005).
[11]V. V. Shklovskaya-Kordi, "An acoustic method of determining the internal pressure in a liquid," Sov. Phys. Acoust. **9**, 82–85 (1963).
[12]J. Wu and J. Tong, "Measurements of the nonlinearity parameter B/A of contrast agents," Ultrasound Med. Biol. **24**, 153–159 (1998).
[13]L. E. Kinsler, A. R. Frey, A. B. Coppens, and J. V. Sanders, *Fundamentals of Acoustics*, 4th ed. (Wiley, New York, 2000), pp. 179–184.
[14]R. J. Fedewa, K. D. Wallace, M. R. Holland, J. R. Jago, G. C. Ng, M. R. Rielly, B. S. Robinson, and J. G. Miller, "Effect of changing the transmit aperture on the spatial coherence of backscatter for the nonlinearly generated second harmonic," *Proceeding of the IEEE Ultrasonics Symposium 02CH37388*, Munich, Germany, 1624–1627 (2002).
[15]P. T. Christopher and K. J. Parker, "New approaches to nonlinear diffractive field propagation," J. Acoust. Soc. Am. **90**, 488–499 (1991).
[16]C. J. Vecchio and P. A. Lewin, "Finite amplitude acoustic propagation modeling using the extended angular spectrum method," J. Acoust. Soc. Am. **95**, 2399–2408 (1994).
[17]M. A. Breazeale, "finite-amplitude waves in solids," in *Handbook of Acoustics*, edited by M. J. Crocker (Wiley, New York, 1998), pp. 211–219.
[18]D. T. Blackstock, M. F. Hamilton, and A. D. Pierce, "Progressive waves in lossless and lossy fluids," in *Nonlinear Acoustics*, edited by M. F. Hamilton and D. T. Blackstock (Academic, San Diego, 1998), pp. 65–150.
[19]K. D. Wallace, Ph.D. thesis, Washington University in Saint Louis (2001).

Tele-infrasonic studies of hard-rock mining explosions

Mihan H. McKenna[a)]
Structural Engineering Branch, Geotechnical and Structures Laboratory, U.S. Army Engineer Research and Development Center, 3909 Halls Ferry Road, Vicksburg, Mississippi 39180

Brian W. Stump[b)]
Department of Geological Sciences, Southern Methodist University, 3225 Daniel Avenue, Dallas, Texas 75275

Sylvia Hayek[c)]
Natural Resources Canada, C.E.F. 930 Carling Ave. 7 Observatory Crescent, Room. 20, Ottawa, Ontario, Canada K1A 0Y3

Jason R. McKenna[d)]
Engineering Geology and Geophysics Branch, Geotechnical and Structures Laboratory, U.S. Army Engineer Research and Development Center, 3909 Halls Ferry Road, Vicksburg, Mississippi 39180

Terry R. Stanton[e)]
Structural Engineering Branch, Geotechnical and Structures Laboratory, U.S. Army Engineer Research and Development Center, 3909 Halls Ferry Road, Vicksburg, Mississippi 39180

(Received 31 October 2006; revised 26 April 2007; accepted 26 April 2007)

The Lac-du-Bonnet infrasound station, IS-10, and the Minnesota iron mines 390 km to the southeast are ideally located to assess the accuracy of atmospheric profiles needed for infrasound modeling. Infrasonic data from 2003 associated with explosions at the iron mine were analyzed for effects of explosion size and atmospheric conditions on observations with well-constrained ground truth. Noise was the determining factor for observation; high noise conditions sometimes prevented unequivocal identification of infrasound arrivals. Observed arrivals had frequencies of 0.5 to 5 Hz, with a dominant frequency of 2 Hz, and generally had durations on the order of 10 s or less. There was no correlation between explosive amount and observability. Tele-infrasonic propagation distances (greater than 250 km) produce thermospheric ray paths. Modeling is based upon MSIS/HWM (Mass Spectrometer Incoherent Scatter/Horizontal Wind Model) and NRL-G2S (Naval Research Laboratory Ground to Space) datasets. The NRL-G2S dataset provided more accurate travel time predictions that the MSIS/HWM dataset. PE modeling for the NRL-G2S dataset indicates energy loss at higher frequencies (around 4 Hz). Additionally, applying the Sutherland/Bass model through the NRL-G2S realization of the atmosphere in InfraMAP results in predicted amplitudes too small to be observed. © *2007 Acoustical Society of America.*
[DOI: 10.1121/1.2741375]

PACS number(s): 43.28.Dm, 43.28.Fp, 43.28.We, 43.28.Hr [VEO] Pages: 97–106

I. INTRODUCTION

This study is designed to quantify long-range infrasonic propagation from mining explosions with ground truth provided by the mine and seismic observations. Analysis of infrasound generated by mining explosions can yield invaluable information about the nature of infrasound propagation since the source is relatively temporally and spatially constrained through records of blast timing and specific shot patterns for each blast. A continually operating hard rock iron mine in Minnesota has provided blast records for 404 events over nearly four years for infrasound events observable at tele-infrasonic distances (greater than 250 km) at the Canadian administered International Monitoring System (IMS) array IS-10 in Lac-du-Bonnet, Manitoba (www.ctbto.org). The blast parameter dataset contains origin times recorded by the mine personnel. These times can be compared to regionally determined seismic arrival times to better pinpoint the actual time of blast for propagation modeling to determine whether the seismic origin time is a better estimate of the actual source time.

This is an ideal situation to assess the accuracy and applicability of the available atmospheric datasets for infrasound modeling, particularly where energy travels through the atmosphere at stratospheric, mesospheric, and thermospheric heights over a range of time periods, from days to weeks, months, or years. This paper will utilize the mine records and associated infrasound recordings at IS-10 to compare two modeling techniques, ray tracing and parabolic equation (PE) modeling, to assess the validity of both the modeling codes as well as the atmospheric models. Ray trace

[a)]Electronic mail: Mihan.H.McKenna@erdc.usace.army.mil
[b)]Electronic mail: Bstump@smu.edu
[c)]Electronic mail: Sylvia.Hayek@nrcan-rncan.gc.ca
[d)]Electronic mail: Jason.R.McKenna@erdc.usace.army.mil
[e)]Electronic mail: Terry.R.Stanton@erdc.usace.army.mil

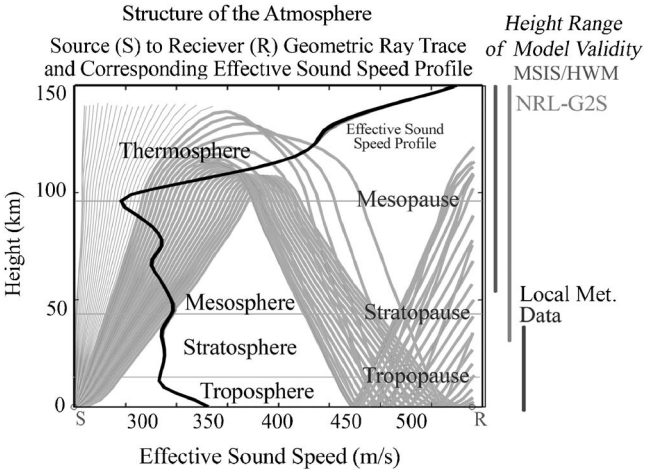


FIG. 1. Schematic of the atmosphere with the corresponding effective sound speed profile. The valid range of three types of atmospheric data shown on the right, with MSIS/HWM and NRL-G2S as subjects in this paper. Local meteorological data is provided for comparison, though is not utilized in this study. Behind effective sound speed profile is the ray trace result for modeling through that atmosphere.

modeling yields insight into travel time and energy pathway while PE modeling gives information on the frequency dependent nature of the amplitude at any given location for the propagating acoustic energy.

Infrasound is low frequency sound waves between 0.1 to 20 Hz, below the range of human hearing from 20 to 20 000 Hz (McKisic, 1997). Known sources of infrasound include avalanches, meteors, ocean waves, severe-weather systems, tornados, turbulence, earthquakes, volcanoes, mining explosions, cultural noise, and even animal vocalization (Bedard and Georges, 2000). Some of the most commonly observed sources observed at infrasound arrays are man-made explosions, such as those recorded at an infrasound array located on the Korean Peninsula (Che *et al.*, 2002). These explosions can be industrial, mining, or related to large single-fired blasts conducted for a variety of reasons (Chael and Whitaker, 2003). Figure 1 shows the sample effective sound speed profile with the regions of the atmosphere labeled, where *S* represents the source location and *R* represents the receiver array location. To the right side of the figure are lines representing the ranges of valid atmospheric parameters associated with the atmospheric models evaluated in this work: MSIS/HWM (Mass Spectrometer Incoherent Scatter/Horizontal Wind Model), NRL-G2S (Naval Research Laboratory Ground To Space), as well as Local Met. Data (Local Meteorological Data), presented for comparison. It should be noted that MSIS/HWM and NRL-G2S are identical in the upper thermosphere. Behind the effective sound speed profile is the corresponding ray trace result for modeling through that atmosphere. The lightest grey lines represent energy that dissipates into the thermosphere and the darker grey lines represent energy that returns to the surface of the Earth. For further information regarding ray trace modeling, see Sec. III A. The atmospheric profiles shown in the figure are explained in depth where they are used; this figure is a useful reference while reading the rest of this paper.

Previous work on infrasound generated by mining explosions includes work done by Sorrells, Herrin and Bonner (1997). They investigated infrasound arrivals at TXAR (Texas Array, Lajitas, TX) from mines in Mexico and the Southwest USA, implementing preliminary modeling studies to explain the seasonal variation in arrivals (Sorrells *et al.*, 1997). Stump *et al.* (2002) assessed infrasound generated at the Minntac taconite mine in the Mesabi Range of Minnesota from eight explosions in August and November of 2000. Five of the eight arrivals were recorded at IS-10, with explosions over 500 000 pounds generating observed signals. However, no source modeling was performed, and the effects of seasonal winds on propagation have not been taken into account (Stump *et al.*, 2002).

A similar investigation of the infrasound generated at the Morenci copper mine in southwest Arizona resulted in 7 of 25 signals with ground truth being observed at LANL, the infrasound array in Los Alamos, New Mexico, as well as several regional signals recorded at a temporary portable array. While initial investigations into back azimuth and frequency content were performed, there was no specific source assessment of these events (Stump *et al.*, 2002). Other investigations into mining infrasound include the work of Hagerty *et al.* (2002) into mining explosions in Kazakstan, but again modeling consisted of ray tracing with no investigation of specific source mechanisms.

The mine studied in this paper had been previously assessed for seismic propagation and source mechanics by Hetzer (2000) and Goforth *et al.* (2006). These authors document the lack of correlation between seismic amplitude and recorded yield at regional distances, similar to the Stump (2002) study and explain it in terms of the constructive and destructive interference between the individual explosions in a delay-fired mining explosion.

II. DESCRIPTION OF EXPERIMENT AND DATA

A. Minnesota hard rock iron mine

Blast statistics from the iron mine have been sent to Southern Methodist University for the time period from June 2001 to September 2004. This set of 404 records includes information on date and time of shots, total yield, number of holes per blast, pounds of blasting agent, pounds per delay, local weather conditions, and how many different shots make up one mine-wide blast record. Blasts can consist of multiple delay-fired shots initiated simultaneously in different locations throughout the mine and the measurement of total tons recorded reflects all shots across the mine in one blast.

A subset of these records was selected for modeling in this study: 126 records from 1/03/03 to 12/30/03. Table I, Column 1, shows the number of events per month during 2003. There is a fairly even distribution of blast days across the year, ranging from eight to twelve blasts a month.

The explosive amounts range from 20 000 to 1 700 000 lbs with the number of holes for delay-fired shots ranging from 6 to 425 holes per shot; this gives an estimate of the temporal finiteness of the blast (Goforth, 2006). The total tons of rock moved range from 22 500 tons to 2 054 215 tons per blast. Blasts occur under a range of local weather condi-

TABLE I. Catalog of observed event characteristics.

Month	Total number of events in the blast record	Number of predicted arrivals	Number of observed events	Number of events with good $f-k$ estimates	Observed time residual from the predicted NRL-G2S times (min)	Average phase velocity of recorded arrivals (m/s)	Average azimuth of recorded arrivals (deg)
Jan	10	10	6	5	−1.43	371.2	144.6
Feb	12	12	9	9	0.18	323.5	141.3
Mar	8	8	8	6	−0.27	334.0	143.7
Apr	12	12	8	7	−0.88	327.4	146.6
May	10	10	8	8	1.43	339.5	144.2
Jun	11	10	8	7	2.31	333.1	142.1
Jul	10	10	6	2	1.84	365.5	140.5
Aug	11	11	2	0	3.63	N/A	N/A
Sep	10	10	6	3	0.72	358.7	139.7
Oct	11	5	4	4	0.25	372.5	139.5
Nov	11	7	5	4	−1.67	295.5	143.0
Dec	9	5	3	3	1.35	348.0	142.7

tions with a slight preference for cloudy days when the wind was out of the south or southeast. This selection is done to mitigate local structural damage by only shooting on days that minimize energy ducting through local inversion layers.

The iron mine is located near the town of Mountain Iron, Minnesota, in the Mesabi Iron Range part of the Biwabik Iron-Formation located at the southwest edge of the Canadian shield (Sims *et al.* 1993, French 1968). The ore body stretches nearly 100 miles and ranges from 1.5 to 3 miles wide and up to 600 feet deep. Though the mine has been in production since the 1890s, today the iron ore is extracted as taconite to form 65 percent iron pellets which are sent to steel processing facilities in the Great Lakes region. Up to 150 000 tons of taconite can be blasted each day. The mine is located at 47.38 N, 92.85 W.

B. Infrasound and seismic stations

The geometry of the array used in this experiment is shown in Fig. 2, along with the relative location of the seismic station used for identifying source event times. Infrasound data was collected at IS-10 (50.2015 N, 96.0269 W for the central station I10H1; other stations are clockwise from the northwestern station I10H2, I10H3, and I10H4) near Lac-du-Bonnet, Manitoba, 390 km away from the mine located along an azimuth of 142°. IS-10 is a four-element International Monitoring System (IMS) infrasound array that records data at 20 Hz with an array aperture of 1 km. The IMS network of arrays was constructed to globally monitor nuclear explosions using a combination of seismic, infrasound, hydro-acoustic, and radionuclide methods. The IS-10 array is ideally located to assess the source to receiver path effects where the energy propagation is over tele-infrasonic distances. Near to IS-10 is the seismic station ULM (50.24987 N, 95.8757 W).

Figure 3 shows a seismic record observed at stations ULM, both filtered and unfiltered, from an event on 29, Jan. 2003. It should be noted that not all of the blasts in the region of the Mesabi Iron Range are associated with the iron mine that provided the ground truth records; there are several mine blasts clustering throughout the Mesabi Range recorded by this system of stations that are not associated with the mine under investigation.

The seismic records and calculated origin times provided by Natural Resources Canada were used to verify the source time and define the record section to examine for the infrasound arrival. The origin times were determined using the standard Canadian velocity model east of the Rocky Mountains with the depth of blast fixed at the surface and were calculated using records from ULM as well as whichever other temporary stations were available, such as EYMN, TBO, and SOLO. The time difference for the origin times as determined by the blast records and the seismic records routinely vary an average of 2 min. In some cases, there was no associated seismic arrival with which to correlate the acous-

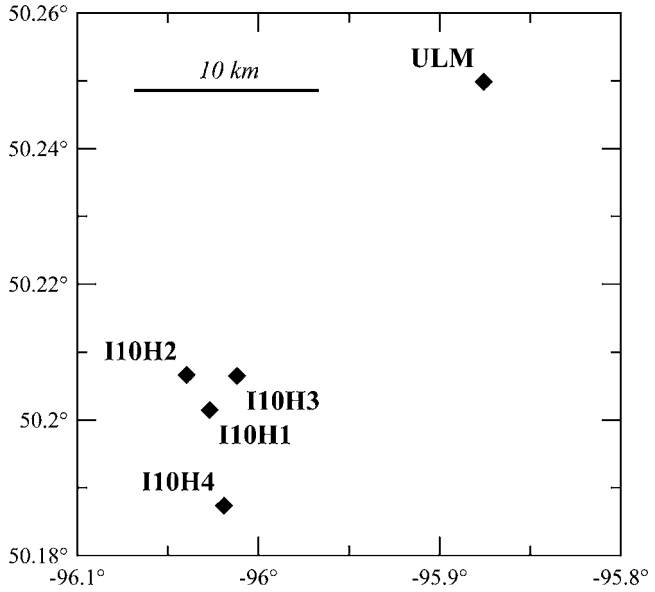


FIG. 2. Locations of the IS-10 acoustic array elements (I10H1, I10H2, I10H3, I10H4) and the ULM seismic array. The array aperture for IS-10 is approximately 1 km.

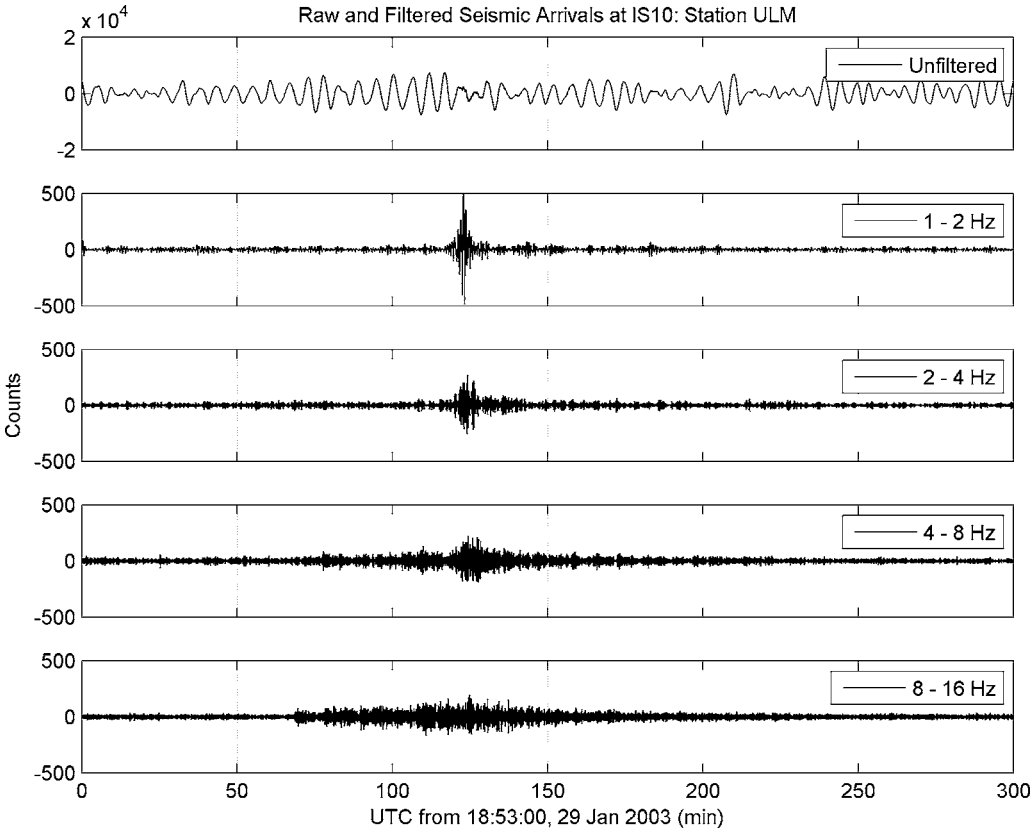


FIG. 3. Seismic record section from the 29 Jan. blast. from ULM, filtered and unfiltered.

tic arrival, or the closest recorded seismic arrival was off as much as 25 min which indicates that the blast is not correctly associated with the seismic records. These missing data were not considered for this paper.

C. Infrasound data

The steps for analyzing the data were as follows:

(1) Data was initially filtered from 0.5 to 5 Hz based on the determination of the frequency band with maximum signal-to-noise ratio from previous infrasound investigations of explosions.

(2) Data was searched for a possible associated arrival for the events catalogued in the blast records. The window of search was ±5 min of predicted infrasound arrival based on the seismic origin time.

(3) Once an arrival was identified in the time domain the phase velocity and azimuth were calculated using frequency-wave number (*f-k*) methods. This analysis uses the frequency-wave number spectrum, the squared modulus of the multidimensional Fourier transform of the array data in time and space (Smart and Finn, 1971), to identify an associated phase velocity and azimuth for a given arrival.

An example arrival corresponding to the seismic arrival on 29, Jan. 2003 presented above is shown in Fig. 4. I10H is the array element, numbered 1 through 4, where the arrival is located at approximately 6.8 min on each trace, evidenced by the uniform frequency character and regular arrival envelope. This plot has been filtered from 1–5 Hz, for ease of visualization. The frequency and time domain character of the arrival in Fig. 4 is similar to most recorded arrivals from the mine. Note the complex character of the waveform and the signal to noise ratio slightly greater than 2.

All four channels are used to estimate phase velocity and azimuth using the *f-k* method. Since this data is broadband and the array geometry is sufficiently large, careful *f-k* calculations provide accurate phase velocity and azimuth estimates (Whitaker *et al.* 2002), though for three observed events in this study no resolvable *f-k* could be determined. Band-pass filtering the data indicates there dominant frequency of the observed arrivals is 2 Hz, with energy visible from 0.5 to 5 Hz, which is consistent with the long travel path from the mine to IS-10. The average observed azimuth for all the mining explosion observations in this study is 142.53° with a standard deviation of 2.19° compared to the actual azimuth of 142°. The average observed phase velocity is 342.63 m/s with a standard deviation of 23.47 m/s.

Infrasound arrivals associated with the mining records are generally on the order of 5 to 10 sec long and can consist of multiple arrivals. In order to definitively conclude that an arrival occurred, the signal must be seen on two or more channels. As wind bursts and other noises were common, there are visible arrivals that could not be fully analyzed since the *f-k* method requires a minimum of three channels. The noise field at IS-10 is complex, alternating from very quiet periods to periods dominated by wind noise, which agrees with findings at globally located infrasound arrays as discussed in Bowman *et al.* 2005. For example, on 23 April 2003, the wind noise obscures any observable arrivals. Figure 5 shows local wind bursts that would obscure any arriv-

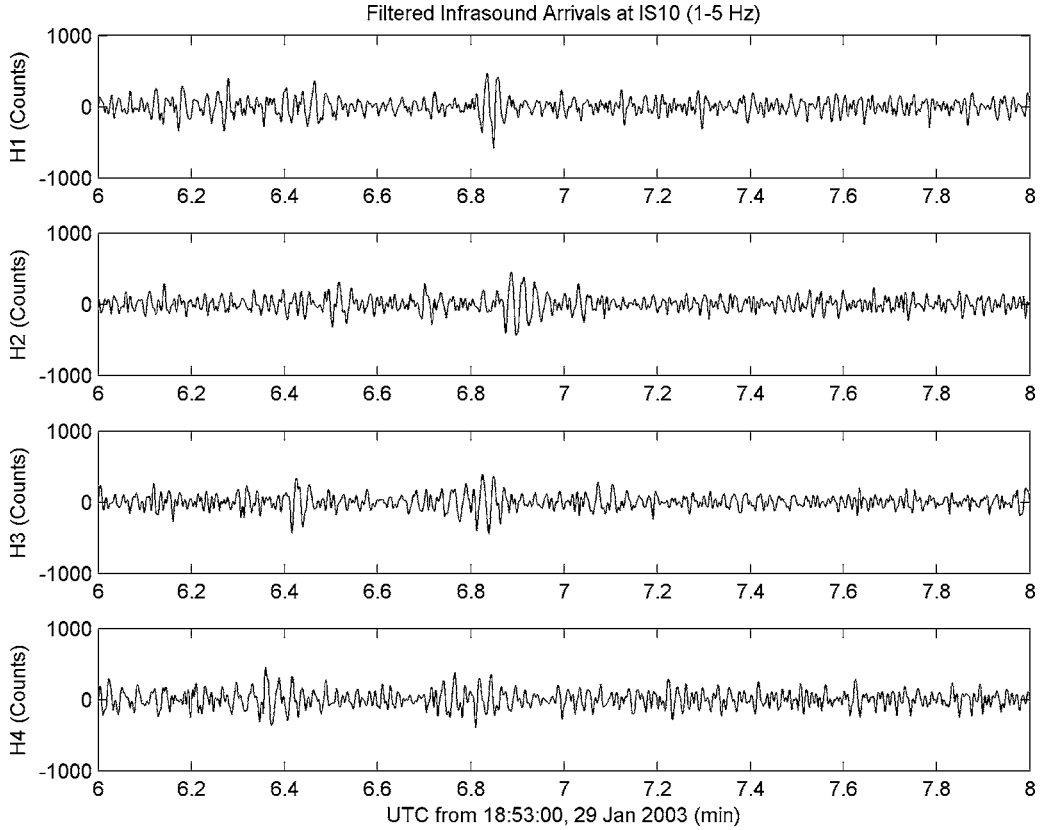


FIG. 4. Sample arrival at IS-10 from the iron mine on 29 Jan. The data has been filtered from 1 to 5 Hz.

als on individual array elements, where the wind is not coherent across the array due to the 1 km array aperture. However, IS-10 is a relatively quiet array (five events during this study were possibly obscured by wind noise) with good signal-to-noise ratios for most of the observed explosions.

The analysis of consecutive twelve months of data is summarized in Table I. The majority of events were observed at IS-10 with relative consistent azimuths but variable arrival times. There appears to be no correlation to total tons moved and the likelihood of an infrasonic observation. Figure 6 shows the correlation between size of the explosive event and whether the event was observed at IS-10. Both the largest and smallest events were observed, though many events with moderate sized explosions were not observed: 58.98% of total events were observed, but 67.61% of predicted arrivals were observed. These unobserved events often occurred on high noise days where wind bursts obscured possible arrival (Fig. 5). This lack of correlation between size of explosion and amplitude of arrival corresponds to what was found for the previous seismo-acoustic study by Stump *et al.* (2002) where there was no correlation from total tons moved to seismic or infrasound amplitude at regional distances. Similar results were found in the seismic-only study by Goforth (2006) of other blasts from this same iron mine in Minnesota. However, it should be noted that a correlation between size of the explosion and dominant period of the observed infrasound signal is well known for atmospheric explosions, dating from the analysis of large atmospheric nuclear explosions in the 1960s (Burg 1972), but less characterized for partially or fully contained explosions.

III. MODELING

Modeling was performed to assess a variety of characteristics of the observed data. The arrival times are compared to the travel times generated by the models for origin times from both the blast records and the calculated seismic arrivals. The two atmospheric datasets, MSIS/HWM and NRL-G2S, are compared to see which provides a better estimate of these travel times. The infrasound energy travel path is assessed to indicate whether the energy traveled thermospheric or stratospheric pathways. Lastly, attenuation effects will be explored to examine both the frequency dependence of the propagation as well as the predicted amplitude at the receiver.

A. Ray path modeling

Ray path modeling in this study utilized the ray tracing program found in InfraMAP, BBN's proprietary infrasound propagation modeling code. This is a three-dimensional ray theory model, Hamiltonian Ray-Tracing Program for Acoustic Waves in the Atmosphere (HARPA), originally developed by Jones *et al.* (1986) and modified by the staff at BBN Technologies for implementation in InfraMAP. This code incorporates a range-dependent atmosphere in a three-dimensional spherical domain to calculate horizontal and vertical refraction, horizontal translation of rays due to moving media, absorption, and attenuation of the wave front by viscous and thermal losses (Gibson and Norris 2002).

MSIS/HWM and NRL-G2S are used as input atmospheric models to compare which predicted travel times were

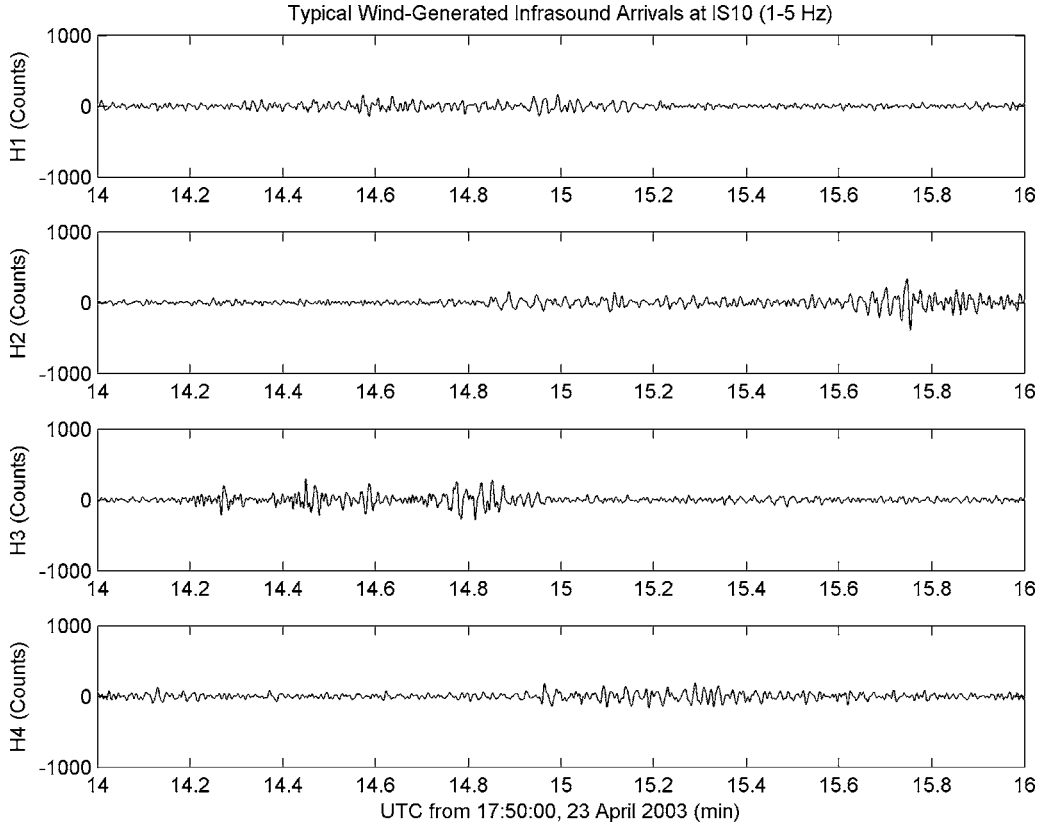


FIG. 5. Wind bursts seen at IS-10 on 23 April, 2003.

most consistent to observed arrivals; this would indicate that the model is a more accurate depiction of the upper atmosphere. The temperature profile MSIS-90 and wind profile HWM utilize ground, rocket, and satellite-based measurements to provide statistical estimates of upper atmospheric temperature, total density, and composition for a specific latitude and longitude (Picone *et al.*, 1997). This model is able to extrapolate profiles for sites not specifically measured by the above-mentioned techniques, though it is only strictly valid above the stratopause (Gibson and Norris, 2002). Both components of the MSIS/HWM profile are generated from 15-year averages and do not take into account immediate, or real-time, temperature and wind information, though for energy turning at thermospheric heights, temporal variation is slight (Garcés *et al.* 1999).

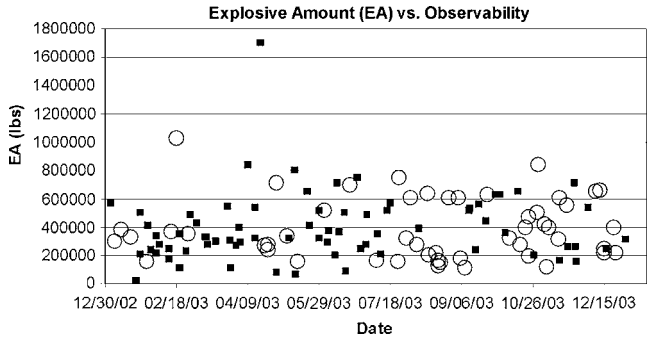


FIG. 6. Comparison of total explosive amount to observed arrivals. The black squares were observed blasts, the open circles were not observed at IS-10.

NRL-G2S is a semi-empirical spectral model which integrates actual temperature and wind measurements into the climatological models and numerical weather prediction (NWP) models (Gibson and Norris, 2004). The NWP components are valid from the surface to 55 km, eliminating the predominant problems of HWM/MSIS dataset, and have a spatial resolution of about one degree (Drob, 2003). This characteristic of the model is important for energy traveling in low thermosphere or stratospheric heights, where the pathways would change on the order of days to months. We will compare the travel time predictions to our observations for each of these models in order to assess their validity. For both profiles, the predominant wind at thermospheric heights during the twelve months was from east to west, toward the infrasound array. For both datasets, UT hour 18 data was utilized for consistency; all events occurred between UT 14 and UT 19. Propagation times were determined from the ray that arrived at the minimum distance from IS-10. A maximum miss distance of 20 km was allowed to take into account the location variation within the mine itself during blasting. Turning rays are considered to be likely to be observed if they are predicted to return within 10 km of the receiver.

For a majority of the modeled days, NRL-G2S and MSIS/HWM yielded similar results, though NRL-G2S predicted more complex turning rays in the stratosphere (multiple propagation paths through stratospheric ducts) and the occasional low altitude arrivals not present in the MSIS/HWM dataset, so NRL-G2S was used for modeling throughout this investigation. Of the 110 predicted arrivals over the

twelve month study, all are thermospheric except one event in June. However, during this twelve month study, there were a total of eight predicted events with additional stratospheric arrivals. There was one predicted event each in both January and August that was not seen. Of the three predicted stratospheric travel paths in June, one was the event with no associated thermospheric arrival and was recorded; one predicted event was not observed, the second predicted event arrived in between the predicted times for the stratospheric and thermospheric arrivals. In July, there were four predicted paired stratospheric and thermospheric arrivals, three of which had single signal recordings which arrived between stratospheric and thermospheric travel times; one was not observed. An example ray trace solution with corresponding NRL-G2S profile is shown in Fig. 7. Figure 7(a) is the effective sound speed at points along the great circle path between source and receiver; Fig. 7(b) is the ray trace solution through that atmosphere, with all possible bounce points shown and the receiver location marked by the letter R; Fig. 7(c) is the parabolic equation solution for the same source to receiver path through that atmosphere, where the color bar indicates the dB loss at a particular location, with a 70 dB loss at the receiver location. See the next section for information on parabolic equation calculations. Most modeled events yielded nearly identical results for the two atmospheric datasets, MSIS/HWM and NRL-G2S.

B. PE modeling

Originally adapted from underwater acoustic studies, PE modeling provides a full wavefield solution for a complete vertical plane at one frequency. PE codes are appropriate for propagation modeling where trapped energy, diffraction, and interface effects related to spherical wave fronts become important. In InfraMAP, the PE code steps forward from a source and calculates an attenuation field for predicting amplitudes along the vertical slice (Gibson and Norris, 2002). In using the PE codes, it is imperative that the computational domain be large enough to include all viable geometric rays, as is seen in ray path modeling (West *et al.*, 1992). The frequency chosen for the majority of study is 2 Hz, though previous work done by the authors indicates that results of PE modeling can change dramatically depending on which frequency was chosen for modeling. A comparison of the raypath solution and PE solution is presented in Fig. 7. The two results agree well with each other, showing strong correlation between the predicted raypaths and the highest amplitude acoustic energy in the PE. This remains the case when different propagation modes are found. However, it should be noted that the PE model consistently predicts energy to return to the ground at a slightly greater range than the ray trace model does. It is unclear at this time why this is, though it is possible that the lack of absorption model in this modeling may play a role.

In some cases some predicted arrivals were not observed, even when high wind noise across the array was not a factor. In these cases, the authors ran PE models for a suite of frequencies to determine whether changing the frequency component of the calculation would affect the predicted observability at the array. The data indicates that infrasound energy is present in the data between 0.5 and 5.0 Hz; 0.5, 1.0, and 4.0 Hz models were run to assess change in model frequency. The change in frequency does not result in a dramatic change in predicted amplitude and in each case where low frequency arrivals were predicted, higher frequency arrivals were also predicted with an equivalent loss of energy.

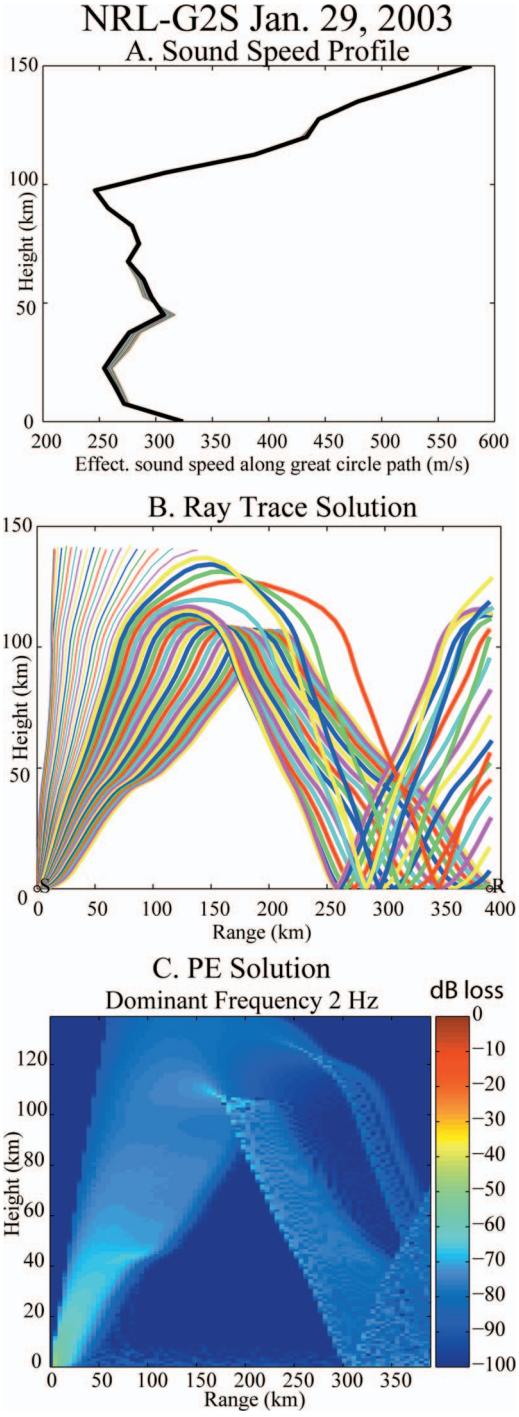


FIG. 7. Comparison of NRL-G2S atmospheric profile (A), raypath solution (B) and PE solution (C) modeling results for 29 Jan. These are the model results which correspond to the signal shown in Fig. 3.

The above PE modeling was performed without absorption. There is a current debate in the infrasound community as to whether or not the Sutherland/Bass model for absorp-

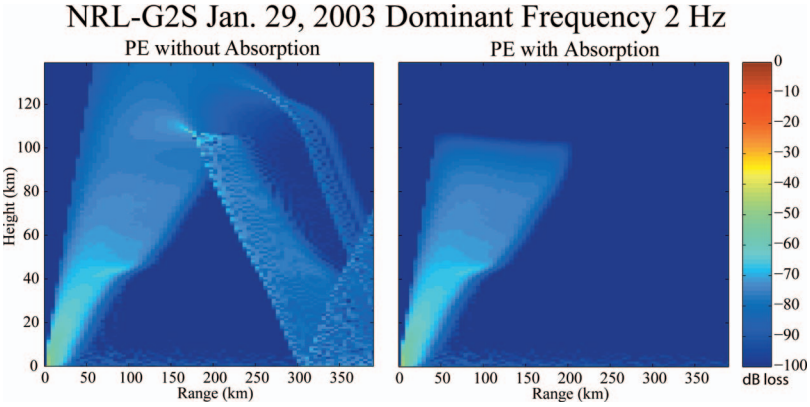


FIG. 8. PE modeling results with and without absorption for 29 Jan.

tion of energy in the atmosphere is appropriate (Norris, 2005; Edelmann *et al.*, 2005). Sutherland and Bass (2004) developed an attenuation model for the loss of energy due to spherical spreading over a given range or specific ray path. It is designed to be used at high altitude for low frequencies and takes into account both classical (translation and diffusion) and relaxational (rotation and vibration) losses (Gibson and Norris, 2004). Figure 8 shows two identical PE models run through the same atmosphere, where no absorption is applied to the calculation in the figure on the left, but the Sutherland/Bass absorption model is applied on the right. In the right panel, note that the energy dissipates completely at approximately 100 km, with no predicted return of acoustic energy to the surface of the earth, at any range. Thus, if the Sutherland/Bass absorption model as implemented as calculated in the modeling package, no thermospheric returns are predicted. However, this event, as well as others, was observed at IS-10 with an arrival time consistent with that of a thermospheric return; for the 29 Jan. case, the arrival was seen 12.6 sec before the predicted arrival time using ray travel times generated with NRL-G2S. The consistent observations at IS-10 from the iron mine ground truth data set suggest that the absorption model as implemented in InfraMAP should be reassessed.

IV. COMPARISON OF OBSERVED AND PREDICTED ARRIVALS

The travel times determined in the modeling section were compared to the arrival times determined from the infrasound data analysis to assess whether the event was observed and whether the travel times were correct. Figure 9 compares the number of events in the catalog provided by the mine, the number of events with predicted arrivals based on ray tracing using the InfraMAP Modeling package, the number of events observed, and the number of observed events with good *f-k* estimates. The number of observed events by month varies from 18 to 100 percent of the total catalog with the smaller percentages found in August, October, and November. The events in August were not recorded due to instrument failure. October and November either predicted arrivals beyond 20 km maximum miss distance (ideally, the miss-distance should not exceed 10 km) due to a change in the prevailing wind direction, or the arrivals were obscured by larger amplitude signals arriving from another direction during the same time period. As stated above, the determining factor as to whether an arrival was observed was the noise level and wind bursts. For four events during this year, there were no associated seismic signal observed, possibly indicating either a problem in the catalog or an event which was poorly coupled to the ground. Fourteen days in July, August, and September experienced instrument failure at IS-10, resulting in between one and all four sensors being offline during this time. If only one element was nonfunctional, then an arrival could be marked but no definitive *f-k* estimate could be resolved. If more than one element was down, no arrivals could be identified.

For arrivals that were both predicted and observed, NRL-G2S provided a more accurate predicted travel time than MSIS/HWM. As an example, the statistics for January are found in Table II, where the negative values indicate that the observed arrival is early with the respect to modeled arrival. The average travel time residual between MSIS/HWM and NRL-G2S is 1.0 min, with predicted differences from −1.4 to 0.6 min.

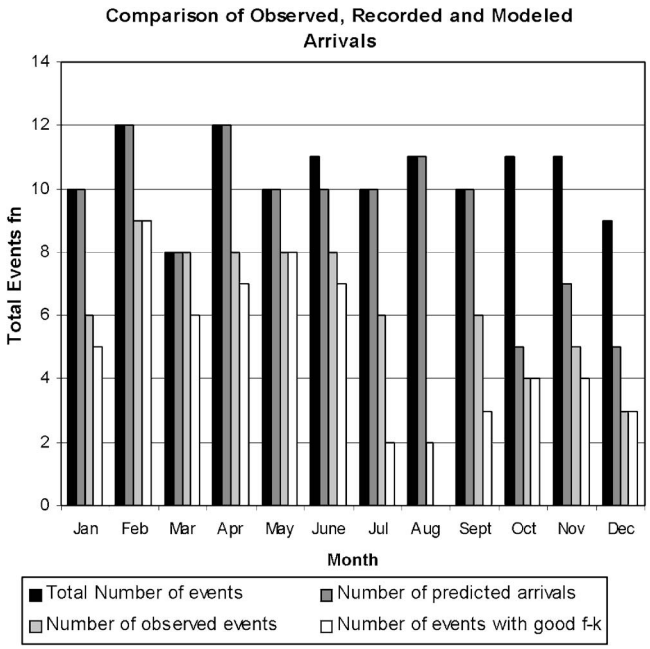


FIG. 9. Graphical representation of number of observed events compared to number predicted by the modeling, the number of recorded blasts from the mine, and the number of recorded events with good *f-k* results.

TABLE II. Travel time residuals between NRL-G2S and MSIS/HWM.

	NRL-G2S	MSIS/HWM
Average travel time	26.47 min	26.13 min
Standard deviation	0.82 min	0.215 min

While it is likely that a component of this miss-time is associated with inaccuracies in the atmospheric parameters, human error cannot be discounted. The seismic records indicate that it was common to have an average of a two minute difference in origin time from that determined from the seismic arrivals to that recorded by the mine personnel. Table III shows the difference in travel time residuals between travel times calculated with the seismic origin times and the blast record times. Using seismic origin times does not improve the miss-time and results in a larger average miss-time for each time period. This indicates that the time determined by the blasters may be only approximate since there is no electronic timing device correlated with the ignition of the explosives; blasters record the time manually from clocks. The negative recorded miss-times indicate arrivals that occurred before the predicted time. For each month, there were arrivals that occurred both before and after the predicted time, though all months indicated a trend towards early arrivals. This trend may indicate that the atmospheric profiles are predicting a travel time that is slow for much of the winter-spring season.

The ray tracing also gives an estimate of expected azimuth deviation. This value can be compared to the actual azimuth deviations determined from the observed data. The actual azimuth from source to receiver is 142°. For the month of January, the average azimuth deviation from the observed data calculated by *f-k* methods was 2.6° while the azimuth deviation predicted by InfraMAP was 4.6°. July through October has arrival azimuths with values lower than the true azimuth, where the remaining months have arrival azimuths values higher than the true azimuth. However, the 12-month average is 142.53°, a difference of 0.53° from true azimuth. This indicates a much lower variability in the actual observed azimuths than was predicted. One possible explanation is the actual horizontal winds are less than those predicted by the NRL-G2S profiles.

V. CONCLUSIONS

For the case of the hard rock iron mine in Minnesota, the observed infrasound arrivals lasted on the order of 10 sec and may consist of multiple parts. It is possible that these multiple arrivals are associated with the multiple shots per blast in the mine or the duration of blasting. The signals had a good signal-to-noise ratio between 1 and 5 Hz. Not all arrivals that were predicted were observed. This can be explained by high noise and the presence of wind bursts. There is no correlation between the signals detected and explosive amount used during each blast.

TABLE III. Travel time residuals for times predicted by the seismic records vs mining records, as modeled through NRL-G2S atmospheres for the month of January.

January 2003	Miss Time	Standard Deviation
Time residual predicted from seismic records	−2.15 min	2.49 min
Time residual predicted from mine records	−1.43 min	1.93 min

Ray trace modeling predicted all recorded blasts and should be seen within a maximum miss distance of 20 km for all months except one shot in October and all shots in November. The majority of observed arrivals were likely to be thermospheric, with one unequivocal stratospheric arrival in June. The PE results indicate that varying the dominant frequency did not change whether the arrivals were predicted though a small loss of signal strength was predicted for the higher frequencies. This correlates well with the low frequency character of the observed arrivals (0.5 to 5 Hz). The Sutherland/Bass attenuation model as formulated in the modeling software appears to be too severe to account for the observed arrivals at IS-10. The implementation of the absorption model for infrasound propagation calculations will need to be reassessed for any future work, including research into how uncertainty in the robustness of the upper atmosphere representation in the model would affect the calculated absorption rates by altering the upper turning ray heights.

Ray trace modeling produced reasonably accurate travel times, though there was more variability in the NRL-G2S travel times than the MSIS/HWM travel times and NRL-G2S more accurately matched the observed travel times. The travel time misses could be explained by either blast record error in origin time, or in inaccuracies in the atmospheric parameters. The systematic overestimation of travel times indicates that the atmospheric profiles are biased and may need to be revised or that the manually recorded blast times were biased early due to human error.

ACKNOWLEDGMENTS

A portion of this manuscript was initially presented in Mihan H. McKenna's unpublished dissertation "Infrasound Wave Propagation over Near-Regional and Tele-Infrasonic Distances." The authors would like to extend their appreciation to Travis Glenn for data preparation. Permission to publish this paper was granted by Director, Geotechnical and Structures Laboratory, U.S. Army Engineer Research and Development Center.

Bedard, A. J., and Georges, T. M. (**2000**). "Atmospheric infrasound," Phys. Today pp. 32–37.

Bowman, J. R., Baker, G. E., and Bahavar, M. (**2005**). "Ambient infrasound noise," Geophys. Res. Lett. **32**, L09803, 10.1029/2005GL022486.

Burg, J. P. (**1972**). "The relationship between maximum entropy spectra and maximum likelihood spectra," Geophysics **37**, 375–376.

Chael, E., and Whitaker, R. (**2003**). *25th seismic research review—Nuclear explosion monitoring: Building the knowledge base* (Infrasound Signal Library, Tucson, Arizona).

Che, I.-Y., Jun, M.-S., Jeon, J.-S., and Min, K. (**2002**). "Analysis of seismo-acoustic events I the Korean Peninsula," Geophys. Res. Lett. **29**, 1589.

Drob, D. P., Picone, J. M., and Garcés, M. (**2003**). "Global morphology of infrasound Propagation," Geophys. Res. Lett. **108**, 10.1029/2002JD003307.

Edlemann, G., Drob, D., Lingevitch, J., and Collins, M. (**2005**). "Modeling volcanic infrasound propagation using the parabolic equation," J. Acoust.

Soc. Am. **117**, 2421.
French, B. (**1968**). *Progressive Contact Metamorphism of the Biwabik Iron-Formation, Mesabi Iron Range, Minnesota* (University of Minnesota Press, Minneapolis, Minnesota).
Garcés, M., Hansen, R., Lindquist, K., Drob, D., and Picone, M. (**1999**). "Variability of the lower atmosphere and its effect on the propagation of infrasonic waves," *Proceedings of the 21nd Annual Seismic Research Symposium: Technologies for Monitoring the Comprehensive Nuclear-Test-Ban Treaty (CTBT)* (Las Vegas, Nevada).
Gibson, R., and Norris, D. (**2002**). *Development of an infrasound propagation modeling tool kit*, BBN Technologies Technical Report. DTRA-TR-99–47. October 2002. Defense Threat Reduction Agency, Fort Belvoir, VA, 22060-6201.
Gibson, R., and Norris, D. (**2004**). *User's Guide for InfraMAP, Version 4.0.* DTRA01–01-C-0084.
Goforth, T., Hetzer, C., and Stump, B. (**2006**). "Characteristics of regional seismograms produced by delay-fired explosions at the Minntac iron mine," Bull. Seismol. Soc. Am. **96**, 272–287.
Hagerty, M. T., Kim, W.-Y., and Martysevich, P. (**2002**). "Infrasound detection of large mining blasts in Kazakstan," Pure Appl. Geophys. **159**, 1063–1079.
Hetzer, C., and Goforth, T. (**2000**). *Seismic Studies of Delay-Fired Mining Explosions in Mountain Iron, Minnesota* (Baylor University Press, Baltimore).
Jones, M., Riley, J., and Georges, T. (**1986**). *A versatile three-dimensional Hamiltonian ray—tracing program for acoustic waves in the atmosphere above irregular terrain*. NOAA Special Report, Wave Propagation Laboratory (Boulder, Colorado).
McKisic, J. Michael (**1997**). *Infrasound and the infrasonic monitoring of atmospheric nuclear explosions: A literature review*, Phillips Laboratory Technical Report, PL-TR-97-2123, Phillips Laboratory, Hanscom AFB, MA 01731-3010.
Norris, D. (**2005**). "Parabolic equation, (PE) model approximations and implications for infrasound," J. Acoust. Soc. Am. **117**.
Picone, J. M., Hedlin, A. E., Coffey, S. L., Lean, J., Drob, D. P., Neal, H., Melendez-Alivra, D. J., Meier, R. R., and Mariska, J. T. (**1997**). *The Naval Research Laboratory Program on Empirical Models of the Neutral Upper Atmosphere*, Naval Research Laboratory Technical Report. AAS-97-632. Naval Research Laboratory, Washington, DC, 20375-5355.
Sims, P., Anderson, J., Bauer, R., Chandler, V., Hanson, G., Kalliokoshi, J., Morey, G., Mudrey, M., Ojakangas, R., Peterman, Z., Schulz, K., Shirey, S., Smith, E., Southwich, D., Van Schmus, W., and Weiblen, P. (**1993**). "The Lake Superior region and trans-Hudson orogen," edited by J. Reed, M. Bickford, R. Houston, P. Link, D. Rankin, P. Sims, and W. Van Schmus, *The Geology of North America, Vol. C-2, Precambrian-Conterminous U.S.: The Geological Society of America*, The Geological Society of America, Boulder, CO.
Smart, E., and Flinn, E. (**1971**). "Fast frequency-wavenumber analysis and fisher signal detection in real-time infrasonic array data processing," Geophys. J. R. Astron. Soc. **26**, 279–284.
Sorrells, G. G., Herrin, E. T., and Bonner, J. L. (**1997**). "Construction of regional ground truth databases using seismic and infrasound data," Seismol. Res. Lett. **68**, 743–752.
Stump, B., Hayward, C., Hetzer, C., and Zhou, R.-M. (**2002**). "Utilization of seismic and infrasound signals for characterizing mining explosions," *Proceedings of the 23nd Annual Seismic Research Symposium: World Wide Monitoring of Nuclear Explosions* (Jackson Hole, Wyoming).
Sutherland, L., and Bass, H. (**2004**). "Atmospheric absorption in the atmosphere up to 160 km," J. Acoust. Soc. Am. **115**, 1021–1032.
West, M., Gilbert, K., and Sack, R. A. (**1992**). "A tutorial on the parabolic equation, (PE) model used for long range sound propagation in the atmosphere," Appl. Acoust. **37**, 31–49.
Whitaker, R., ReVelle, D., and Sandoval, T. (**2002**). "On Infrasound Detection and Location Strategies," *24th Seismic Research Review: Nuclear Explosion Monitoring—Innovation and Integration* (Ponte Vedra Beach, Florida).

Padé approximation in time-domain boundary conditions of porous surfaces

Vladimir E. Ostashev
NOAA/Earth System Research Laboratory, Boulder, Colorado 80305 and Department of Physics, New Mexico State University, Las Cruces, New Mexico 88003

Sandra L. Collier
U.S. Army Research Laboratory, Maryland 20783-1197

D. Keith Wilson
U.S. Army Engineer Research and Development Center, Hanover, New Hampshire 03755

David F. Aldridge and Neill P. Symons
Geophysics Department, Sandia National Laboratories, Albuquerque, New Mexico 87185

David Marlin
U.S. Army Research Laboratory, White Sands Missile Range, New Mexico 88002

(Received 22 February 2007; revised 27 April 2007; accepted 28 April 2007)

Formulation and implementation of time-domain boundary conditions (TDBCs) at the surface of a reactive porous material are made challenging by the slow decay, complexity, or noncausal nature of many commonly used models of porous materials. In this paper, approaches are described that improve computational efficiency and enforce causality. One approach involves approximating the known TDBC for the modified Zwikker-Kosten impedance model as a summation of decaying exponential functions. A second approach, which can be applied to any impedance model, involves replacing the characteristic admittance with its Padé approximation. Then, approximating fractional derivatives with decaying exponentials, a causal and recursive TDBC is formulated © *2007 Acoustical Society of America.* [DOI: 10.1121/1.2743153]

PACS number(s): 43.28.En, 43.28.Js [KA] Pages: 107–112

I. INTRODUCTION

Finite-difference time-domain (FDTD) simulation is a very promising technique to account for the effects of many factors affecting sound propagation outdoors, such as terrain, obstacles, atmospheric stratification, and turbulence.[1–4] A challenging problem in such a simulation is the formulation and numerical implementation of a time-domain boundary condition (TDBC) at the ground surface. (See Ref. 5 and references therein.) Typically, models of the characteristic impedance are formulated in the frequency domain (e.g., Refs. 6–8). Due to complexity of these models, many of them do not have analytically tractable inverse Fourier transforms to the time domain. Others are noncausal (see Sec. II A) or have a very slowly decaying impulse response.[5]

The goal of the present paper is twofold. First, we describe a numerically efficient evaluation of the TDBC for the causal, modified impedance model of Zwikker and Kosten (ZK).[5] This method (described in Sec. II) is based on a direct approximation of $z(t)$, the inverse Fourier transform of the characteristic impedance $Z(\omega)$ of a porous medium, by a summation of decaying exponential functions. Here, t is time and ω is angular frequency. Second, we show how Padé approximations of $Z(\omega)$ and fractional derivatives can be used to formulate numerically efficient TDBC calculations when the function $z(t)$ is not directly available (Sec. III). As examples, we consider the modified ZK and Attenborough[6] impedance models.

II. ZWIKKER-KOSTEN MODEL AND RECURSIVE APPROXIMATIONS

A. Basic equations

Let $p(t)$ be the sound pressure at the interface between the open air and a porous material (such as the ground) and $w(t)$ be the component of particle velocity normal to this interface. The relationships between $p(t)$ and $w(t)$ and their spectral amplitudes $P(\omega)$ and $W(\omega)$ are given by

$$p(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-i\omega t} P(\omega) d\omega,$$

$$w(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-i\omega t} W(\omega) d\omega. \tag{1}$$

In the frequency domain, the boundary condition at the interface between the air and porous material directly relates $P(\omega)$ to $W(\omega)$ and is well known,

$$P(\omega) = Z(\omega) W(\omega). \tag{2}$$

In this paper, we consider locally reacting surfaces for which the characteristic impedance $Z(\omega)$ does not depend on the angle at which a sound wave is incident upon the boundary. Applying the integral operator $\int_{-\infty}^{\infty} e^{-i\omega t} d\omega$ to both sides of Eq. (2), we obtain the following TDBC relating $p(t)$ and $w(t')$:

$$p(t) = \int_{-\infty}^{\infty} w(t-t')z(t')dt', \tag{3}$$

in which

$$z(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-i\omega t}Z(\omega)d\omega. \tag{4}$$

The convolution operation in Eq. (3) complicates development and implementation of TDBCs. Many realistic impedance models do not have analytically tractable inverse Fourier transforms $z(t)$ and might have noncausal TDBCs when $p(t)$ in Eq. (3) depends on the values of $w(t_1)$ for $t_1 > t$. For example, a widely used model of Delany and Bazley[7] is noncausal.[9] The use of noncausal impedance models in time-domain calculations is problematic.

B. Time-domain boundary condition

The ZK phenomenological model[10] has been used in FDTD simulations of outdoor sound propagation, e.g., Ref. 1. For this model, the characteristic impedance Z has the following frequency dependence:[11]

$$Z(\omega) = Z_\infty\sqrt{\frac{1-i\omega\tau}{-i\omega\tau}}, \tag{5}$$

where Z_∞ and τ are two parameters. It is well known that the ZK impedance model fails to predict correct values of $Z(\omega)$ for high frequencies (or high flow resistivity). To overcome this problem, Ref. 5 suggested to use the following values of the impedance Z_∞ in the limit $\omega\tau\to\infty$ and the time constant τ:

$$Z_\infty = \frac{\varrho cq}{\Omega}, \quad \tau = \frac{\varrho q^2\gamma}{\sigma_0\Omega}, \tag{6}$$

which are different from those for the ZK phenomenological model. Here, ϱ is the density of the air, c is the sound speed in the air, γ is the ratio of specific heats for air, Ω is the porosity of the ground, q is tortuosity, and σ_0 is the flow resistivity in the limit of low frequency. The impedance model, Eq. (5), with Z_∞ and τ given by Eq. (6), is called a modified ZK impedance model. It was shown in Ref. 5 that the real and imaginary parts of $Z(\omega)$ are almost indistinguishable from those of more realistic models characterized by more parameters, over a broad frequency range.

For the modified ZK model, the function $Z(\omega)$ converts to the time domain as follows:[5]

$$z(t) = Z_\infty[\delta(t) + g(t)], \tag{7}$$

where $\delta(t)$ is the Dirac delta function, and

$$g(t) = \frac{\exp(-t/2\tau)}{2\tau}\left[I_1\left(\frac{t}{2\tau}\right) + I_0\left(\frac{t}{2\tau}\right)\right]H(t) \tag{8}$$

describes the decaying response of the porous medium. Here, I_n is the modified Bessel function of order n, and $H(t)$ is the Heaviside function, which is zero for $t<0$ and 1 otherwise. The availability of a causal, analytic, time-domain formulation, and its close agreement to other, more sophisticated models, makes the modified ZK model an attractive choice for time-domain calculations. However, the response function $g(t)$ decays very slowly: as $t^{-1/2}$ for $t \gg \tau$. This decay is a general characteristic of realistic TDBCs for dissipative porous media. In numerical evaluations,[5] it was found necessary to retain a very large number (up to hundreds) of time steps in the calculation of the convolution integral, which is a substantial computational burden both in terms of calculation time and memory. In the following, we consider a recursive algorithm that can substantially reduce the computational burden.

TABLE I. Parameters a_k and γ_k for exponential series fitted to the TDBC response function. A six-term approximation ($k=1,\ldots,6$) is shown. Fits for the full modified ZK response function are presented in the second and third columns; fourth and fifth columns are for a low-frequency approximation to the modified ZK model.

k	a_k, full ZK	γ_k, full ZK	a_k, approx.	γ_k, approx.
1	−0.555 90	0.575 59	0.279 41	0.825 97
2	0.137 51	0.011 57	−0.079 50	0.040 10
3	0.482 79	0.635 07	0.165 38	0.334 55
4	0.148 98	0.113 57	0.194 67	0.016 27
5	0.286 96	0.419 58	0.173 42	0.105 57
6	−0.000 37	2.637 62	0.531 60	2.319 09

The main idea behind the recursive algorithm is to approximate the response function $g(t)$ with a summation of decaying exponential functions. This approach was considered for the acoustic TDBC in Ref. 12 and is also well known in the electromagnetic and seismic literatures. We first define a new function $f=\tau g(t)$, which depends only on the normalized time $\bar{t}=t/\tau$,

$$f(\bar{t}) = \frac{\exp(-\bar{t}/2)}{2}\left[I_1\left(\frac{\bar{t}}{2}\right) + I_0\left(\frac{\bar{t}}{2}\right)\right]H(\bar{t}). \tag{9}$$

Next, we set

$$f(\bar{t}) \approx \sum_{k=1}^{K} a_k e^{-\gamma_k\bar{t}}H(\bar{t}), \tag{10}$$

where K is the number of terms in the approximation and a_k and γ_k are to be determined by fitting the exponential series to $f(\bar{t})$. Some example fitted values of a_k and γ_k, for $K=6$, are given in Table I. These values were generated by minimizing the root-mean-square (rms) difference between the left and right sides of Eq. (10) over the range $0\leqslant\bar{t}\leqslant 30$. The MATLAB fminsearch function (which uses a simplex search method) was used with randomized initial conditions, and the best fit from 1000 trials is presented in the table. The rms error corresponding to these values is 6.77×10^{-5}. In Fig. 1, the exponential series approximation is compared to Eq. (9).

By transformation of Eq. (10), it can be shown that the approximation for the time-domain response function is equivalent to a pole expansion of the impedance in the frequency domain,

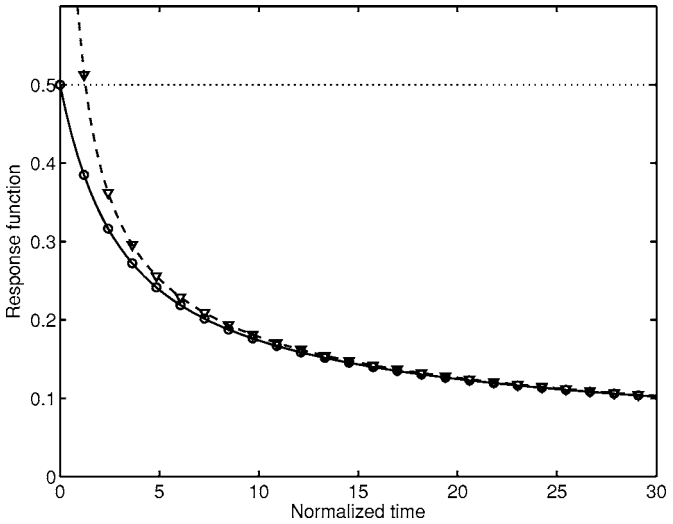


FIG. 1. Normalized response function for the porous medium, $f(\bar{t})$, where $\bar{t}=t/\tau$ is the normalized time. The response for the full modified ZK model is shown as a solid line. Circles are an approximation to this response based on a summation of decaying exponential functions. A low-frequency approximation to the modified ZK model is shown as the dashed line; triangles are the corresponding approximation by exponentials. A high-frequency approximation to the modified ZK model is shown as the dotted line.

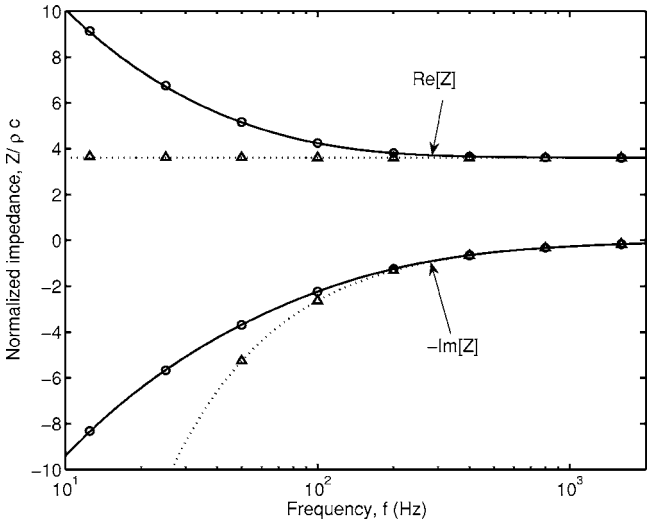


FIG. 2. Comparisons of numerical TDBC calculations to theory for the following porous material parameters: $\sigma_0=10^4$ Pa s m^{-2}, $\Omega=0.5$, and $q=1.8$. Solid line is the theoretical impedance for the full modified ZK model given by Eq. (5); circles are corresponding numerical calculations based on approximation of the time-domain response as a summation of decaying exponential functions. Dotted line is a high-frequency approximation to the modified ZK model; triangles are corresponding numerical calculations based on an integration procedure described in the text.

$$Z(\omega) \approx Z_\infty \left[1 + \sum_{k=1}^{K} \frac{a_k}{\gamma_k - i\omega\tau} \right]. \tag{11}$$

Based on the values in Table I, a direct comparison of Eq. (11) to Eq. (5) finds excellent agreement for $\omega\tau \gtrsim 10^{-2}$. Viewed another way, the reaction of the medium to an excitation with a dominant frequency $f_0=\omega_0/2\pi$ has essentially vanished within a time interval $1/\omega_0$ after the excitation has ended. (See, for example, Figs. 6 and 7 in Ref. 5.) Hence, an approximation that is valid for $\bar{t} \lesssim 100$ (such as Table I) can be reasonably applied when $\omega\tau \gtrsim 10^{-2}$. For very small $\omega\tau$, the characteristic impedance given by Eq. (11) approaches a constant real value, $Z(\omega) \approx Z_\infty(1+\Sigma_{k=1}^{K} a_k/\gamma_k) = 13.4 Z_\infty$. Although this behavior departs substantially from Eq. (5), the inaccuracies in the impedance may not be practically important, since the impedance at low frequencies is much higher than open air and almost all of the incident energy is reflected.

Substituting the exponential series, given by Eq. (10), into Eq. (7), and then into Eq. (3), we find that

$$p(t) = Z_\infty \left[w(t) + \sum_{k=1}^{K} \int_0^\infty w(t-t') a_k e^{-\gamma_k t'/\tau} \frac{dt'}{\tau} \right]. \tag{12}$$

Suppose we wish to solve Eq. (12) at a number of discrete time steps $t=t_n=n\Delta t, n=1,2,\ldots$. Let us define the auxiliary variable Ψ_k^n as the value of the integral in Eq. (12) at $t=t_n$. It can then be shown that

$$\Psi_k^n = e^{-\gamma_k \Delta\bar{t}} \Psi_k^{n-1} + \int_0^{\Delta t} w(t_n - t') a_k e^{-\gamma_k t'/\tau} \frac{dt'}{\tau}. \tag{13}$$

Here, $\Delta\bar{t}=\Delta t/\tau$. If $w(t)$ is approximated as being constant over $t_{n-1} < t \leq t_n$, the integration results in

$$\begin{aligned} \Psi_k^n &= e^{-\gamma_k \Delta\bar{t}} \Psi_k^{n-1} + \frac{a_k w(t_n)}{\gamma_k} (1 - e^{-\gamma_k \Delta\bar{t}}) \\ &\approx e^{-\gamma_k \Delta\bar{t}} \Psi_k^{n-1} + a_k w(t_n) \Delta\bar{t}. \end{aligned} \tag{14}$$

The final form assumes that $\gamma_k \Delta\bar{t}$ is small. The utility of Eq. (14) is that Ψ_k^n can be recalculated from Ψ_k^{n-1}, without knowledge of the field at earlier time steps. Substituting into Eq. (12) and solving for $w(t_n)$, we have

$$w(t_n) = \frac{1}{1 + \sum_{k=1}^{K} a_k \Delta\bar{t}} \left[\frac{p(t_n)}{Z_\infty} - \sum_{k=1}^{K} e^{-\gamma_k \Delta\bar{t}} \Psi_k^{n-1} \right]. \tag{15}$$

This result can be used in implementing TDBC in an FDTD calculation as described in Ref. 5. It replaces Eq. (37) in that article.

Results from a one-dimensional, FDTD calculation, which uses Eq. (15), are shown in Fig. 2. The calculation is basically the numerical analog of an impedance tube experiment, with a harmonic source (with a variable frequency f) at a distance of 7.25 wavelengths (in air) from a porous material. The parameters for the porous material are $\sigma_0 = 10^4$ Pa s m^{-2}, $\Omega=0.5$, and $q=1.8$, implying $\tau=1.1$ ms. The frequency range for the calculations, $12.5 \text{ Hz} \leq f = \omega/2\pi \leq 8000$ Hz, corresponds to $0.086 \leq \omega\tau \leq 11$. In the FDTD calculation, the characteristic impedance $Z(\omega)$ is deduced by observing the amplitude and phase of the pressure signal at two points outside the material. Additional details of the calculation are provided in Ref. 5. The figure shows that the exponential series approximation provides excellent results; they are nearly indistinguishable from the theory and direct evaluations of the convolutions as described in Ref. 5. Note that the method works well for small or large $\omega\tau$. It is therefore applicable to calculations involving a wide range of flow

resistivities. In particular, it can be used for outdoor sound propagation calculations above ground surfaces with a high flow resistivity (small $\omega\tau$).

C. Low- and high-frequency approximations

In many previous works, TDBCs were formulated in a limited frequency range (e.g., Refs. 13 and 14) which simplified the formulations or made them possible. Therefore, it is interesting to consider low- and high-frequency approximations to the broadband TDBC formulated above. For low frequencies ($\omega\tau \ll 1$), Eq. (5) becomes

$$Z(\omega) \approx Z_\infty \sqrt{\frac{1}{-i\omega\tau}}. \tag{16}$$

By applying the inverse Fourier transform to Eq. (2), it can be shown that

$$p(t) = \frac{Z_\infty}{\sqrt{\pi\tau}} \int_{-\infty}^{t} \frac{w(t')}{\sqrt{t-t'}} dt' = \frac{Z_\infty}{\sqrt{\tau}} D^{-1/2} w(t). \tag{17}$$

In the preceding, we have written the integral as a fractional derivative, defined as follows:[15–18]

$$D^\nu w(t) = \frac{1}{\Gamma(-\nu)} \int_{-\infty}^{t} \frac{w(t')}{(t-t')^{1+\nu}} dt'. \tag{18}$$

As before, we can approximate the response with an exponential series. Instead of Eqs. (7) and (8), we have $z(t) \approx Z_\infty g(t)$, with $g(t) \approx H(t)/\sqrt{\pi t}$. Hence, $f(\bar{t}) = \tau g(t) \approx H(t)/\sqrt{\pi \bar{t}}$. Resulting values for α_k and γ_k are given in Table I. These values were generated by minimizing the rms difference between $1/\sqrt{\pi \bar{t}}$ and the right side of Eq. (10) over the range $1 \leqslant \bar{t} \leqslant 30$; otherwise, the same numerical approach was used as before. The results, which have an rms error of 2.08×10^{-4}, are shown in Fig. 1. When this low-frequency approximation was implemented in the FDTD code, however, the TDBC became numerically unstable. Artificially increasing the real part of the impedance (so it is greater than the imaginary part) stabilized the calculation. We are uncertain why this condition seems to be necessary for stability. In any case, the low-frequency approximation is no more numerically efficient than directly approximating Eq. (7).

For high frequencies ($\omega\tau \gg 1$), Eq. (5) becomes

$$Z(\omega) \approx Z_\infty \left(1 - \frac{1}{2i\omega\tau}\right). \tag{19}$$

Taking the inverse transform, we have

$$p(t) = Z_\infty \left[w(t) + \frac{1}{2\tau} \int_{-\infty}^{t} w(t') dt' \right]. \tag{20}$$

Hence, at high frequencies, the real part of the impedance transforms to a constant (Z_∞), whereas transformation of the imaginary part produces an integral of $w(t)$. Equation (20) corresponds to $g(t) \approx H(t)/2\tau$, or $f(\bar{t}) \approx H(\bar{t})/2$. This response function is plotted in Fig. 1 as a dotted line. Discretizing the integral, with $w(t)$ approximated by its value at the end of the time step, we have

$$p(t_n) \approx Z_\infty \left[w(t_n) + \frac{\Delta\bar{t}}{2} \sum_{m=0}^{\infty} w(n\Delta t - m\Delta t) \right]. \tag{21}$$

The running sum can be calculated recursively as follows:

$$p(t_n) \approx Z_\infty \left[w(t_n) + R^n \right], \tag{22}$$

where

$$R^n = \frac{\Delta\bar{t}}{2} \sum_{m=0}^{\infty} w(n\Delta t - m\Delta t) = \frac{\Delta\bar{t}}{2} w(n\Delta t) + R^{n-1}. \tag{23}$$

A numerical evaluation of the high-frequency approximation is shown in Fig. 2. It is more computationally efficient than the direct approximation to Eq. (7), since only one memory (auxiliary) variable is required.

III. FREQUENCY-DOMAIN, PADÉ APPROXIMATIONS

The main goal of different impedance models is to capture the dependence of Z on ω for parameters describing a porous material. A small change in this dependence is insignificant since most of the models are semiempirical or semianalytical. This idea was used in Ref. 19 to express $Z(\omega)$ in the Attenborough model[6] as a Padé approximation. It was shown that the difference between the model and its Padé approximation can be made insignificantly small. In this section, we will also use a Padé approximation for the ground impedance $Z(\omega)$ appearing in Eq. (2). In principle, this approach provides a causal TDBC for *any* impedance model. For concreteness, we will consider the impedance $Z(\omega)$ for the modified ZK and Attenborough models of the ground.

A. Zwikker-Kosten model

Let $x = \sqrt{-i\omega\tau}$. Then, the inverse of the modified ZK characteristic impedance (i.e., the characteristic admittance) can be written as

$$\frac{1}{Z} = \frac{x}{Z_\infty} \frac{1}{\sqrt{1+x^2}}. \tag{24}$$

The second fraction on the right-hand side of this equation may be replaced with a Padé approximation,

$$\frac{1}{\sqrt{1+x^2}} \approx \frac{1 + a_2 x + a_3 x^2}{1 + b_1 x + b_2 x^2 + b_3 x^3}, \tag{25}$$

where a_i and b_i are coefficients to be determined. The fraction $1/\sqrt{1+x^2}$ is approximately equal to $1 - x^2/2$ and $1/x$ for small and large x, respectively. Assuming that the right-hand side of Eq. (25) is also approximately equal to $1 - x^2/2$ and $1/x$ for these limiting values of x, we obtain the following relationships between the coefficients: $a_2 = b_1, a_3 = b_3$, and $b_2 = b_3 + 1/2$. These relationships reduce the number of independent coefficients to two, say b_1 and b_3. The values $b_1 = 0.992$ and $b_3 = 0.566$ were determined numerically by the following approach. For fixed values of b_1 and b_3, let d_r be the maximum value of the real part of the difference between the left- and right-hand sides of Eq. (25) in the range $0 \leqslant \omega\tau \leqslant 10^2$. Similarly, d_{im} is the maximum value of the imaginary part of this difference. Then, the sum $d_r^2 + d_{im}^2$ was

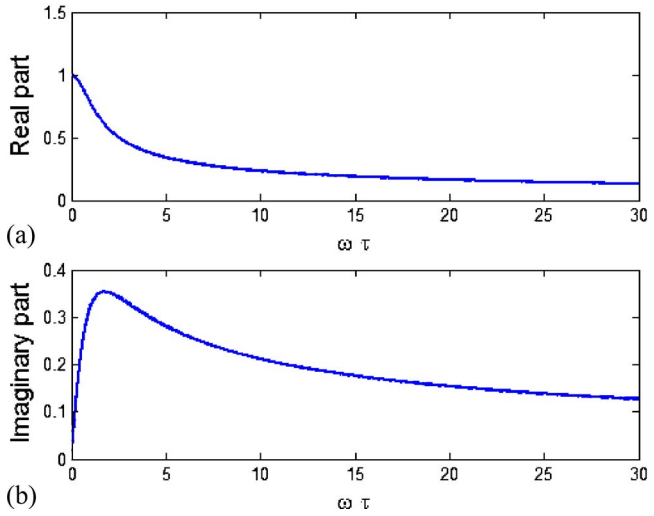


FIG. 3. (Color online) (a) The real parts of the left-hand side (solid line) and right-hand side (dashed line) of Eq. (25) vs normalized frequency $\omega\tau$. (b) The imaginary parts of the left-hand side (solid line) and right-hand side (dashed line) of Eq. (25) vs normalized frequency $\omega\tau$.

considered as a function of the coefficients b_1 and b_3. The values of these coefficients, for which the sum $d_r^2+d_{im}^2$ was minimal, were chosen as values of b_1 and b_3 in Eq. (25).

In Fig. 3(a), the real parts of the left- and right-hand sides of Eq. (25) are plotted as solid and dashed lines, respectively, and in Fig. 3(b) the imaginary parts are plotted. It follows from the figure that the left- and right-hand sides of Eq. (25) are almost indistinguishable. Note that we also tried to approximate $1/\sqrt{1+x^2}$ by a Padé approximation with order less than that in Eq. (25). However, in this case, the agreement between $1/\sqrt{1+x^2}$ and its Padé approximation was not satisfactory.

Using Eqs. (24) and (25) in the frequency domain the boundary condition, Eq. (2) can be written as

$$\frac{1}{Z_\infty}\frac{x+a_2x^2+a_3x^3}{1+b_1x+b_2x^2+b_3x^3}P(\omega)=W(\omega). \tag{26}$$

Replacing x with its value, this equation is expressed in the following equivalent form:

$$\frac{1}{Z_\infty}\left[1+a_2\frac{-i\omega\tau}{\sqrt{-i\omega\tau}}-a_3i\omega\tau\right]P(\omega)=\left[\frac{1}{\sqrt{-i\omega\tau}}+b_1-b_2\frac{i\omega\tau}{\sqrt{-i\omega\tau}}-b_3i\omega\tau\right]W(\omega). \tag{27}$$

Applying the integral operator $(1/2)\pi\int_{-\infty}^{\infty}e^{i\omega t}d\omega$ to both sides of Eq. (27) and taking into account that the term $-i\omega$ in the integrand can be replaced with the term d/dt in the front of the integral, we have

$$\frac{1}{Z_\infty}\left[\left(1+a_3\tau\frac{d}{dt}\right)p(t)+a_2\tau^{1/2}\frac{d}{dt}\frac{1}{2\pi}\int_{-\infty}^{\infty}e^{-i\omega t}\frac{P(\omega)}{\sqrt{-i\omega}}d\omega\right]=\left(b_1+b_3\tau\frac{d}{dt}\right)w(t)+\left(\tau^{-1/2}+b_2\tau^{1/2}\frac{d}{dt}\right)\frac{1}{2\pi}\times\int_{-\infty}^{\infty}e^{-i\omega t}\frac{W(\omega)}{\sqrt{-i\omega}}d\omega. \tag{28}$$

The use of Eq. (18) allows us to write Eq. (28) in the following form:

$$\frac{1}{Z_\infty}\left[1+a_3\tau\frac{d}{dt}+a_2\tau^{1/2}\frac{d}{dt}D^{-1/2}\right]p(t)=\left[b_1+b_3\tau\frac{d}{dt}+\left(\tau^{-1/2}+b_2\tau^{1/2}\frac{d}{dt}\right)D^{-1/2}\right]w(t). \tag{29}$$

This is the desired causal TDBC for the modified ZK model.

An algorithm for calculation of the fractional derivative can be developed from the Grünwald formula:[15–17]

$$D^\nu w(t)=\lim_{\Delta t\to 0}\frac{1}{(\Delta t)^\nu}\sum_{m=0}^{\infty}\frac{\Gamma(m-\nu)}{\Gamma(-\nu)\Gamma(m+1)}w(t-m\Delta t). \tag{30}$$

In a numerical calculation, the limit $\Delta t\to 0$ is approximated with a finite time step. For $m\gg 1$, the ratio $\Gamma(m+1/2)/\Gamma(m+1)\sim m^{-1/2}$. This parallels the slowly decaying time-domain impedance ($t^{-1/2}$) observed in Sec. II. As suggested by Liebler *et al.*,[16] however, a recursive algorithm can be developed by approximating the ratio $\Gamma(m+1/2)/\Gamma(m+1)$ with an exponential series,

$$\frac{\Gamma(m+1/2)}{\Gamma(m+1)}\approx\sum_{k=1}^{K}b_ke^{-\lambda_km}. \tag{31}$$

Here, b_k and λ_k are the coefficients to be determined by matching the left- and right-hand sides of this equation to a good accuracy (10^{-4} in Ref. 16) for different values of m ranging from 1 to several thousand (depending on a signal duration). Following the same procedure as earlier, we define the auxiliary variable

$$\Phi_k^n=\sum_{m=0}^{\infty}b_ke^{-\lambda_km}w(n\Delta t-m\Delta t)\sqrt{\Delta\bar{t}}, \tag{32}$$

from which we find that

$$\Phi_k^n=b_kw(n\Delta t)\sqrt{\Delta\bar{t}}+e^{-\lambda_k}\Phi_k^{n-1}. \tag{33}$$

Hence, only K levels in time must be stored and recursively updated to evaluate each partial derivative.

B. Attenborough model

It was suggested in Ref. 19 to use the following Padé approximation for the characteristic impedance Z in the Attenborough model[6] of rigid-frame porous media:

$$\frac{1}{Z}=\frac{x}{Z_0}\frac{1+a_2x+a_3x^2+a_4x^3}{1+b_1x+b_2x^2+b_3x^3+b_4x^4}. \tag{34}$$

Here, numerical coefficients a_i and b_i are determined in Table I from Ref. 19, $Z_0=q\varrho c/(\Omega\sqrt{8\gamma})$, and $x=\sqrt{-i\omega\tau}$, where $\tau=2\varrho q^2/(\Omega\sigma s_p^2)$. Furthermore, σ is the flow resistivity and s_p is the pore shape factor. Let us substitute the value of $1/Z$ given by Eq. (34) into Eq. (2). Then, using an approach similar to that to derive Eq. (29), we obtain the following causal TDBC:

$$\frac{1}{Z_0}\left[1 + a_3\tau\frac{d}{dt} + \left(a_2\tau^{1/2}\frac{d}{dt} + a_4\tau^{3/2}\frac{d^2}{dt^2}\right)D^{-1/2}\right]p(t)$$
$$= \left[b_1 + b_3\tau\frac{d}{dt} + \left(\tau^{-1/2} + b_2\tau^{1/2}\frac{d}{dt} + b_4\tau^{3/2}\frac{d^2}{dt^2}\right)D^{-1/2}\right]w(t). \quad (35)$$

The TDBCs for the modified ZK and Attenborough model given by Eqs. (29) and (35) could be implemented in FDTD simulations of outdoor sound propagation and situations involving porous materials. For other models of the characteristic impedance Z, causal TDBCs can be derived similarly. The computational burden would be similar to the approach in Sec. II, which was based on a formula for the time-domain impedance.

IV. CONCLUSIONS

In this paper, we have discussed two methods for efficiently formulating and evaluating causal TDBCs for sound propagation near an impedance surface. Both are based on approximating response functions for the porous medium with summations of decaying exponential functions. This procedure leads to recursive solutions. One approach involved analytical determination and subsequent approximation of a time-domain impedance model, namely the modified ZK impedance model. The second approach could be applied to any impedance model. It is based on Padé approximations of the frequency-domain impedance. Then, using fractional derivatives, a causal TDBC is formulated. The fractional derivatives can be efficiently calculated numerically using a recursive algorithm suggested in Ref. 16. Here, we used this approach to formulate TDBCs for the modified ZK and Attenborough models, Eqs. (29) and (35), respectively. Causal TDBCs can be similarly derived for other impedance models.

[1]E. Salomons, R. Blumrich, and D. Heimann, "Eulerian time-domain model for sound propagation over a finite-impedance ground surface. Comparison with frequency-domain methods," Acust. Acta Acust. **88**, 483–492 (2002).

[2]T. V. Renterghem and D. Botteldooren, "Numerical simulation of the effect of trees on downwind noise barrier performance," Acust. Acta Acust. **89**, 764–778 (2003).

[3]V. E. Ostashev, D. K. Wilson, L. Liu, D. F. Aldridge, N. P. Symons, and D. H. Marlin, "Equations for finite-difference, time-domain simulation of sound propagation in moving inhomogeneous media and numerical implementation," J. Acoust. Soc. Am. **117**, 503–517 (2005).

[4]L. Liu and D. G. Albert, "Acoustic pulse propagation near a right-angle wall," J. Acoust. Soc. Am. **119**, 2073–2083 (2006).

[5]D. K. Wilson, S. L. Collier, V. E. Ostashev, D. F. Aldridge, N. P. Symons, and D. H. Marlin, "Time-domain modeling of the acoustic impedance of porous surfaces," Acust. Acta Acust. **92**, 965–975 (2006).

[6]K. Attenborough, "On the acoustic slow wave in air-filled granular media," J. Acoust. Soc. Am. **81**, 93–102 (1987).

[7]M. E. Delany and E. N. Bazley, "Acoustical properties of fibrous absorbent materials," Appl. Acoust. **3**, 105–116 (1970).

[8]D. K. Wilson, "Simple, relaxational models for the acoustical properties of porous media," Appl. Acoust. **50**, 171–188 (1997).

[9]K. V. Horoshenkov and S. N. Chandler-Wilde, "On the behaviour of some impedance models for the acoustic properties of rigid frame porous media," in *Proceedings of the 6th International Congress on Sound and Vibration*, pp. 653–660, Copenhagen, July 1999.

[10]C. Zwikker and C. W. Kosten, *Sound Absorbing Materials* (Elsevier, New York, 1949).

[11]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**, 173–200 (2007).

[12]S. L. Collier, D. K. Wilson, V. E. Ostashev, and D. F. Aldridge, "Time domain equations for sound propagation over rigid, porous ground surfaces," in *Proceedings of 11th International Symposium on Long Range Sound Propagation*, Fairlee, VT, 1–3 June (2004), p. 46–68.

[13]K. Heutschi, M. Horvath, and J. Hofmann, "Simulation of ground impedance in finite difference time domain calculations of outdoor sound propagation," Acust. Acta Acust. **91**, 35–40 (2005).

[14]C.-H. Lam, B. J. Kooij, and A. T. D. Hoop, "Impulsive sound reflection from an absorptive and dispersive planar boundary," J. Acoust. Soc. Am. **116**, 677–685 (2004).

[15]K. S. Miller and B. Ross, *An Introduction to the Fractional Calculus and Fractional Differential Equations* (Wiley, New York, 1993).

[16]M. Liebler, S. Ginter, T. Dreyer, and R. E. Riedlinger, "Full wave modeling of therapeutic ultrasound: Efficient time-domain implementation of the frequency power-law attenuation," J. Acoust. Soc. Am. **116**, 2742–2750 (2004).

[17]K. B. Oldman and J. Spanier, *The Fractional Calculus* (Academic, New York, 1974).

[18]Z. E. A. Fellah and C. Depollier, "Transient acoustic wave propagation in rigid porous media: A time-domain approach," J. Acoust. Soc. Am. **107**, 683–688 (2000).

[19]S. N. Chandler-Wilde and K. V. Horoshenkov, "Padé approximants for the acoustical characteristics of rigid porous media," J. Acoust. Soc. Am. **98**, 1119–1129 (1995).

Influence of scattering, atmospheric refraction, and ground effect on sound propagation through a pine forest

Michelle E. Swearingen[a] and Michael J. White
U.S. Army Engineer Research and Development Center, Construction Engineering Research Laboratory, Champaign, Illinois 61822

(Received 17 November 2006; revised 3 April 2007; accepted 3 April 2007)

Sound propagation through a forest is affected by the microclimate in the canopy, scattering by trunks and stems, and ground reflection. Each of these effects is such a strong contributor to the attenuation of sound that mutual interactions between the phenomena could become important. A sound propagation model for use in a forest has been developed that incorporates scattering from trunks and branches and atmospheric refraction by modifying the effective wave number in the Green's function parabolic equation model. The ground effect for a hard-backed pine straw layer is approximated as a local reaction impedance condition. Comparisons to experimental data are made for frequencies up to 4 200 Hz. Cumulative influences of the separate phenomena are examined. The method developed in this paper is compared to previously published methods. The overall comparison with spectral transmission data is good, suggesting that the model captures the necessary details. [DOI: 10.1121/1.2735108]

PACS number(s): 43.28.En, 43.28.Js, 43.28.Bj, 43.20.El [VEO] Pages: 113–119

I. INTRODUCTION

The acoustics of forests has been a topic of study since Eyring[1] studied the Panamanian rain forest in 1946. Much of the work prior to the early 1980's was purely experimental. After that point, various attempts were made to define the unique propagation characteristics within a forest. Bullen and Fricke[2] modeled propagation through a large number of trees as a classical diffusion problem. They derived a general differential equation and compared the predictions to a small scale model study and to field measurements. The model worked reasonably well, but had significant restrictions in tree diameter and spacing. Martens and Huisman[3] developed a ray tracing program called RAYFLUX that included the effects of vegetation. It worked reliably for windless, isothermal conditions. The vegetative absorption was based on measurements. Price, Attenborough, and Heap[4] developed a point to point model that incorporated scattering via Twersky's model, although the number density of trees was reduced by 60% to obtain a better fit to data. This model assumed an arbitrary, but range dependent, cutoff frequency for ground effects. Below this frequency (2 kHz for ranges less than 40 m), the scattering and ground contributions were added arithmetically. Above this cutoff, only scattering contributions were considered. This worked reasonably well in predicting the shape of the attenuation spectrum, but did not accurately predict the levels. The model was compared to three different forest stands. Huisman and Attenborough[5] used a stochastic particle bounce model to simulate the multiple scattering present in the forest. Because this failed to predict the strong height dependence shown in their attenuation measurements, they concluded that the interaction between ground effect and trunk scattering is more complicated than a simple addition can account for. In 2002, Defrance *et al.*[6] published a diffusion model based on Felbacq's theory of diffusion for sound propagation through forests. They incorporated tree trunk scattering into the Green's function parabolic equation (GFPE). This was done by computing attenuation from tree scattering using the Twersky method and subtracting this value from the calculated spectral levels at specific ranges. Realistic sound speed profiles were used. Results compared satisfactorily to experimental data.

This paper describes a model for sound propagation through a forest, also based on the GFPE. However, the approach to the scattering contribution is different. Instead of using the path attenuation method of Defrance *et al.*, this paper incorporates the scattering effect by modifying the effective bulk wavenumber to include an imaginary part that provides attenuation due to multiple scattering. This attenuation comes from the multiple scattering approach described by Linton and Martin.[7] Section II describes each component of the model. The next section briefly describes measurements taken in a pine forest. Section IV compares the model output to the field data. Finally, the cumulative contributions of each component of the model are examined, and model results are compared to previously published methods for determining the scattering contribution.

II. THE FOREST MODEL

The model described in this paper combines several different physical effects acting upon the wave field in a forest. The wave field computations are accomplished by the GFPE, a numerical solver of a parabolic equation, with axial symmetry and outgoing waves assumed. Since the GFPE method is well described in the literature, the derivation will not be included here. Instead, the reader is referred to Gilbert and Di[8] and Salomons[9], two excellent papers describing the GFPE. In this model, the trunk and canopy scattering are incorporated into the GFPE by modifying the effective wave

[a]Electronic mail: michelle.e.swearingen@erdc.usace.army.mil

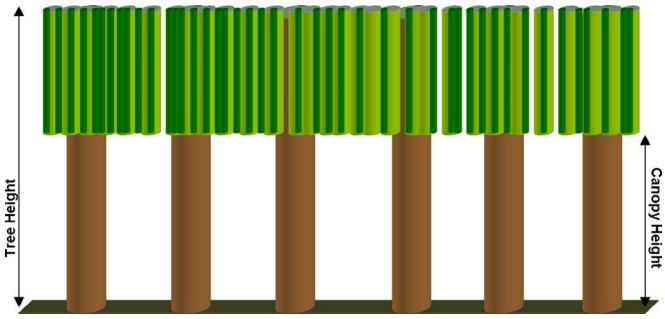


FIG. 1. (Color online) Geometry of the model. Trunks extend to the tree height. Two different sizes of cylinders are used to construct the canopy.

number so that it includes attenuation due to scattering. The following sections will describe the geometry, vertical sound speed profile, ground impedance, and trunk and canopy scattering.

A. Geometry

The geometry of this model represents an idealized pine forest. An even-aged stand of pine trees planted in a slightly randomized square-lattice pattern with negligible undergrowth and a well-defined forest canopy is assumed. Figure 1 is a graphical representation of this assumed geometry. For consideration of wave scattering, the pine trees are modeled as an assortment of truncated right-circular cylinders. Trunk cylinders are all perpendicular to the ground and extend to the tree height. The canopy is also made up of cylinders, which are oriented parallel to the trunks for conceptual simplicity. They extend from the canopy height to the tree height. Only two estimated branch sizes are considered. The ground surface is assumed to be flat, level, and homogeneous. Propagation begins in the forest and travels only through the forest; no edges are present. The actual quantities associated with each parameter, such as the number of trees per unit area and trunk and branch diameters, are treated as variables to be initialized at the beginning of a computational realization.

B. Atmospheric variation with height

The presence of the forest canopy serves as a target for solar heating, causing the ground to be partly shaded. In the absence of solar heating, radiative cooling within the canopy is reduced as well, because the upper canopy is at a much higher temperature than the open night sky. The presence of trees appears to the wind as additional roughness elements, increasing the drag. Because leaf areas, heat capacities, and transpiration rates of the various surfaces differ widely between tree types, the detailed microclimate within the forest is complex. Tunick[10] has developed a two-dimensional model for this problem, which is used here to generate the vertical effective sound speed profiles.

Tunick's model requires several inputs regarding the base weather conditions and the forest characteristics. The base weather condition inputs include cloud amount, height of cloud base, latitude and longitude of location, Julian day number, GMT time, maximum wind speed at three times the forest height, canopy top temperature, relative humidity, and surface pressure. Forest characteristic inputs include tree and canopy heights, leaf area index, canopy clumping factor, average leaf size, leaf drag coefficient, and tree type.

C. Ground impedance

Ground impedance under the tree canopy is typically affected most by the detritus composing the surface layer. The forest floor reflection is frequently represented as that from a hard-backed porous layer (see, for example, Price *et al.*[4] and Huisman and Attenborough[5]). In this particular case, Attenborough's[11] hard-backed layer impedance two parameter model (2PA) has been utilized, in which porosity decreases exponentially with depth. In this method, the ratio of surface impedance to that for the air of a porous layer of depth d_f in meters overlying a perfect reflector can be expressed as

$$Z(d_f) \simeq Z_1 + ic/(2\gamma\omega\Omega_1 d_f), \tag{1}$$

where c is the sound speed at the surface (m/s), γ is the ratio of specific heats for air, $\omega=2\pi f$, f is frequency (Hz), and Ω_1 is the porosity of the uppermost part of the layer. The relative characteristic impedance is

$$Z_1 = \left(\frac{4T}{3\Omega_1} + \frac{4i\Omega_1\sigma_e}{\omega\rho}\right)(k_g)^{-1}, \tag{2}$$

where T is tortuosity, ρ is the density of air (kg/m^3), and the propagation constant normalized with respect to that in air is

$$k_g = \sqrt{\gamma[aT + 4i\Omega_1^2\sigma_e/(\omega\rho)]}, \tag{3}$$

with σ_e as the effective flow resistivity (Pa s m^{-2}), and

$$a = \frac{4}{3} - [(\gamma-1)/\gamma]N_{Pr}, \tag{4}$$

where N_{Pr} is the Prandtl number. Note that for sound absorption to occur within the layer the time dependence $\exp(-i\omega t)$ must be assumed and the square root in Eq. (3) must select the branch with the positive imaginary part.

D. Trunk and canopy scattering

The trunk and canopy scattering are both based on Linton and Martin's[7] multiple scattering model. This formula for the bulk sound wave number k_b assumes a dilute random array of identical scatterers and utilizes Lax's[12] quasicrystalline approximation. The following equation is a discretized and simplified version of Eq. (86) in Linton and Martin:[7]

$$k_b^2 = k^2 + 4iN\sum_{n=-\infty}^{\infty} A_n + \frac{8N^2}{k^2}\sum_{n=-\infty}^{\infty}\sum_{s=-\infty}^{\infty} A_n A_s (n-s)\mathrm{sgn}(n-s), \tag{5}$$

where $k=\omega/c$, c is the sound speed in the medium, N is the number of scatterers per unit area (for example, trees/m^2), and

$$A_n = \frac{qJ_n'(ka)J_n(\kappa a) - J_n(ka)J_n'(\kappa a)}{qH_n'(ka)J_n(\kappa a) - H_n(ka)J_n'(\kappa a)} = A_{-n}. \quad (6)$$

In Eq. (6), $\kappa = \omega/\tilde{c}$ and $q = \tilde{\rho}\tilde{c}/(\rho c)$, where $\tilde{\rho}$ and $\tilde{c}$ are the density and sound speed, respectively, inside the cylinders. The cylinder radius is a. J_n is the Bessel function and H_n is the Hankel function of the first kind, and J_n' and H_n' are their first derivatives with respect to the argument. Note that Linton and Martin use Z_n instead of A_n for the scattering coefficients. Because the trees seem effectively rigid, it is tempting to use the rigid cylinder approximation to Eq. (6). For example, $\tilde{\rho} = 450$ kg/m^3 and $\tilde{c} = 3500$ m/s for pine.[13] These values result in a change in k_b of less than 8×10^{-10} for trunk diameters, indicating that the rigid cylinder approximation $Z_n \to J_n'(ka)/H_n'(ka)$ and $q \to \infty$ is appropriate.

For both the trunk and canopy, the scattering contributions assume the approximation of scattering from infinite length cylinders using Eq. (5), scaled appropriately for diameter and number density of scatterers. These scattering contributions are only applied in the layer containing the scatterers. The trunk layer extends from the ground to the tree height. The canopy layer extends from the canopy height to the tree height and includes large and small branches scaled in size and number from the tree trunk parameters. Scale variables relate the areal number density and the diameter of limbs to the respective areal trunk density and diameter. The canopy cylinders are assumed to be aligned parallel to the trunk cylinders. Because local sound speed is utilized in these calculations, the scattering contributions can vary slightly with height. In the numerical realization of the forest, the measured value of the areal number density of trunks is used. The number and size of large and small limbs are estimates based on the number and size of trunks. Pine needle scattering contributions are neglected.

In practice, the infinite sums in Eq. (5) need to be truncated. These are truncated by the condition $|n| \leqslant 2ka + 20$. A sum confined to $|n| \leqslant 2ka$ is typically sufficient, but the additional terms are included to ensure that the sum has converged.

III. THE FIELD EXPERIMENT

A field experiment was conducted to examine acoustic propagation in a field versus a forest. Two different weights of Composition C-4 explosive, 1.25 and 5 lb, were suspended 2 m above the ground at several different locations and detonated individually. Microphones were located at a height of 1.2 m above the ground. For the purposes of this paper, only the source points labeled TC3 and TC4 in Fig. 2 will be considered. Two microphone locations, 174 and 315 m into the forest, will be used to compare experimental data to model results. These correspond to the path from TC4 to WA3 and TC3 to WA7, respectively. Microphones at these locations were Brüel and Kjaer (B&K) Type 4143 half-inch condenser microphones with Type 2639 preamplifiers and Type 2804 power supplies. Signals were recorded on a Sony PC208A eight-channel digital audio tape (DAT) recorder in eight-channel mode with 2× recording speed. The effective

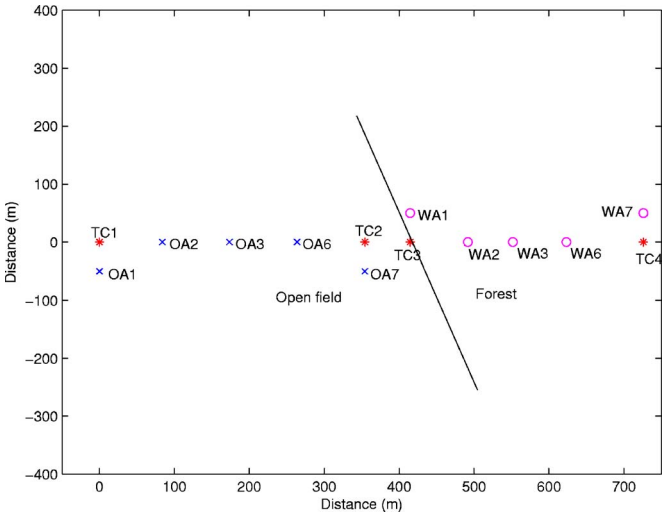


FIG. 2. (Color online) Layout schematic of the experimental setup. (*) denotes a source point, (×) denotes a microphone in the open field, and (○) denotes a microphone in the forest.

sampling rate was 24 kS/s per channel. For a complete description of the experimental setup and results, see Swearingen and White.[14]

The forest used in this experiment was predominantly slash pine. There was little growth in the understory. The ground had a layer about 3 cm thick of fine litter, with hard-packed sand underneath. The average forest tree height was 15.8 m and the average canopy height was 10 m. Trunks averaged 0.096 m in diameter and the average stem areal density (number of stems per unit area) was 0.0124 trees/m^2. The (trunk:large branch:small branch) diameters were estimated to be in the ratio of $1:\frac{1}{2}:\frac{1}{8}$ and the areal densities followed the ratio of 1:6:30. As indicated in Sec. II D, the trunks and branches are assumed to be rigid cylinders. The source and receiver heights were 2 and 1.2 m, respectively. Detailed meteorological measurements were not taken during this field experiment. However, local conditions at the time of the experiment were clear skies, no discernible wind, and temperatures around 32–36 °C and 50–60 % relative humidity. Estimated profiles are generated using Tunick's model (described above).

IV. COMPARISONS OF FOREST GFPE TO DATA

Comparing theoretical predictions to field data can lead to interesting observations about both the prediction and the field test. The measured data are presented in 1/3-octave band sound exposure level (SEL) spectra (dB). Predictions are formed from adding the Friedlander[15] source sound exposure spectrum to the forest GFPE calculated transmission loss in decibels for the indicated frequency. The Friedlander spectrum approximates the source strength of Composition C-4 explosions. The data and model are presented in 1/3-octave band SELs. The starting field used to initialize the forest GFPE follows that in Cooper.[16]

Several inputs are required to create the meteorological profiles using Tunick's[10] two-dimensional (2D) model. For determining solar input, the cloud amount was set equal to 0.2, cloud base at 1 km, latitude and longitude were 33.4°N,

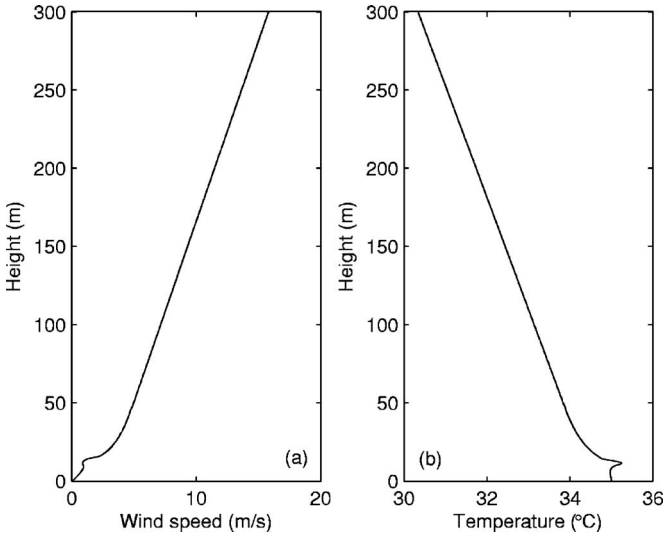


FIG. 3. Temperature and wind profiles used in calculations.

94.1°W, on day 205 at 13 h local time, retarded 5 h re GMT. The influence of the canopy on the microclimate was parametrized by tree height equal to 15.8 m, leaf area index of 3, clumping factor of 0.6, typical pine needle dimensions equal to 1 cm by 15 cm, and a coefficient of drag set to 0.15. Leaf area index defines the relative density of leaves with respect to height within the canopy. The clumping factor describes the heterogeneous distribution of leaves on tree branches. The maximum wind speed at $z=47.4$ m was set to 4.8 m/s, the canopy top temperature was 35.0 °C, with a 50% relative humidity and a surface pressure equal to 101 900 Pa. The resulting wind and temperature profiles are presented in Fig. 3. The corresponding effective sound speed profiles are shown in Fig. 4 for upwind and downwind sound propagation.

Further inputs describing the physical forest structure are necessary in order to run the forest GFPE. For determining the ground reflection properties, Attenborough's[11] 2PA model was used. The effective flow resistivity (σ_e) was set to

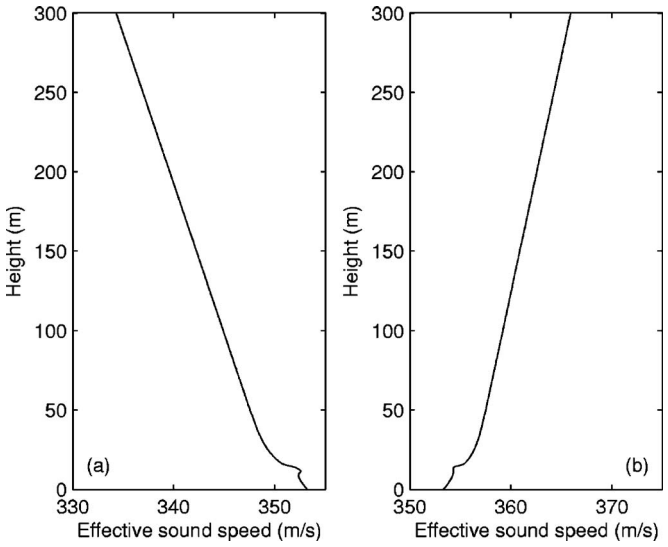


FIG. 4. Effective sound speed profiles used in calculations: (a) upwind and (b) downwind.

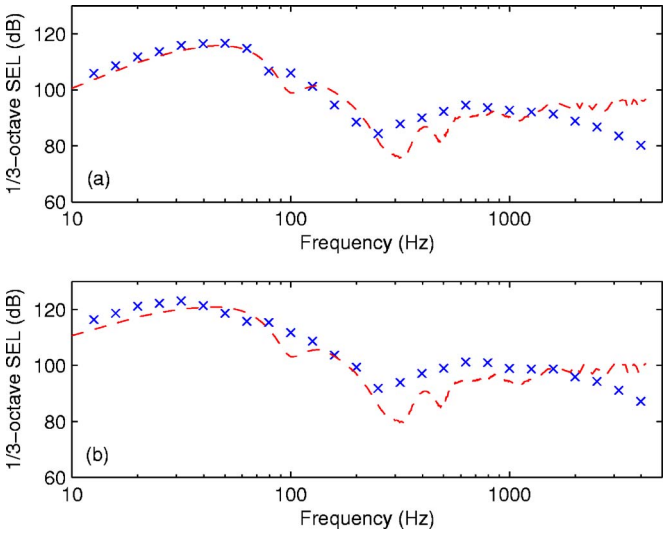


FIG. 5. (Color online) Comparison of experimental data and model results at 174 m from the source. (×) is the data and (- -) is the upwind case. The source is 1.25 lb Composition C-4 in (a) and 5 lb Composition C-4 in (b).

7 500 Pa s m^{-2}, the layer depth was 0.03 m, and the tortuosity and porosity were set to $T=2$ and $\Omega=0.4$, respectively. Values for σ_e and T are taken from Table II in Attenborough,[11] Ω is an estimate, and the layer depth is measured. To suppress reflections from the upper boundary in the PE, the bottom of the absorbing layer was set to 300 m and the top of the absorbing layer was set at 1024 m. The calculational area of interest extends from the ground to a height of 300 m.

Figure 5 shows the comparison between the forest GFPE calculations and the measured forest propagation data at 174 m from the source. Overall winds during the experiment were light but variable in the cloud-free sky during midday. This type of condition generally indicates an upward refracting situation. Therefore, the upwind condition shown in Fig. 4 is used for comparisons to the experimental data. Agreement between measurement and model realizations is reasonable at this distance. The ground dip is accurately predicted around 250–300 Hz. The model diverges in shape from the measurement at frequencies above 1 100 Hz, indicating that the scattering algorithm is not accurately predicting the attenuating value of trunks and limbs at those higher frequencies. It should be noted that in order to resolve and present the acoustic field for the short propagation distance used in the comparison it was necessary to decrease the horizontal range step size to $\lambda/2$ for frequencies below 100 Hz and to λ for frequencies below 400 Hz. At all higher frequencies, the typical GFPE horizontal range step of 6λ was used.

Figure 6 shows the same comparison as above at a distance of 315 m. At this distance, the agreement is not as good. The ground dip is not accurately modeled, and again the shape of the curve diverges from the experimental data above 1 100 Hz. The sharp dips near 95 and 1 000 Hz are caused by interference between the refracted and ground bounce components of the field.

V. CUMULATIVE COMPONENT CONTRIBUTIONS

Because the forest GFPE incorporates all of the different components into the differential equations for the boundary

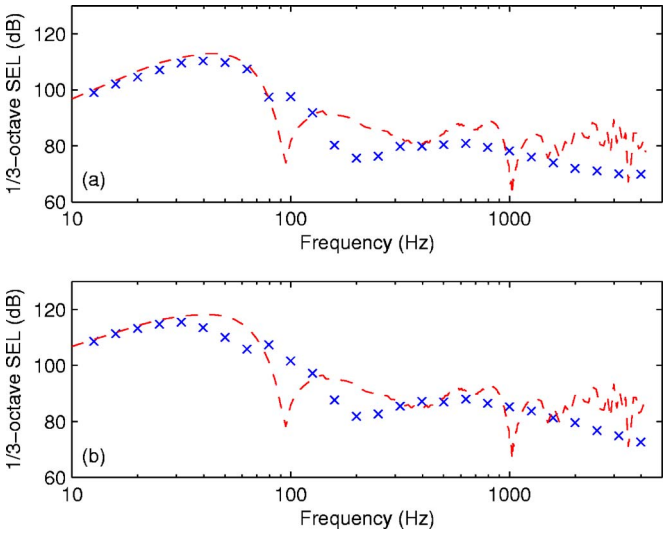


FIG. 6. (Color online) Same as Fig. 5, but for a distance of 315 m between source and receiver.

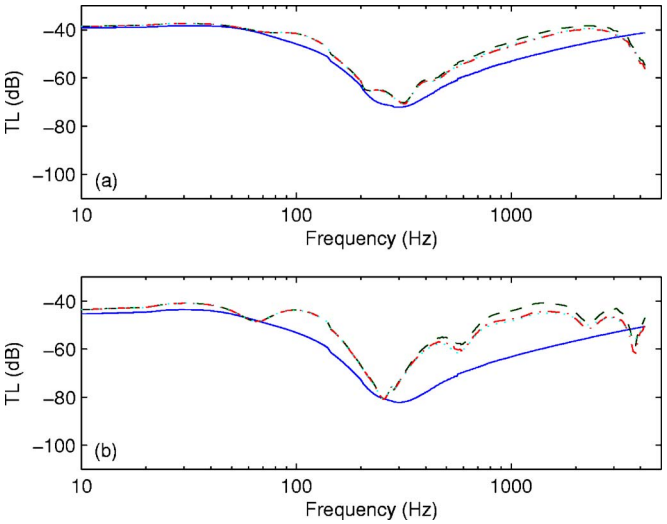


FIG. 8. (Color online) Same as Fig. 7, but for the case of downwind propagation.

conditions and sound field in a unified manner, it is interesting to examine the cumulative contributions of the components. In all cases, the basic parameters used are identical to those used in the field data comparison, although this time both upwind and downwind cases are examined. Figures 7 (upwind) and 8 (downwind) show transmission loss in dB for several realizations of the same set of parameters, with each component added in one at a time. The forest ground impedance used in Sec. IV is assumed in all cases. The progression is as follows: (1) homogeneous atmosphere, (2) forest atmosphere (upwind or downwind), (3) trunk scattering, and (4) canopy scattering. In this progression, all previous components are maintained in the next realization, so that by the time the canopy scattering is added, all components are present. Thus, the progression of curves goes as (1), (1)+(2), (1)+(2)+(3), and (1)+(2)+(3)+(4). As can be seen,

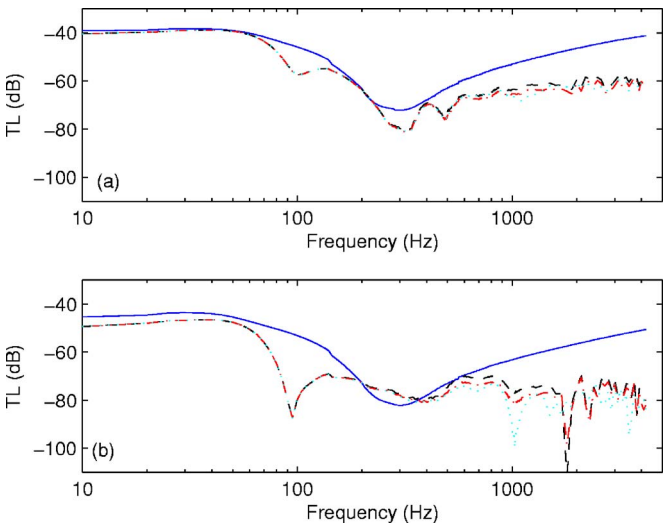


FIG. 7. (Color online) Cumulative contributions of components, upwind case. In each graph, (—) is a homogeneous atmosphere with forest ground impedance. (- -) adds the upwind forest profile, (-·) adds trunk scattering, and (···) adds canopy scattering. The source to receiver distance is 174 m in (a) and 315 m in (b).

the atmospheric profile has a profound influence on the propagation; the trunk scattering affects frequencies primarily above 400 Hz. In the upwind case, the canopy scattering generally contributes to the overall attenuation, but also causes variation in the shape of the curve, particularly at the 315 m distance. In the downwind case, the canopy has almost no contribution to the overall attenuation. These realizations are shown at 174 and 315 m. Trunk and canopy scattering become increasingly significant with increasing distance.

VI. COMPARISON TO PREVIOUS METHODS

The forest GFPE presented in this paper deviates from the historical usage of Twersky's[17] multiple scattering method to account for trunk scattering. Historically, the best fit to data occurs when the number of trees per unit area is taken at 40% of the actual measured value for modeling purposes.[4] However, this is a nonphysical representation of the actual environment, indicating that the Twersky method overemphasizes the scattering attenuation. Linton and Martin's[7] paper states that the scattering attenuation coefficient of Twersky is based on an inappropriate choice of pair-correlation function. Using 100% of the number density of trees with the Linton-Martin method yields satisfactory results.

In order to compare to previous methods such as Price[4] and Defrance *et al.*,[6] two types of calculation methods are performed. In method 1, the scattering attenuation is included in the forest GFPE as an attenuation coefficient incorporated into a complex value wave number, substituting k_b for k. This is the method developed in this paper. In method 2, the bulk attenuation for the scattering medium is added to the propagation through a medium free of scatterers but involving the forest atmospheric profiles and ground conditions. The attenuations at the two distances 174 and 315 m are now examined. Canopy scattering is neglected in this examination. In method 1, scattering attenuation is applied only to waves within the forest layer. Method 2 assumes that the tree trunks extend from the ground to the top of the

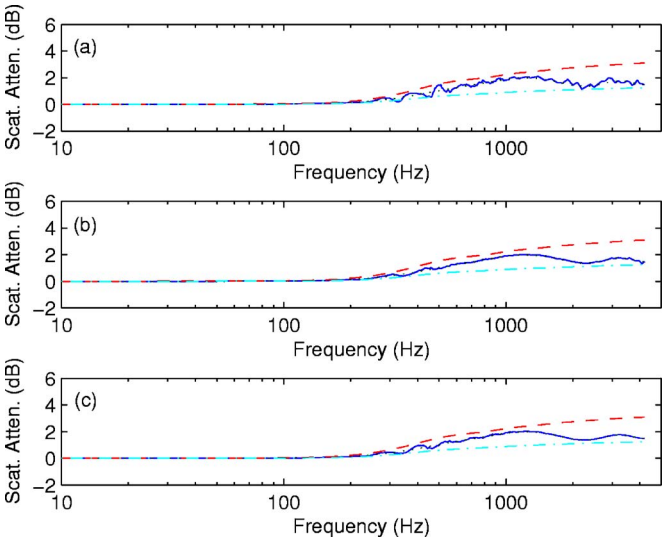


FIG. 9. (Color online) Comparison between method 1 and method 2 (described in the text) for a 174 m propagation distance. Graphs are of scattering attenuation calculated by method 1 and scattering attenuation calculated by method 2, in decibels. Method 1 is represented by (—). The three method 2 scattering attenuation spectra are calculated using Linton-Martin (···), Twersky with 100% of the trees (- -), and Twersky with the usual 40% of the trees (-·). The refraction case is upwind in (a), downwind in (b), and homogeneous in (c).

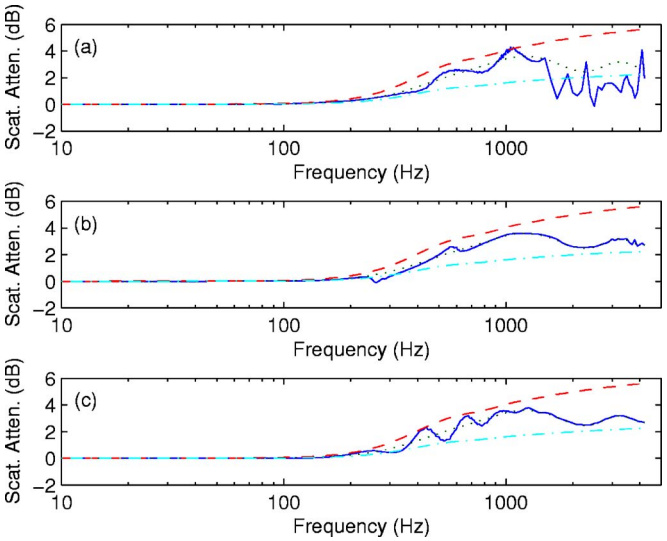


FIG. 10. (Color online) Same as Fig. 9, but for a propagation distance of 315 m.

calculation space. This is not physically correct for longer distances. However, it gave reasonable fits to data in Price *et al.* for distances up to 24 m (they measured out to 96 m but only show model comparisons out to 24 m). In method 2, at any distance d from the source, attenuation in decibels due to trunk scattering only is deducted from the overall level as a function of frequency using the following relation:

$$Attn(f) = 20\,\mathrm{Im}(k_b(f))d/\ln(10). \tag{7}$$

In these figures, a positive number indicates attenuation. Method 1 attenuations are calculated by subtracting the forest GFPE results from a run including only refraction and ground from results from a run including refraction, ground, and trunks. Method 2 attenuations are calculated using Eq. (7). Interestingly, the method by which the scattering attenuation is incorporated has almost no impact at 174 m, and a more significant impact at 315 m. Figures 9 and 10 show the scattering attenuation for methods 1 and 2 for upwind, downwind, and homogeneous profiles for each method at two distances 174 and 315 m, respectively. The differences in scattering attenuation are small in all cases. The method 1 attenuation generally tracks along the Linton and Martin method 2 scattering attenuation curves. Additional oscillations in the method 1 scattering attenuation are due to interference between modes.

VII. CONCLUSIONS

A modified Green's function parabolic equation model for forest propagation has been presented and compared to experimental data. The forest GFPE shows reasonable agreement with short-range measurements. The trunk scattering contribution has been examined using two different methods: as an integrated modification to the wave number and as an added attenuation at a distance d. The integrated modification to the wave number yields acceptable results and allows for all possible paths, including those that travel above the forest for some amount of time. It is important to note that this realization only takes the wave front into account. While this yields acceptable results, i.e., it satisfactorily predicts the received acoustic level, it does not take the reverberant field into account.

ACKNOWLEDGMENTS

The authors would like to thank Mr. Arnold Tunick at the Army Research Laboratory for providing the meteorological profiles used in this paper. The authors also thank Mr. Patrick Guertin of the U.S. Army Engineer Research and Development Center, Construction Engineering Research Laboratory (ERDC-CERL) for sharing his expertise in forestry. The authors gratefully acknowledge the assistance of Dr. Larry Pater and Mr. Jeff Mifflin, both from ERDC-CERL, for assisting in the measurements presented here. The authors would like to thank Mr. Ryan Lee, formally from ERDC-CERL, for reducing the measured data. Finally, the authors gratefully thank the editor and two anonymous reviewers for their extremely detailed and helpful comments. The U.S. Army Corps of Engineers, Engineer Research and Development Center, Construction Engineering Research Laboratory funded this work through Project No. 622720D048.

[1]C. F. Eyring, "Jungle acoustics," J. Acoust. Soc. Am. **18**, 257–270 (1946).
[2]R. Bullen and F. Fricke, "Sound propagation through vegetation," J. Sound Vib. **80**, 11–23 (1982).
[3]M. J. M. Martens and W. H. T. Huisman, "Ecosystem acoustics research," in *Proceedings of the Workshop on Sound Propagation in Forested Areas and Shelterbelts*, edited by M. J. M. Martens, pp. 13–24 (Fac. Sciences, Njimegen) (1986).
[4]M. A. Price, K. Attenborough, and N. W. Heap, "Sound attenuation through trees: Measurements and models," J. Acoust. Soc. Am. **84**, 1836–1844 (1988).
[5]W. H. T. Huisman and K. Attenborough, "Reverberation and attenuation in a pine forest," J. Acoust. Soc. Am. **90**, 2664–2677 (1991).
[6]J. Defrance, N. Barriere, and E. Premat, "A diffusion model for sound

propagation through forests," in *Forum Acusticum, Sevilla, Spain* (2002).
[7]C. M. Linton and P. A. Martin, "Multiple scattering by random configurations of circular cylinders: Second-order corrections for the effective wave number," J. Acoust. Soc. Am. **117**, 3413–3423 (2005).
[8]K. E. Gilbert and X. Di, "A fast Green's function method for one-way sound propagation in the atmosphere," J. Acoust. Soc. Am. **94**, 2343–2352 (1993).
[9]E. M. Salomons, "Improved Green's function parabolic equation method for atmospheric sound propagation," J. Acoust. Soc. Am. **104**, 100–111 (1998).
[10]A. Tunick, "Calculating the micrometeorological influences on the speed of sound through the atmosphere in forests," J. Acoust. Soc. Am. **114**, 1796–1806 (2003).
[11]K. Attenborough, "Ground parameter information for propagation modeling," J. Acoust. Soc. Am. **92**, 418–427 (1992).
[12]M. Lax, "Multiple scattering of waves," Rev. Mod. Phys. **23**, 287–310 (1951).
[13]L. E. Kinsler, A. R. Frey, A. B. Coppens, and J. V. Sanders, *Fundamentals of Acoustics* (John Wiley & Sons, New York, 1982).
[14]M. E. Swearingen and M. J. White, "Effects of forest on blast noise," Technical Report TR-05-29, ERDC/CERL, Champaign, IL (2005).
[15]F. G. Friedlander, "The diffraction of sound pulses," Proc. R. Soc. London, Ser. A **186**, 322–367 (1946).
[16]J. L. Cooper, "Parameter selection in the Green's function parabolic equation for outdoor sound propagation over varied terrain," Ph.D. thesis, The Pennsylvania State University (2003).
[17]V. Twersky, "On scattering of waves by random distributions. I. Free-space scatterer formalism," J. Math. Phys. **3**, 700–715 (1962).

Experimental characterization of onset of acoustic instability in a nonpremixed half-dump combustor

Satyanarayanan R. Chakravarthy[a] and Obla J. Shreenivasan
Department of Aerospace Engineering, Indian Institute of Technology–Madras, Chennai–600 036, India

Benjamin Boehm, Andreas Dreizler, and Johannes Janicka
Faculty for Energy and Powerplant Technology, Technical University of Darmstadt, Petersenstrasse 30, Darmstadt 64287, Germany

(Received 25 September 2006; revised 25 April 2007; accepted 26 April 2007)

This paper reports work on a nonpremixed half-dump combustor, in which methane is injected at the backward-facing step, and mixes and burns with the air flowing past the step in the unsteady recirculation zone. The flow and geometric parameters are widely varied, to gradually change from conditions of low-amplitude noise to excitation of high-amplitude discrete tones. The purpose of the work is to focus on the transition from the former condition to the latter, and to mark the onset of instability. Dimensionless groups such as the Helmholtz and Strouhal numbers are formed based on the observed dominant frequencies, whose variation with the air flow Reynolds number is used to identify the oscillations as those due to the natural acoustic modes or the vortex shedding process. High-speed chemiluminescence imaging reveals shedding of vortical structures in the flame zone. With variation in the conditions, flow-acoustic lock-on and transition from one vortex shedding mode to another is marked by nonlinearity in the corresponding amplitude variations. Such conditions are identified as the onset of instability in terms of the ratio of the flow time scale to the acoustic time scale and mapped against the operating fuel-air equivalence ratio of the combustor.
 [DOI: 10.1121/1.2741374]

PACS number(s): 43.28.Kt, 43.28.Py, 43.28.Ra, 43.20.Ks [RR] Pages: 120–127

I. INTRODUCTION

Unsteady and unstable combustion has been a subject of investigation for a long time, owing to problems of noise generation and excitation of acoustic instabilities that leads to structural damage caused by enhanced vibrations or heat transfer in combustors of gas turbines, furnaces, etc. Candel[1] has recently reviewed the literature pertaining to different aspects of combustion dynamics and control. On the one hand, elementary processes of flame interactions with flow disturbances in the combustor are involved,[2] and on the other, the role of heat release fluctuations in the combustion zone on the acoustic oscillations[3] is equally significant.

One of the important driving mechanisms of combustion instability has been that due to vortex shedding, as highlighted by Schadow and Gutmark,[4] who showed that the development of coherent flow structures and their breakdown into fine-scale turbulence can lead to periodic heat release. Many workers have investigated combustor geometries that include a predominant role for vortex shedding, such as dump combustors involving axisymmetric backward-facing step[5–9] and bluff-body flame-holders.[10] They have deduced the phase relationship between the vortex roll-up sequence and the heat release fluctuations under conditions of excitation of intense oscillations.

Bloxsidge *et al.*[11] have utilized the observations of heat release rate fluctuations in terms of the flame chemiluminescence reported in Ref. 10 to predict regimes of instability of combustor operation. Dowling[12] extended this to include saturation of the heat release rate fluctuations depending upon the instantaneous direction of the flow at the flameholder, which leads to prediction of hysteresis and nonlinear behavior including limit-cycle oscillations. Similarly, Dowling[13] developed a kinematic model of the flame oscillations with the flame anchor point fluctuating at the flameholder depending upon whether the flame speed instantaneously exceeded the flow velocity.

Lieuwen *et al.*[14] point out that the convective and chemical heat release time scales of the fuel-air mixture within the combustor may match the acoustic time scales for excitation of strong acoustic oscillations, and this could be amplified by equivalence ratio fluctuations[15] due to the response of the fuel feed line to the acoustic oscillations present in the combustion chamber. The vortex shedding at a location such as the dump plane in a dump combustor, for instance, would carry the fuel-air mixture in packets and modulate the heat release fluctuations. Hubbard and Dowling[16] have modeled the flame fluctuation in response to equivalence ratio fluctuations that are coupled to the acoustic oscillations in the combustor. Lieuwen[17] has reviewed in detail the flame-acoustic interaction in the context of modeling of combustion instability in premixed combustors. Lieuwen[18] has also reported experimental characterization of limit-cycle oscillations in a premixed dump combustor, including examples of transition from stable to unstable combustor operation in the form of supercritical and subcritical bifurcations.

[a]Author to whom correspondence should be addressed; electronic mail: src@ae.iitm.ac.in

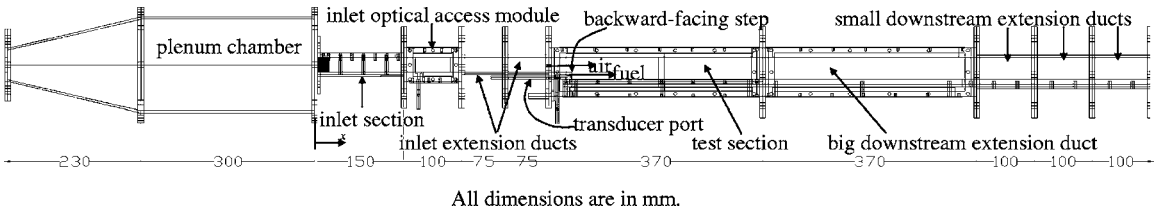


FIG. 1. Schematic of the experimental setup.

All the above-mentioned works are on premixed combustion, performed in the context of stationary gas turbines used for power generation, which have stringent emission control norms, and therefore employ lean premixed prevaporized combustors, but they are also prone to intense instability problems. Very little attention is devoted to instabilities in nonpremixed combustion systems such as those currently adopted in gas turbines for propulsion applications. Although the general view is that nonpremixed combustion is quite stable, it is not unconditionally so, however.

Further, almost all the above-mentioned works focus on investigations under conditions when the combustion oscillations are unstable. With a limited exception of Lieuwen[18] recently, there is no systematic work on a wide variation of geometric parameters of the combustor and flow conditions that span from a regime of low-intensity noise generation without appreciable acoustic feedback from the combustion chamber on the combustion process, to a regime of excitation of high-intensity discrete tones symptomatic of combustion instability. The Rayleigh criterion delineates regimes of unstable combustion from those of stable combustion, but it is only a necessary condition which could be met by a variety and combination of physical mechanisms. These mechanisms would gradually vary in predominance and interplay with each other, leading to a transition from low-amplitude noise to high-amplitude instability conditions. A systematic variation would prompt investigation on the mechanisms that dictate the onset of instability. The data can also serve to identify precursors to instability that can be utilized in actively deploying certain passive control measures in practical combustion systems.

The present work involves the operation of a rectangular half-dump combustor, wherein the fuel is injected at the corner of the backward-facing step to co-flow with the main flow of air past the step. The fuel-air mixing occurs in the recirculation zone downstream of the dump plane, followed by unsteady combustion in the oscillating shear layer of that zone. The length of the combustor is closely varied, while the length of the inlet duct up to the dump plane is held fixed, resulting in different locations of fuel injection and combustion heat release relative to the natural acoustic mode of the duct, which ensures different driving conditions. The air and fuel flow rates are also closely varied to obtain a range of overall equivalence ratio on the lean side, where blow-out limits are observed with interesting dependence on the duct length. Most of the measurements are on the unsteady pressure oscillations along the length of the combustor on either side of the backward-facing step. Limited experiments are performed on time-resolved imaging of the unsteady chemiluminescence of the CH^* radical in the flame zone to qualitatively examine the role of vortex shedding in the heat release fluctuations. The unsteady pressure data obtained are utilized in forming dimensionless quantities such as the Helmholtz and Strouhal numbers to distinguish fluctuations that are acoustic in nature from those due to the vortex shedding process. The data lend themselves to investigating the transition from conditions of low-intensity noise generation to those when high-amplitude discrete tones are excited, by correlation of jumps in the trends shown by the dimensionless quantities with "bifurcations" observed in the acoustic amplitude as exemplified in Ref. 18.

II. EXPERIMENTAL DETAILS

A schematic of the laboratory-scale combustor developed for this investigation is shown in Fig. 1. It is a rectangular geometry of constant width 60 mm downstream of the plenum chamber, to provide access for optical diagnostic investigation later. The plenum contracts abruptly into an inlet section with a contraction ratio of 17.45:1 by area, so as to provide a nominally acoustically open upstream boundary condition. The inlet section consists of flow-straightening tubes and mesh screens to reduce flow nonuniformities, particularly due to the presence of the sudden contraction. The inlet section has a height of 30 mm, and is followed by three sections of the same cross section until the backward-facing step is reached, namely, an inlet optical-access module and two upstream extension ducts. The test section contains the backward-facing step, whose height has been maintained at 30 mm in the present study. A number of optional downstream extension ducts of different lengths and the same cross-sectional dimensions of the test section (60 mm $\times$ 60 mm) are used in combination to vary the overall length of the combustor. Provision is made for injection of fuel through a single port of 4 mm $\times$ 4 mm dimension, 1 mm below the top edge of the step along the central plane of the test section. Methane is used as fuel.

The maximum range of variation of the air flow rate is 350–3550 liters per minute (l/min), in steps of 50 mm; the upper limit depends on the blow-off conditions at a given fuel flow rate. This gives a velocity range of approximately 3–30 m/s and a corresponding Reynolds number range of 6000–60 000, based on the step height. The fuel flow rate is varied at the levels of 6, 9, 13, 17, 21, 25, and 29 l/min for each air flow rate, which corresponds to fuel mass flow rate $\dot{m}_f$ in the range of 65–316 mg/s. This spans an equivalence ratio range of 0.02–0.79. The flow rates are measured by conventional rotameters with a measurement error within 2%.

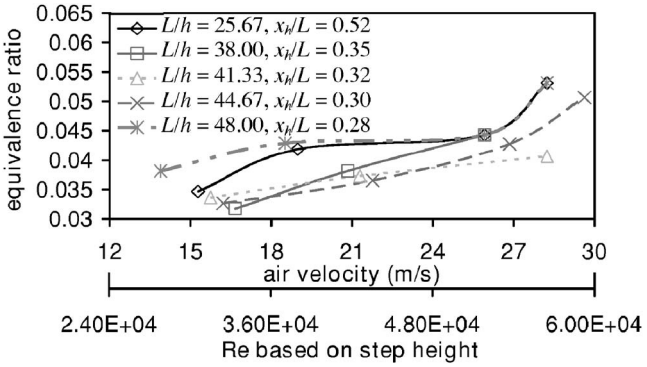


FIG. 2. Blow-off limits.

Five wall-mounted acoustic transducer measurements are made simultaneously in each experimental run. The transducers used are of the piezoelectric type, model No. 106B of PCB Piezotronics, with a sensitivity of 500 mV/psi and a resolution of 0.45 Pa. The sampling for each of the transducers is done at the rate of 10 kHz to obtain 32 768 (2^{15}) samples per channel. Fast Fourier transform is performed on the time-series data, with a bin size of 0.3 Hz. The dominant frequencies and their corresponding amplitudes are identified from the Fourier-transformed spectra. Multiple dominant frequencies are identified as local maxima in the spectra. The axial location of the transducer ports are marked as x_P in the figures presenting the results, where x is the axial distance measured from the entry to the inlet section (Fig. 1). One of the transducers is always mounted on the face of the step, at 400 mm from the entry to the inlet section.

Time-resolved imaging of the unsteady chemiluminescence of the CH^* radicals is performed with a high-speed CMOS camera (model HiSIS 2002 manufactured by KSV Instruments, Finland), at a framing rate of 2250 frames/s with a spatial resolution of 256×256 pixels for a duration of 3.64 s. An optical bandpass filter with peak transmittivity centered around 431 nm wavelength of light is fitted to the camera to admit the chemiluminescence of CH^* radicals in the flame zone, whose intensity is indicative of the region of chemical heat release that acts as the acoustic source in the present problem.

III. RESULTS AND DISCUSSION

A. Blow-off limits

The blow-off limits define the envelope of test conditions for investigating the acoustic characteristics at every length of the combustor. Figure 2 shows the blow-off limits for the different lengths of the combustor L, indicated in terms of the step height h. The axial location of the step x_h as a fraction of the combustor length is also indicated for each length in Fig. 2. It can be seen that both the cases of the shortest and the longest duct exhibit relatively higher overall equivalence ratios at blow-off than those of the intermediate lengths. Note that since the length of the inlet portion of the combustor upstream of the step is maintained constant in all the cases, the x_h/L decreases as L/h increases. For the smallest duct length, x_h/L is 0.52. This is a region where the acoustic velocity reaches a maximum for the dominant natural acoustic mode of the duct. This causes unstable combustion, leading to blow-off. On the other hand, the largest duct length exhibits excitation of acoustic oscillations of high amplitude, as seen later, which is the cause for blow-off at a relatively high equivalence ratio. For intermediate lengths, particularly in the lower range of the air velocity shown in Fig. 2, the blow-off limit is not sensitive to the duct length, which indicates a process such as blow out that is not affected by the acoustic oscillations in the duct.

B. Dominant frequencies and amplitudes

It is well known that broadband combustion noise is typically of lower amplitude than those of discrete tones observed under instability conditions in a given combustor. The transition between the former and latter conditions typically occurs as the air flow and fuel flow rates are increased, and in the present case, when the duct length is increased. Under the former conditions, multiple peaks of comparable (low) amplitudes are observed. For short ducts, the frequency of the fundamental natural acoustic mode is high when compared to the frequency of the shear layer oscillations in the flow downstream of the backward-facing step, particularly at low air flow rates, which results in the multiple peaks. As the conditions are varied toward excitation of instability, these multiple peaks move closer to each other and gain in amplitude, marking the onset of instability at some stage, as will be seen later.

Regardless of the multiple peaks, even when they are of comparable amplitude, the most dominant frequency in the observed amplitude spectra is selected under each flow condition for the present analysis. The trend in the dominant frequency with air flow rate is consolidated in the form of the dimensionless quantities, Helmholtz and Strouhal numbers, for different lengths of the duct. The Helmholtz number is defined as $\mathrm{He}=fL/c$, where f is the observed dominant frequency, and c is the speed of sound at the combustor inlet, which are the same as the ambient conditions, taken as a reference. The consideration of the reference speed of sound under ambient conditions as opposed to combustion conditions is discussed later. A nearly constant value of the Helmholtz number would signify that the observed dominant frequency corresponds to the natural acoustic mode of the duct, although not necessarily with any ideal acoustic boundary conditions. The Strouhal number is defined as $\mathrm{St}=fh/U$, where U is the mean air flow velocity at the inlet. A constant value of the Strouhal number would indicate that the observed dominant frequency corresponds to that due to the shear layer instability such as a vortex shedding process in the flow field. Note that when the Helmholtz number is formed based on an observed dominant frequency that corresponds to vortex shedding, it would vary linearly with the air flow Reynolds number, whereas when the Strouhal number is formed based on a frequency corresponding to a natural acoustic mode of the duct, it would vary hyperbolically with the air-flow Reynolds number. These variations enable distinction of the different dominant oscillations as due to either fluid mechanical or acoustic processes.

Figure 3 shows the variation of the Helmholtz and

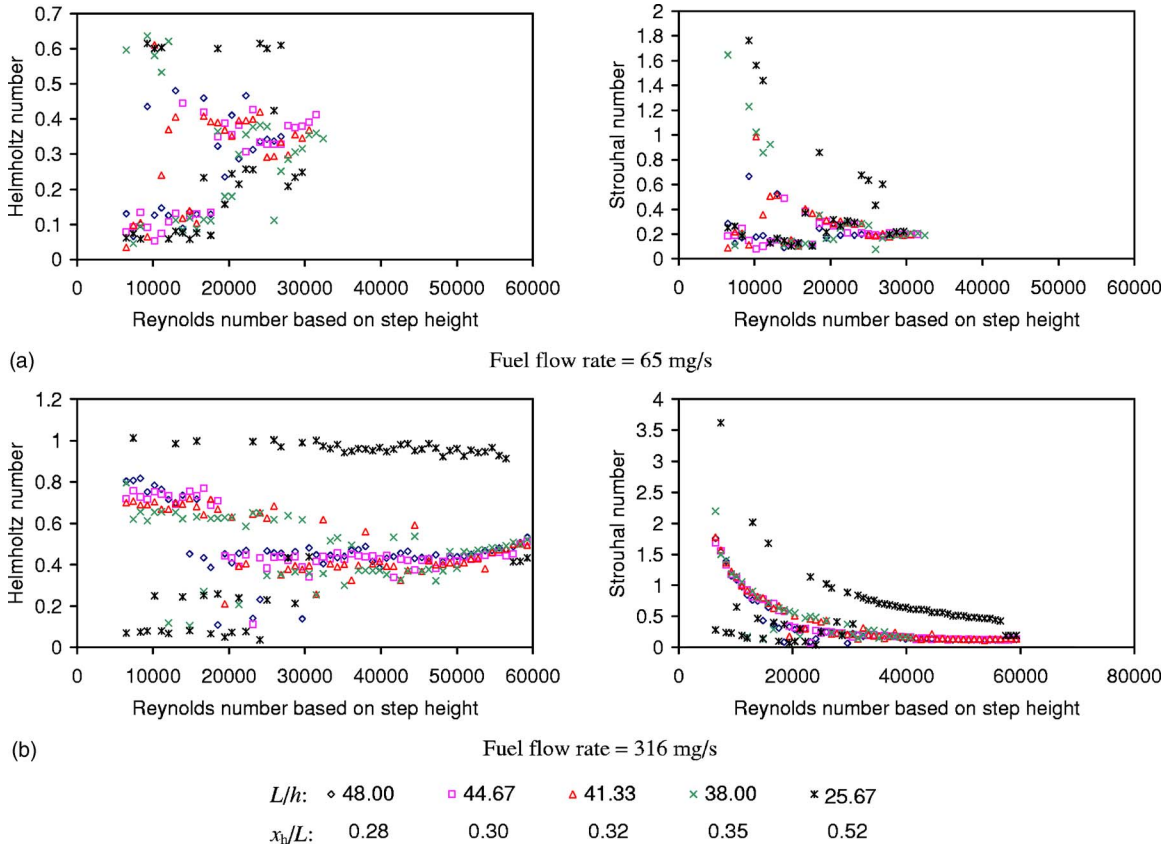


FIG. 3. (Color online) Helmholtz and Strouhal numbers based on the dominant frequencies observed at the backward-facing step.

Strouhal numbers based on the observed dominant frequencies with the Reynolds number (Re) based on the step height for all duct lengths tested, for two sample cases of extreme fuel flow rates. All the data in Fig. 3 correspond to those recorded at the base of the backward-facing step. This needs to be seen together with the amplitude at the dominant frequency observed at different locations along the duct, which is shown as a function of the air flow velocity at the above-mentioned two fuel flow rates, for two extreme duct lengths, in Fig. 4. It can be seen from Fig. 3(a) that, for the low fuel flow rate, most of the data fall in two clusters, one predominantly at low Helmholtz number for low Reynolds numbers, and another at a higher Helmholtz number for intermediate Reynolds numbers beyond which the flame blows off. However, the two are roughly in the same Strouhal number range around 0.2. The one at the intermediate Reynolds numbers asymptotes to this value hyperbolically. In Fig. 3(b) for high fuel flow rate, there are again two clusters of Helmholtz number data, with the one around 0.75 mildly decreasing with increase in the Reynolds number for the intermediate duct lengths. Most of the observed dominant frequency, however, transitions to the second cluster at a lower Helmholtz number, where it is nearly of constant value up to a Reynolds number of around 45 000, but linearly increases with further increase in Reynolds number. The latter corresponds to a constant Strouhal number of 0.2 again, and also registers the highest amplitudes [Fig. 4(b)(ii)]. These conditions exhibit the discrete high-amplitude tones as in combustion instability behavior as opposed to the broadband multiple peaks of comparable, low amplitudes observed at low Reynolds numbers. This behavior of a nearly constant Helmholtz number transitioning to a linearly increasing trend (or a hyperbolic trend in Strouhal number asymptoting to a constant value), and a corresponding nonlinear rise in the amplitudes with increase in the flow Reynolds number, is referred to here as the flow-acoustic lock-on. The shear layer oscillations downstream of the backward-facing step modulate the heat-release fluctuations that are the source of acoustic oscillations, and the frequency of the former oscillations matches one of the natural acoustic modes of the duct. The rise and the subsequent fall of the amplitudes are symptomatic of resonance between the two processes.

In both the cases of low and high fuel flow rates (Fig. 3), there is a third data set at quite high and nearly constant Helmholtz number for the shortest duct, and its Strouhal number varies hyperbolically but without asymptoting to a constant value. The latter feature indicates an absence of flow-acoustic lock-on for this duct length within the test conditions. The flame is located in the downstream part of the duct ($x_h/L=0.52$), so the second mode is excited, as the Rayleigh criterion is satisfied then.[19] The amplitudes increase nonlinearly for this mode at high fuel flow rate [Fig. 4(a)(ii)], but it is limited by the absence of a lock-on of this mode with the shear layer oscillations in the test range of air flow rates, unlike in the case of the longest duct.

C. Flow-acoustic lock-on

Besides the Strouhal number approaching a constant value under the flow-acoustic lock-on regime identified ear-

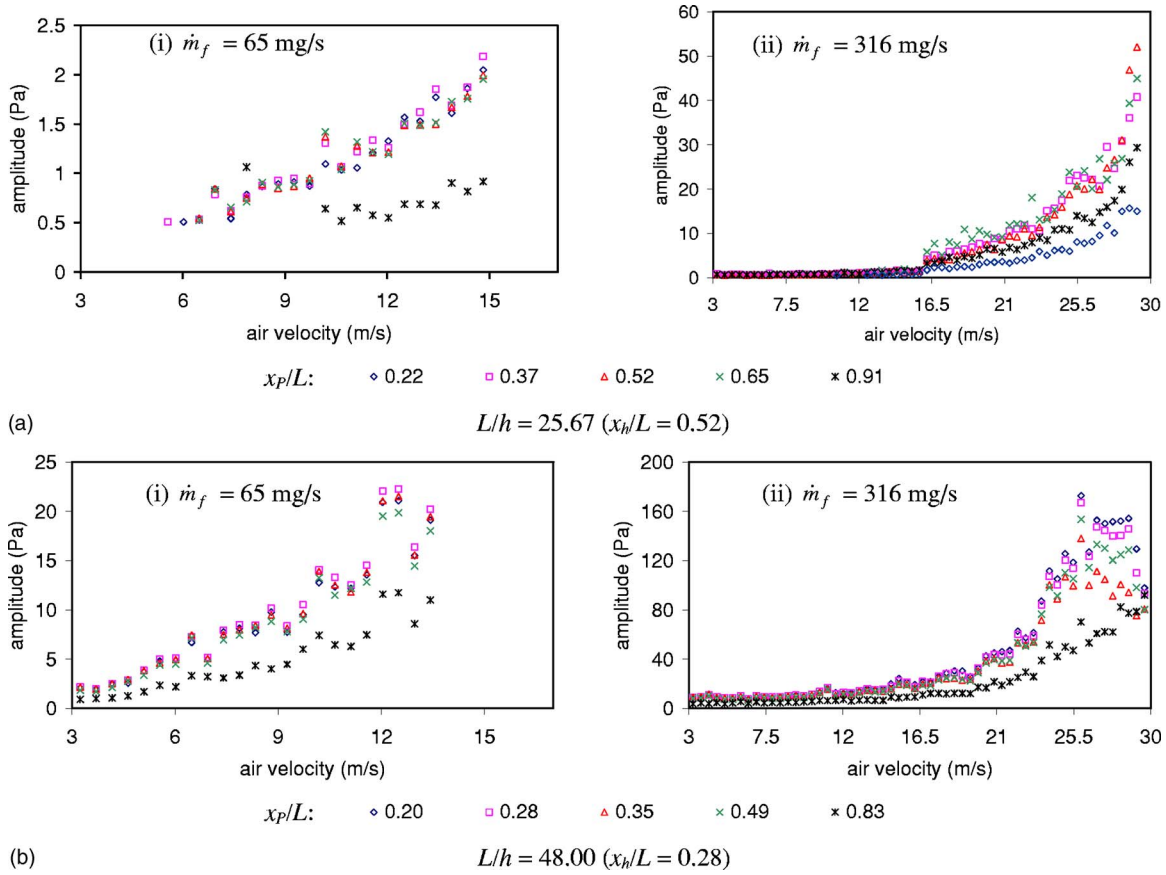


FIG. 4. (Color online) Amplitude of pressure fluctuations for different fuel flow rates and duct lengths.

lier, a direct indication, although qualitative, that fluid mechanical oscillations are involved as the driving mechanism of the observed acoustic oscillations is provided by the high-speed imaging of the CH^* chemiluminescence in the combustion zone. A time-resolved sequence of images is shown in Fig. 5 under the instability conditions. The outline of the flame shown in the region downstream of the backward-step indicates the presence of coherent structures corresponding to large-scale vortex roll-up in the shear layer originating from the step. Since the images are that of the flame, it follows that the heat release rate fluctuations generating the sound are actually modulated by the vortex shedding process.

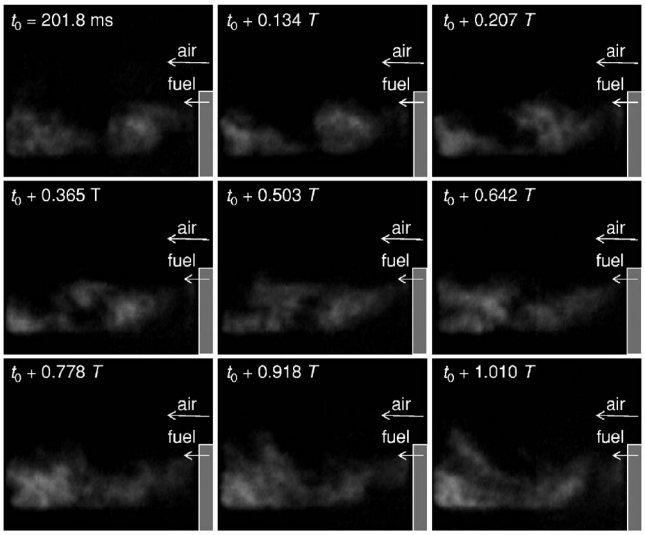


FIG. 5. Time-resolved sequence of CH^* chemiluminescence images in the combustion zone showing vortical patterns; T=9.637 ms, L/h=48(x_h/L =0.28), Re=54 000, and $\dot{m}_f$=316 mg/s.

Figure 6(a) shows the dominant frequencies measured simultaneously at the different transducer locations as a function of the air flow Reynolds number for the specific case of the longest duct and the highest fuel flow rate in the interest of clarity. Considerable scatter is observed at the low and intermediate Reynolds number, but the transducers at all the locations record the same dominant frequencies in the high Reynolds number range corresponding to the flow-acoustic lock-on regime. The linearly increasing trend in this range suggests that it is the vortex shedding frequency. While it is quite likely that the transducers in the vicinity of the backward-facing step record the fluid dynamic pressure fluctuations and hence the vortex shedding frequency, those near the ends of the duct, particularly the upstream end recording the same frequencies indicates that the natural acoustic mode of the duct varies as the vortex shedding frequency with increase in the air flow rate in this regime, although at lower amplitude [Fig. 4(b)(ii)].

To examine this further, the subdominant peak (local maxima) frequency in the observed amplitude spectra along the linearly increasing trend with air flow rate is also catalogued in Fig. 6(a) in the Reynolds number range lower than that of the lock-on regime, from the three transducers at and just downstream of the backward-facing step. This subdomi-

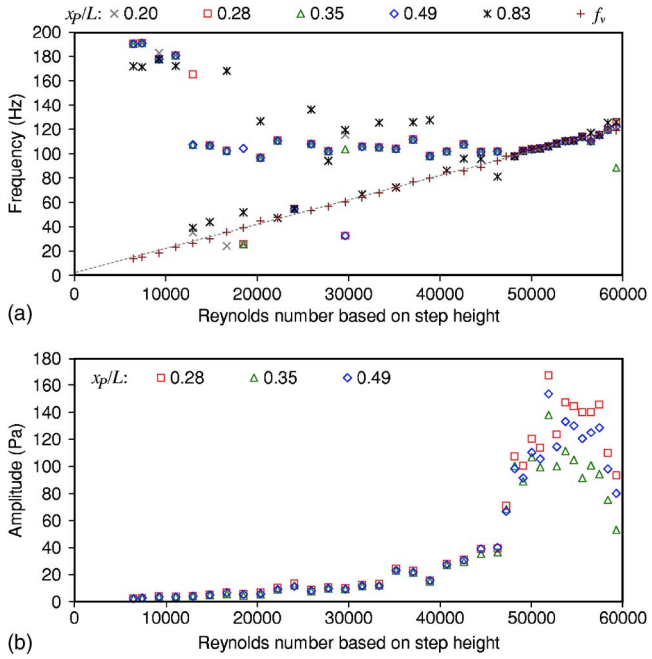


FIG. 6. (Color online) (a) Helmholtz number based on the dominant frequencies observed at different locations in the duct for $L/h=48$ ($x_h/L=0.28$) and $\dot{m}_f=316$ mg/s; f_v is the observed vortex shedding frequency observed at $x_p/L=0.28$, 0.35, and 0.49. (b) Amplitude at the vortex shedding frequency.

nant peak is observed in all three spectra at the same frequency at a given air flow Reynolds number, represented by a single plot symbol in Fig. 6(a). However, in general, the amplitudes are different at this frequency, which are plotted in Fig. 6(b). Comparing these with those in Fig. 4(b)(ii), it can be seen that the vortex shedding process amplifies the acoustic oscillations in the duct at the latter's natural frequency in the intermediate Reynolds number range prior to the lock-on regime, but the natural acoustic mode starts to follow the trend of the vortex shedding frequency with the onset of lock-on, accompanied by a nonlinearly steep rise in the amplitudes.

Since it is possible that the linearly increasing trend of the vortex shedding frequency with air flow rate is observed predominantly in the lock-on regime due to the fluid mechanical pressure fluctuations in the vicinity of the backward-facing step instead of the duct-acoustic fluctuations, the data from pairs of adjacent transducers are cross correlated to obtain the phase difference between the oscillations. Figure 7 shows the phase difference obtained between the transducers at $x_p/L=0.28$ (base of the step) and 0.35, and that between $x_p/L=0.49$ and 0.35, both at the observed dominant frequency as well as the subdominant vortex shedding frequency identified in Fig. 6(a) over the entire test Reynolds number range. It can be seen that all the phase differences are within a small range of $\sim 10°-20°$. If the oscillations were fluid dynamic in nature, their convection over the different distances between the two pairs of transducers would yield markedly different phase differences, at least at the vortex shedding frequency. The observation to the contrary indicates that it is the duct acoustic oscillations that are amplified by the heat release oscillations modulated by the vortex shedding process.

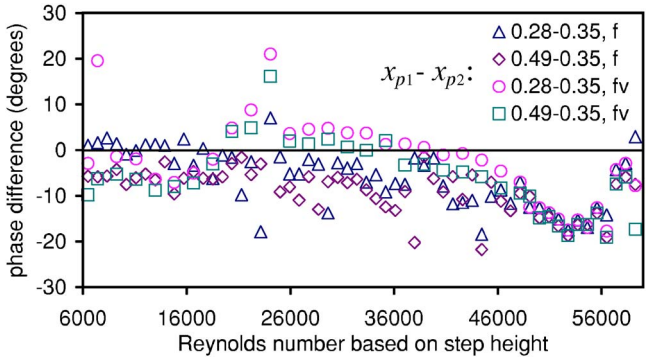


FIG. 7. (Color online) Phase differences in the transducer measurements between pairs of adjacent ports x_{p1} and x_{p2} measured at the dominant frequency f and the vortex shedding frequency f_v.

The above-presented discussion implies that the duct acoustic mode actually locks on to the vortex shedding frequency during the flow-acoustic lock-on in the present work rather than the other way. This is the opposite of what is observed in sound generated by cold-flow vortex shedding in a duct.[20] In the present case, the sound generation is mainly due to the heat release rate fluctuations and not significantly due to vorticity fluctuations. But the vortex shedding process modulates the heat release rate fluctuations by promoting the periodic mixing of fuel and air in the recirculation zone.

D. Mode transitions and onset of instability

The data obtained over a wide range of conditions in the present study are consolidated in the form of a map of onset of instability. This is first illustrated by considering a single representative data set shown in Fig. 8. The horizontal line in Fig. 8(a) denotes the natural acoustic mode of the duct prior to flow-acoustic lock-on, and the inclined lines denote differ-

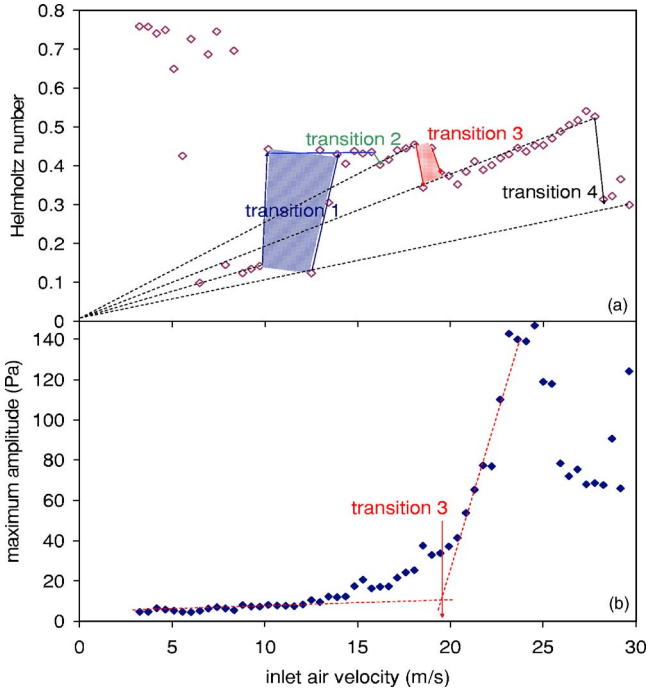


FIG. 8. (Color online) Mode transitions depicted for $L/h=44.67$ ($x_h/L=0.30$) and $\dot{m}_f=229$ mg/s.

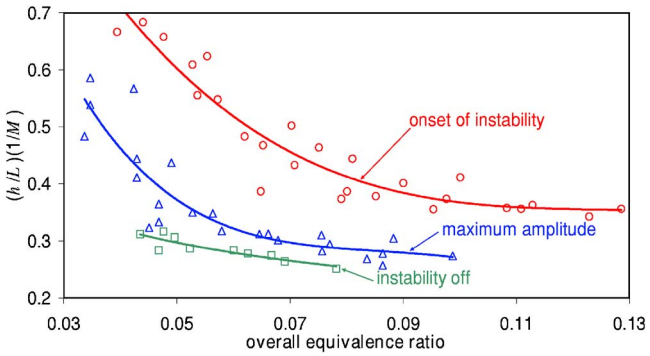


FIG. 9. (Color online) Conditions for onset of instability, excitation of maximum amplitude, and instability off.

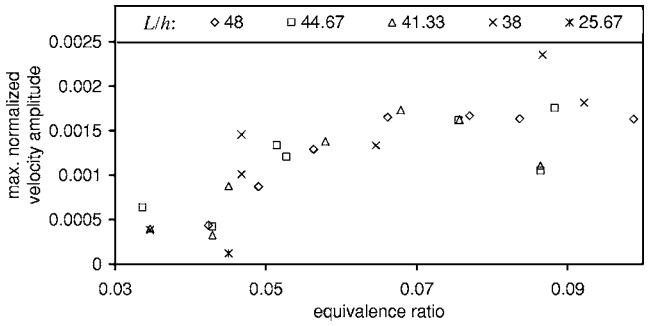


FIG. 10. Normalized maximum amplitude of fluctuations as a function of equivalence ratio.

ent possible modes of vortex shedding. The latter correspond to different constant Strouhal numbers obtained possibly due to such processes as vortex pairing/merging or splitting,[21,22] which needs to be confirmed by further optical diagnostic investigations. Four transitions are marked in Fig. 8(a), some over a range of air flow rates,[18] the first of which is similar to the flow-acoustic lock-on under cold-flow conditions.[20] The remaining ones are between successive vortex shedding modes within the flow-acoustic lock-on regime under combustion conditions identified earlier in Fig. 3(b). Of these, transition three is the most significant, as can be noticed by the sudden turn in the amplitude at that condition, in Fig. 8(b).

It is pertinent at this point to consider the effect of temperature during combustion on the speed of sound used in the definition of the Helmholtz number. Theoretically, the combustion temperature decreases with increase in the air flow rate for a fixed fuel flow rate. So, the speed of sound may accordingly decrease, and therefore the Helmholtz number would increase with increase in air flow rate. However, in reality, the combustion is distributed over a large part of the combustor under low air flow rate conditions when flow-acoustic lock-on does not prevail, but becomes increasingly compact with the increase in the acoustic amplitudes due to lock-on promoting fuel-air mixing. On the whole, this results in the maximum temperature attained being more or less constant over the air flow rate range. When the maximum temperature in the combustion zone is used for the speed of sound, the Helmholtz number merely decreases uniformly over the Reynolds number range in Fig. 8(a) relative to when the inlet temperature is used. So, the latter is retained, since it is of engineering significance as one of the control parameters of the flow conditions.

The conditions of transition three are collected from all the tests of the present study to form the ratio of the Strouhal number to the Helmholtz number, $\mathrm{St/He}=(h/L)(1/M)$, where M is the Mach number of the air flow at the combustor inlet. The above-noted ratio is that of the unsteady time scale of the flow to the acoustic time scale in the combustor. Strong flow-acoustic interaction and resonance is expected to occur when this ratio is of the order of unity. The above-presented quantity is plotted in Fig. 9 against the fuel-air equivalence ratio, so that all the flow and geometric parameters of the problem are represented by such a plot. The plot shows that all the conditions of onset of instability identified over a wide variation of parameters fall nearly on a single curve, with the $(h/L)(1/M)$ values being less than unity. The values are as expected because, in reality, the time scales of flow unsteadiness and acoustic modes are off by factors of ~ 0.2 and ~ 0.5, respectively. Besides, the actual flow velocity is lower and the actual length scale higher in the combustion zone than what is adopted here. Figure 9 also shows the flow and geometric conditions when the maximum amplitude is attained, after the onset of the instability, and also when the amplitude decreases to lower values, signaling the end of the flow-acoustic lock-on regime following transition three. The limited data on the latter are due to the flame blow-off under some conditions and reaching the higher limit of the test range of air flow rate under some other conditions.

Figure 10 shows the maximum amplitude obtained following the onset of the instability, in terms of the normalized acoustic velocity amplitude as $\hat{u}/U \sim (\frac{1}{2}M\hat{p})/(\frac{1}{2}\rho_0 U^2)$, where $\hat{u}$ and $\hat{p}$ are the acoustic velocity and pressure amplitudes, respectively, and ρ_0 is the mean flow density at the inlet. It can be seen that most of the data points fall into a single curve that levels off with increase in the equivalence ratio, suggesting that the thermal energy available from increased fuel flow rate does not indefinitely increase the amplitude levels, but may rather contribute to exciting higher modes.

The equivalence ratio range where significant variations are observed in Figs. 9 and 10 are quite low, as expected of nonpremixed combustors. These variations are likely to be observed at higher equivalence ratios when the fuel is injected ahead of the flame stabilization zone and premixed with the air before combustion.

IV. CONCLUSION

This paper reports unsteady pressure measurements in a nonpremixed half-dump combustor over a wide range of operating conditions and lengths of the combustor duct, ranging from generation of low-amplitude noise without appreciable feedback from the natural acoustic modes of the duct, to excitation of high-amplitude discrete tones due to intense flow-acoustic lock-on. Since these unsteady pressure signals could predominantly arise from the natural acoustic modes of the duct or due to local flow fluctuations in the vortex shedding process downstream of the dump plane, dimensionless groups such as the Helmholtz and Strouhal numbers are formed, whose variations help distinguish the duct acoustic

modes from the vortex shedding modes. As the flow velocity is increased, the observed dominant frequencies shift from those corresponding to the latter to those of the former, whereupon the flow-acoustic lock-on shifts between one vortex shedding mode to the other, as opposed to the vortex shedding locking on to the natural acoustic mode of the duct observed under cold-flow conditions. These are considered to be due to such processes as vortex pairing and merging that influence the heat release fluctuations, which demands a separate investigation. A map of the onset of instability, excitation of maximum amplitude, and the end of instability is presented for the conditions tested. The map is in a domain of the equivalence ratio of the fuel and air flows to the ratio of the Strouhal number to the Helmholtz number, i.e., the ratio of the flow time scale to the acoustic time scale. These two quantities contain all the parameters of the problem, and indicate an instability map for a particular location of injection of the fuel, i.e., mode of combustion, in a given combustor geometry.

ACKNOWLEDGMENT

This work was funded by the Deutsche Forschungsgemeinschaft, Germany, under Contract Nos. JA-544/30-1 and JA-544/31-1.

[1]S. Candel, "Combustion dynamics and control: Progress and challenges," Proc. Combust. Inst. **29**, 1–28 (2002).
[2]S. Ducruix, T. Schuller, D. Durox, and S. Candel, "Combustion dynamics and instabilities: Elementary coupling and driving mechanisms," J. Propul. Power **19**, 722–734 (2003).
[3]S. Kato, T. Fujimori, A. P. Dowling, and H. Kobayashi, "Effect of heat release distribution on combustion oscillation," Proc. Combust. Inst. **30**(2), 1799-1806 (2007).
[4]S. C. Schadow and E. J. Gutmark, "Combustion instability related to vortex shedding in dump combustors and their passive control," Prog. Energy Combust. Sci. **18**, 117–132 (1992).
[5]E. J. Gutmark, K. C. Schadow, S. Sivasegaram, and J. H. Whitelaw, "Interaction between fluid dynamic and acoustic instabilities in combusting flows within ducts," Combust. Sci. Technol. **79**, 161–166 (1991).
[6]S. Sivasegaram and J. H. Whitelaw, "Oscillations in axisymmteric dump combustors," Combust. Sci. Technol. **52**, 413–426 (1987).
[7]K. H. Yu, A. C. Trouve, and J. W. Daily, "Low frequency pressure oscillations in a model ramjet combustor," J. Fluid Mech. **232**, 47–72 (1991).
[8]T. J. Poinsot, A. C. Trouve, D. P. Veynante, S. M. Candel, and E. J. Esposito, "Vortex-driven acoustically coupled combustion instabilities," J. Fluid Mech. **177**, 265–292 (1987).
[9]J. D. Sterling and E. E. Zukoski, "Non-linear dynamics of laboratory combustor pressure oscillations," Combust. Sci. Technol. **77**, 225–238 (1991).
[10]P. J. Langhorne, "Reheat buzz: An acoustically coupled combustion instability. 1. Experiment," J. Fluid Mech. **193**, 417–443 (1988).
[11]G. J. Bloxsidge, A. P. Dowling, and P. J. Langhorne, "Reheat buzz: An acoustically coupled combustion instability. 2. Theory," J. Fluid Mech. **193**, 445–473 (1988).
[12]A. P. Dowling, "Nonlinear self-excited oscillations of a ducted flame," J. Fluid Mech. **346**, 271–290 (1997).
[13]A. P. Dowling, "A kinematic model of a ducted flame," J. Fluid Mech. **394**, 51–72 (1999).
[14]T. C. Lieuwen, H. Torres, C. Johnson, and B. T. Zinn, "A mechanism of combustion instability in lean premixed gas turbine combustors," J. Eng. Gas Turbines Power **123**, 182–190 (2001).
[15]T. C. Lieuwen and B. T. Zinn, "The role of equivalence ratio oscillations in driving combustion instabilities in low NO_x gas turbines," Proc. Combust. Inst. **27**, 1809–1816 (1998).
[16]S. Hubbard and A. P. Dowling, "Acoustic resonances of an industrial gas turbine combustion system," J. Eng. Gas Turbines Power **123**, 766–773 (2001).
[17]T. C. Lieuwen, "Modeling premixed combustion – Acoustic wave interactions: A review," J. Propul. Power **19**, 765–781 (2003).
[18]T. C. Lieuwen, "Experimental investigation of limit-cycle oscillations in an unstable gas turbine combustor," J. Propul. Power **18**, 61–67 (2002).
[19]M. Tyagi, S. R. Chakravarthy, and R. I. Sujith, "Unsteady combustion response of a ducted non-premixed flame and acoustic coupling" Combust. Theory Modell. **11**(2), 205–226 (2007).
[20]B. Karthik, S. R. Chakravarthy, and R. I. Sujith, "Jet forking driven by pipe tone," J. Acoust. Soc. Am. **113**, 2091–2094 (2003).
[21]C.-M. Ho and L.-S. Huang, "Subharmonics and vortex-merging in mixing layers," J. Fluid Mech. **119**, 443–473 (1982).
[22]C. D. Winant and F. K. Browand, "Vortex pairing: The mechanism of turbulent mixing-layer growth at moderate Reynolds number," J. Fluid Mech. **63** 237–255 (1974).

Vorticity dynamics and sound generation in two-dimensional fluid flow

Raymond J. Nagem[a] and Guido Sandri
Department of Aerospace and Mechanical Engineering, Boston University, Boston, Massachusetts 02215

David Uminsky
Department of Mathematics and Statistics, Boston University, Boston, Massachusetts 02215

(Received 12 December 2006; revised 10 April 2007; accepted 12 April 2007)

An approximate solution to the two-dimensional incompressible fluid equations is constructed by expanding the vorticity field in a series of derivatives of a Gaussian vortex. The expansion is used to analyze the motion of a corotating Gaussian vortex pair, and the spatial rotation frequency of the vortex pair is derived directly from the fluid vorticity equation. The resulting rotation frequency includes the effects of finite vortex core size and viscosity and reduces, in the appropriate limit, to the rotation frequency of the Kirchhoff point vortex theory. The expansion is then used in the low Mach number Lighthill equation to derive the far-field acoustic pressure generated by the Gaussian vortex pair. This pressure amplitude is compared with that of a previous fully numerical simulation in which the Reynolds number is large and the vortex core size is significant compared to the vortex separation. The present analytic result for the far-field acoustic pressure is shown to be substantially more accurate than previous theoretical predictions. The given example suggests that the vorticity expansion is a useful tool for the prediction of sound generated by a general distributed vorticity field. © *2007 Acoustical Society of America.* [DOI: 10.1121/1.2736513]

PACS number(s): 43.28.Ra, 43.20.Rz [RMW] Pages: 128–134

I. INTRODUCTION

The prediction of aerodynamic sound depends on establishing a connection between an unsteady fluid flow, which is most commonly modeled as incompressible, and the small compressible fluctuations associated with an acoustic field. The most successful approach to this fundamentally unsolved problem has identified the *vorticity* of the incompressible fluid flow as an important source for the generation of sound. This theory, which takes its starting point from the Lighthill acoustic analogy,[1] has provided many examples of sound generation by a given vorticity distribution in a fluid flow.[2]

In this paper, we begin in Sec. II with the general equation for the time evolution of the vorticity in a two-dimensional incompressible flow. We expand the vorticity field in a series of derivatives of a Gaussian vortex. We then derive a set of ordinary differential equations for the time-dependent coefficients of our series. A solution of these ordinary differential equations provides a solution for the time-dependent vorticity and velocity fields. In Sec. III, we apply our method to analyze the motion of two corotating Gaussian vortices, and show that the second-order term of our vorticity expansion rotates in space with a frequency which reduces, as the vortex core size and the viscosity approach zero, to the rotation frequency predicted by the Kirchhoff point vortex theory. In Sec. IV, we show how our solution of the vorticity equation can be used in the Lighthill equation for sound generation. We use the Lighthill equation to compute the sound produced by the Gaussian vortex pair, and compare our result to a previous high Reynolds number numerical computation which is based on a numerical solution of the compressible Navier-Stokes equations. It is shown that our analytic result for the far-field pressure is substantially more accurate than previous theoretical predictions.

Our example suggests that the second-order term of our vorticity expansion is sufficient to provide a good approximation to the sound generated by a two-dimensional distributed vorticity field. The higher-order terms in our expansion, which are not considered in detail here, may give additional insight into the general structure (for example, vortex core distortion and vortex core interaction) of two-dimensional vortex dynamics.

II. VORTICITY DYNAMICS FOR TWO-DIMENSIONAL INCOMPRESSIBLE FLOW

The continuity and momentum equations for two-dimensional incompressible fluid flow can be written in the form

$$\frac{\partial u_i}{\partial x_i} = 0, \tag{1}$$

$$\dot{\omega} + u_i \frac{\partial \omega}{\partial x_i} = \nu \frac{\partial^2 \omega}{\partial x_i \partial x_i}, \tag{2}$$

where $u_i(x_1, x_2, t)$ is the fluid velocity vector, ν is the kinematic viscosity, and

[a]Electronic mail: nagem@bu.edu

 0001-4966/2007/122(1)/128/7/$23.00

$$\omega(x_1,x_2,t)=\varepsilon_{jk}\frac{\partial u_k}{\partial x_j},\quad \varepsilon_{12}=-\varepsilon_{21}=1,\quad \varepsilon_{11}=\varepsilon_{22}=0 \tag{3}$$

is the single component of the fluid vorticity. Throughout this paper, the superposed dot indicates the time derivative, alphabetic subscripts take the value 1 or 2, and the summation convention for repeated subscripts is assumed.

A particular solution of Eqs. (1)–(3) is the Gaussian vortex[3] defined by the vorticity and velocity fields

$$\omega=\gamma\delta^{\lambda}(x_1,x_2,t)\equiv\gamma\frac{e^{-r^2/\lambda^2}}{\pi\lambda^2} \tag{4}$$

and

$$u_i=\gamma U_i^{\lambda}(x_1,x_2,t)\equiv\gamma\frac{1-e^{-r^2/\lambda^2}}{2\pi r}\hat{\theta}_i=\gamma\frac{1-e^{-r^2/\lambda^2}}{2\pi r}(-\varepsilon_{ij}\hat{r}_j) \tag{5}$$

where

$$\lambda=\sqrt{\lambda_0^2+4\nu t}=\lambda_0\sqrt{1+4\nu t/\lambda_0^2}, \tag{6}$$

$r=\sqrt{x_1^2+x_2^2}$, and $\hat{r}_i$, $\hat{\theta}_i$ are the ith components of the radial and tangential unit vectors corresponding to the r, θ polar coordinate system in the x_1, x_2 plane. The vorticity distribution $\gamma\delta^{\lambda}(r,t)$ is a two-dimensional radially symmetric Gaussian whose total integral over the x_1, x_2 plane is equal to the constant γ. The spatial broadness of the Gaussian at time $t=0$ is characterized by the constant λ_0, and the spatial broadness $\lambda(t)$ increases in proportion to $\sqrt{t}$. As $\lambda\to 0$, $\gamma\delta^{\lambda}(r,t)$ approaches the two-dimensional point vortex $\gamma\delta(\mathbf{r})$. The tangential velocity γU^{λ} is zero at $r=0$, and has a radially symmetric core region in which the tangential velocity increases to a maximum value of approximately $(0.102)\gamma/\lambda$ at a radius of approximately $(1.121)\lambda$. For $r\gg\lambda$, the tangential velocity γU^{λ} approaches the value $\gamma/2\pi r$ associated with the point vortex $\gamma\delta(\mathbf{r})$.

To obtain a more general solution of Eqs. (1)–(3), we construct the series

$$\omega=\Gamma(t)\delta^{\lambda}+D_j(t)\frac{\partial}{\partial x_j}\delta^{\lambda}+Q_{jk}(t)\frac{\partial^2}{\partial x_j\partial x_k}\delta^{\lambda}+\cdots, \tag{7}$$

$$u_i=\Gamma(t)U_i^{\lambda}+D_j(t)\frac{\partial}{\partial x_j}U_i^{\lambda}+Q_{jk}(t)\frac{\partial^2}{\partial x_j\partial x_k}U_i^{k}+\cdots, \tag{8}$$

which satisfy Eqs. (1) and (3) identically. The series coefficients $\Gamma(t)$, $D_j(t)$, and $Q_{jk}(t)$ are to be determined by vorticity equation (2). The coefficient Q_{jk} is assumed, without loss of generality, to be symmetric in j and k. A similar series for the vorticity and velocities fields has been used previously to derive long-time asymptotics of the two-dimensional Navier-Stokes and vorticity equations.[4]

The series (7) is a convergent expansion in the two-dimensional Gaussian derivatives, a complete and orthogonal set of spatial basis functions for functions which have spatial moments of all orders.[5] Using the orthogonality property of these basis functions, the first three time-dependent series coefficients in Eq. (7) are given in terms of ω by the moments

$$\Gamma=\int\int d^2\mathbf{r}\omega, \tag{9}$$

$$D_j=-\int\int d^2\mathbf{r}x_j\omega, \tag{10}$$

$$Q_{jk}=\int\int d^2\mathbf{r}\left(\frac{x_jx_k}{2}-\frac{\lambda^2\delta_{jk}}{4}\right)\omega, \tag{11}$$

where δ_{jk} is the Kronecker delta.

We refer to the successive terms on the right-hand side of Eq. (7) as the monopole, dipole, and quadrupole components of the vorticity field. This terminology is conventional when the two-dimensional Dirac delta function $\delta(\mathbf{r})$ is used in place of the Gaussian $\delta^{\lambda}(r,t)$. The corresponding terminology is used for the successive terms in the velocity field given by Eq. (8). By using $\delta^{\lambda}(r,t)$ instead of $\delta(\mathbf{r})$, we avoid the singularities associated with the distribution $\delta(\mathbf{r})$ and, as shown in Eq. (14), simultaneously account for the viscous diffusion that is explicitly built into the function $\delta^{\lambda}(r,t)$ via the parameter λ defined in Eq. (6).

Substitution of the series (7) and (8) into Eq. (2) gives

$$\begin{aligned}&\dot{\Gamma}\delta^{\lambda}+\Gamma\dot{\delta}^{\lambda}+\dot{D}_j\frac{\partial\delta^{\lambda}}{\delta x_j}+D_j\frac{\partial\dot{\delta}^{\lambda}}{\partial x_j}+\dot{Q}_{jk}\frac{\partial^2\delta^{\lambda}}{\partial x_j\partial x_k}+Q_{jk}\frac{\partial^2\dot{\delta}^{\lambda}}{\partial x_j\partial x_k}\\&\quad+\cdots+\frac{\partial}{\partial x_i}\left[\left(\Gamma U_i^{\lambda}+D_j\frac{\partial U_i^{\lambda}}{\partial x_j}+Q_{jk}\frac{\partial^2 V_i^{\lambda}}{\partial x_j\partial x_k}+\cdots\right)\right.\\&\quad\left.\times\left(\Gamma\delta^{\lambda}+D_m\frac{\partial\delta^{\lambda}}{\partial x_m}+Q_{mn}\frac{\partial^2\delta^{\lambda}}{\partial x_m\partial x_n}+\cdots\right)\right]\\&\quad=\nu\frac{\partial^2}{\partial x_i\partial x_i}\left(\Gamma\delta^{\lambda}+D_j\frac{\partial\delta^{\lambda}}{\partial x_j}+Q_{jk}\frac{\partial^2\delta^{\lambda}}{\partial x_j\partial x_k}+\cdots\right),\end{aligned} \tag{12}$$

where we have used

$$u_i\frac{\partial\omega}{\partial x_i}=\frac{\partial}{\partial x_i}(u_i\omega) \tag{13}$$

on the basis of Eq. (1). Since

$$\dot{\delta}^{\lambda}=\nu\frac{\partial^2}{\partial x_i\partial x_i}\delta^{\lambda}, \tag{14}$$

Eq. (12) reduces to the much simpler equation

$$\begin{aligned}&\dot{\Gamma}\delta^{\lambda}+\dot{D}_j\frac{\partial\delta^{\lambda}}{\partial x_j}+\dot{Q}_{jk}\frac{\partial^2\delta^{\lambda}}{\partial x_j\partial x_k}+\cdots+\frac{\partial}{\partial x_i}\left[\left(\Gamma U_i^{\lambda}+D_j\frac{\partial U_i^{\lambda}}{\partial x_j}\right.\right.\\&\quad\left.+Q_{jk}\frac{\partial^2 U_i^{\lambda}}{\partial x_j\partial x_k}+\cdots\right)\left(\Gamma\delta^{\lambda}+D_m\frac{\partial\delta^{\lambda}}{\partial x_m}+Q_{mn}\frac{\partial^2\delta^{\lambda}}{\partial x_m\partial x_n}\right.\\&\quad\left.\left.+\cdots\right)\right]=0.\end{aligned} \tag{15}$$

Differential equations for the temporal functions $\Gamma(t)$, $D_j(t)$, and $Q_{jk}(t)$ can now be obtained from Eq. (15) by using the orthogonality properties of the two-dimensional Gaussian derivatives, as discussed earlier in conjunction with Eqs. (9)–(11). Integration of Eq. (15) over the entire x_1,x_2 plane gives

$$\dot{\Gamma} = -\int\int d^2\mathbf{r}\frac{\partial}{\partial x_i}(u_i\omega), \tag{16}$$

which after an integration by parts gives

$$\dot{\Gamma} = 0. \tag{17}$$

The coefficient Γ, according to Eq. (9), is the total spatial integral of the vorticity field. Equation (17) thus states the known result that this total spatial integral of the vorticity is constant in time.

Taking the first spatial moment of Eq. (15) gives

$$\dot{D}_j = \int\int d^2\mathbf{r}x_j\frac{\partial}{\partial x_i}(u_i\omega) = -\delta_{ij}\int\int d^2\mathbf{r}u_i\omega, \tag{18}$$

where the last equality is obtained by an integration by parts. It is shown in Appendix A that the integral on the right-hand side of Eq. (18) vanishes, so that

$$\dot{D}_j = 0. \tag{19}$$

We thus obtain the known result that the dipole coefficient D_j in the series (7), which according to Eq. (10) is the negative of the first spatial moment of the vorticity field, is also constant in time.

The time derivative of the quadrupole coefficient Q_{jk} in Eq. (7) is obtained from Eq. (15) as

$$\dot{Q}_{pq} = -\int\int d^2\mathbf{r}\left(\frac{x_jx_k}{2} - \frac{\lambda^2\delta_{jk}}{4}\right)\left\{\frac{\partial}{\partial x_i}\left[\left(\Gamma U_i^\lambda + D_p\frac{\partial U_i^\lambda}{\partial x_p} + Q_{pq}\frac{\partial^2 U_i^\lambda}{\partial x_p\partial x_q} + \cdots\right)\times\left(\Gamma\delta^\lambda + D_m\frac{\partial\delta^\lambda}{\partial x_m} + Q_{mn}\frac{\partial^2\delta^\lambda}{\partial x_m\partial x_n} + \cdots\right)\right]\right\}. \tag{20}$$

Each term in the infinite series on the right-hand side of Eq. (20) can be integrated in closed form. We have evaluated the terms up the quadrupole order explicitly; the result of these evaluations is

$$\dot{Q}_{jk} = -\frac{Q_{ii}}{8\pi\lambda^4}(Q_{jm}\varepsilon_{mk} + Q_{km}\varepsilon_{mj}) + \frac{\Gamma}{8\pi\lambda^2}(Q_{jm}\varepsilon_{mk} + Q_{km}\varepsilon_{mj}) + \frac{1}{16\pi\lambda^2}(D_jD_m\varepsilon_{km} + D_kD_m\varepsilon_{jm}). \tag{21}$$

If we set $j=k$ in Eq. (21), we obtain

$$\dot{Q}_{kk} = 0, \tag{22}$$

so that

$$T = Q_{kk} = Q_{11} + Q_{22} = \mathrm{Tr}(Q) = \text{constant}. \tag{23}$$

Equations (21)–(23) are based on a truncation of the series (7) and (8), but it can be shown directly from Eq. (11) (see Appendix B) that the result $\mathrm{Tr}(Q)$=constant remains valid when the entire series (7) and (8) are retained.

The constants Γ and D_j, together with a solution of Eq. (21), give, via Eqs. (7) and (8), an approximate solution for the vorticity and velocity fields. The values of the constants Γ and D_j and the initial values of Q_{jk} are obtained in terms of the initial value of the vorticity distribution by means of Eqs. (9)–(11). At $t=0$, the parameter λ reduces to the constant λ_0, which is chosen on the basis of the length scale associated with the initial vorticity distribution.

III. MOTION OF A GAUSSIAN VORTEX PAIR

As an example, we consider the initial vorticity distribution

$$\omega(x,y,0) = \gamma\delta^a(x_1 - \ell, x_2) + \gamma\delta^a(x_1 + \ell, x_2), \tag{24}$$

where the function δ^a is defined by Eq. (4) with λ equal to the constant value a. Equation (24) represents two equal corotating Gaussian vortices on the x_1 axis, separated by the distance 2ℓ. The initial core size of the vortices is determined by the parameter a. If the core size a is zero, the Gaussians become point vortices. In the absence of viscosity, the Kirchhoff theory of two-dimensional point vortex motion[6] predicts that two such point vortices will rotate around the origin of the x_1,x_2 plane on a circle of radius 2ℓ with angular velocity $\gamma/4\pi\ell^2$. We will use the series (7) to predict the approximate time evolution of the Gaussian vortices, including the effects of viscosity and the finite size of the vortex core regions.

For the initial vorticity distribution (24), Eqs. (9)–(11) give, with $\lambda_0=\ell$, the monopole, dipole, and and initial quadrupole coefficients,

$$\Gamma = 2\gamma, \tag{25}$$

$$D_1 = D_2 = 0, \tag{26}$$

$$Q_{11}(0) = \gamma\left(\frac{a^2}{2} + \frac{\ell^2}{2}\right), \tag{27}$$

$$Q_{12}(0) = 0, \tag{28}$$

$$Q_{22}(0) = \gamma\left(\frac{a^2}{2} - \frac{\ell^2}{2}\right). \tag{29}$$

Here

$$T = Q_{11} + Q_{22} = \gamma a^2, \tag{30}$$

and the components of Eq. (21) become

$$\dot{Q}_{11} = \frac{Q_{12}}{4\pi\lambda^2}\left(\frac{T}{\lambda^2} - \Gamma\right), \tag{31}$$

$$\dot{Q}_{12} = -\frac{Q_{11} - Q_{22}}{8\pi\lambda^2}\left(\frac{T}{\lambda^2} - \Gamma\right), \tag{32}$$

$$\dot{Q}_{22} = -\frac{Q_{12}}{4\pi\lambda^2}\left(\frac{T}{\lambda^2} - \Gamma\right). \tag{33}$$

Letting

$$S = Q_{11} - Q_{22}, \tag{34}$$

Eqs. (31)–(33) imply that

$$\dot{S} = \frac{Q_{12}}{2\pi\lambda^2}\left(\frac{T}{\lambda^2} - \Gamma\right), \tag{35}$$

$$\dot{Q}_{12} = -\frac{S}{8\pi\lambda^2}\left(\frac{T}{\lambda^2} - \Gamma\right), \tag{36}$$

which in turn imply that

$$-4Q_{12}\dot{Q}_{12} = S\dot{S}, \tag{37}$$

or

$$S^2 + 4Q_{12}^2 = 4C_1^2, \tag{38}$$

where C_1 is a constant. Equations (36) and (38) now give

$$\frac{dQ_{12}}{\sqrt{4C_1^2 - 4Q_{12}^2}} = \frac{dt}{8\pi\lambda^2}\left(\frac{T}{\lambda^2} - \Gamma\right). \tag{39}$$

Equation (39) can be integrated to give Q_{12}, and Eq. (36) then gives S. Using Eq. (23), the results for the quadrupole coefficients are, assuming $C_1 \neq 0$,

$$Q_{11} = \frac{Q_{11}(0) + Q_{22}(0)}{2} + \frac{Q_{11}(0) - Q_{22}(0)}{2}\cos 2\Omega t + Q_{12}(0)\sin 2\Omega t, \tag{40}$$

$$Q_{12} = Q_{12}(0)\cos 2\Omega t + \frac{Q_{11}(0) - Q_{22}(0)}{2}\sin 2\Omega t, \tag{41}$$

$$Q_{22} = \frac{Q_{11}(0) + Q_{22}(0)}{2} - \frac{Q_{11}(0) - Q_{22}(0)}{2}\cos 2\Omega t - Q_{12}(0)\sin 2\Omega t, \tag{42}$$

where the time-dependent frequency Ω is given by

$$\Omega = \frac{1}{8\pi}\left[\frac{\Gamma}{4\nu t}\ln\left(1 + \frac{4\nu t}{\lambda_0^2}\right) - \frac{T}{\lambda_0^4}\frac{1}{1 + 4\nu t/\lambda_0^2}\right]. \tag{43}$$

For the initial conditions of our example, Eqs. (40)–(44) reduce to

$$Q_{11} = \frac{\gamma a^2}{2} + \frac{\gamma \ell^2}{2}\cos 2\Omega t, \tag{44}$$

$$Q_{12} = \frac{\gamma \ell^2}{2}\sin 2\Omega t, \tag{45}$$

$$Q_{22} = \frac{\gamma a^2}{2} - \frac{\gamma \ell^2}{2}\cos 2\Omega t, \tag{46}$$

$$\Omega = \frac{1}{8\pi}\left[\frac{\gamma}{2\nu t}\ln\left(1 + \frac{4\nu t}{\ell^2}\right) - \frac{\gamma a^2}{\ell^4}\frac{1}{1 + 4\nu t/\ell^2}\right]. \tag{47}$$

Equations (40)–(42) can be written as

$$\begin{bmatrix} Q_{11} & Q_{12} \\ Q_{12} & Q_{22} \end{bmatrix} = \begin{bmatrix} \cos\Omega t & -\sin\Omega t \\ \sin\Omega t & \cos\Omega t \end{bmatrix}\begin{bmatrix} Q_{11}(0) & Q_{12}(0) \\ Q_{12}(0) & Q_{22}(0) \end{bmatrix} \times \begin{bmatrix} \cos\Omega t & \sin\Omega t \\ -\sin\Omega t & \cos\Omega t \end{bmatrix}. \tag{48}$$

Equation (48) has the familiar form of the transformation equation for a second-rank tensor under a rotation of spatial coordinates. *The quadrupole term in the vorticity expansion (7) rotates in space at the frequency* Ω. For small viscosity ν, the frequency Ω in Eq. (43) decreases slowly with time. In the inviscid point vortex limit $\nu \to 0$ and $a \to 0$, the rotation frequency Ω reduces to the Kirchhoff theory value

$$\Omega = \frac{\gamma}{4\pi\ell^2}. \tag{49}$$

Thus, the quadrupole term in our vorticity expansion (7) is sufficient to capture the rotation associated with the Kirchhoff point vortex model. For finite vortex core size a and nonzero viscosity ν, the quadrupole term of our expansion also captures the modification of the rotation frequency due to the finite core size, the viscous retardation of the rotation frequency, and the viscous diffusion of the vorticity field.

We note that Eq. (48) is the solution of Eq. (21) for any initial conditions that imply $D_j = 0$. For $D_j \neq 0$, the solution of Eq. (21) consists of Eq. (48) plus a particular solution corresponding to the nonhomogeneous dipole term.

IV. LIGHTHILL EQUATION FOR SOUND GENERATION

The Lighthill equation for sound generated by a homentropic low Mach number fluid flow is[2]

$$\left(\frac{1}{c_0^2}\frac{\partial^2}{\partial t^2} - \nabla^2\right)p = \rho_0\frac{\partial^2(u_i u_j)}{\partial x_i \partial x_j}, \tag{50}$$

where p is the acoustic pressure perturbation, ρ_0 is the density of the undisturbed fluid, c_0 is the speed of sound in the undisturbed fluid, and u_i is the incompressible component of the fluid velocity vector. The source term on the right-hand side of Eq. (50) can be written as

$$S = \rho_0\frac{\partial^2(u_i u_j)}{\partial x_i \partial x_j} = \rho_0\,\boldsymbol{\nabla}\cdot(\boldsymbol{\omega}\times\mathbf{u}) + \rho_0\boldsymbol{\nabla}^2\left(\frac{1}{2}u^2\right), \tag{51}$$

where $\boldsymbol{\omega}$ is the vorticity vector corresponding to the velocity vector $\mathbf{u}$. It is argued in Ref. 2 that the first term on the right-hand side of Eq. (51) is the principal source of sound at low Mach numbers. Our series expansions for $\boldsymbol{\omega}$ and $\mathbf{u}$ in Sec. II can therefore be used directly in the Lighthill equation to compute the sound generated by a two-dimensional vorticity field. We will use the Lighthill equation to compute the sound produced by the Gaussian vortex pair in the previous section, and compare our result to a previous result[2,7] which assumes the Kirchhoff theory of point vortex motion.

In order to obtain a far-field solution for the acoustic pressure p in Eq. (50), we expand the principal source term on the right-hand side as

$$S = \rho_0\,\boldsymbol{\nabla}\cdot(\boldsymbol{\omega}\times\mathbf{u}) = S^{(0)}(t)\delta(\mathbf{r}) + S_p^{(1)}(t)\frac{\partial}{\partial x_p}\delta(\mathbf{r}) + S_{pq}^{(2)}(t) \times \frac{\partial^2}{\partial x_p \partial x_q}\delta(\mathbf{r}) + \cdots, \tag{52}$$

where

$$S^{(0)}(t) = \int\int d^2\mathbf{r}S, \tag{53}$$

$$S_p^{(1)}(t) = -\int\int d^2\mathbf{r}x_p S, \tag{54}$$

$$S_{pq}^{(2)}(t) = \int\int d^2\mathbf{r}\frac{x_p x_q}{2}S. \tag{55}$$

For the Gaussian vortex pair, the coefficients $S^{(0)}$, $S_p^{(1)}$, and $S_{pq}^{(2)}$ in Eqs. (53)–(55) become

$$S^{(0)} = 0, \tag{56}$$

$$S_p^{(1)} = 0, \tag{57}$$

$$S_{pq}^{(2)}(t) = \frac{\rho_0\delta_{pq}}{24\pi}\left(-3\Gamma^2 + \frac{3\Gamma Q_{kk}}{\lambda^2} - \frac{(Q_{kk})^2}{\lambda^4} - \frac{Q_{kl}Q_{kl}}{12\pi\lambda^4}\right) + \rho_0\left(-\frac{\Gamma Q_{pq}}{4\pi\lambda^2} + \frac{Q_{kk}Q_{pq}}{12\pi\lambda^4} + \frac{Q_{pk}Q_{kq}}{6\pi\lambda^4}\right), \tag{58}$$

where Γ and Q_{jk} are given for this example by Eqs. (25) and (44)–(46), respectively. The terms proportional to δ_{pq} in Eq. (58) are nonoscillatory in time, and thus will not contribute to a propagating sound field.

The solution of the three-dimensional wave equation (50) is[2]

$$p(\mathbf{R},t) = \int\int\int d^3\mathbf{R}'\frac{S(\mathbf{R}',t-|\mathbf{R}-\mathbf{R}'|/c_0)}{4\pi|\mathbf{R}-\mathbf{R}'|} = \frac{\partial^2}{\partial x_p \partial x_q}\int_{-\infty}^{\infty} dx_3'\frac{S_{pq}^{(2)}(t-\sqrt{x_1^2+x_2^2+(x_3-x_3')^2}/c_0)}{4\pi\sqrt{x_1^2+x_2^2+(x_3-x_3')^2}}, \tag{59}$$

where $\mathbf{R}=(x_1,x_2,x_3)$, $\mathbf{R}'=(x_1',x_2',x_3')$, and the source S defined by Eqs. (51) and (52) has been used. For small viscosity, the source expansion term $S_{pq}^{(2)}(t)$ defined by Eqs. (58) and (44)–(46) contains sinusoidal functions which vary slowly in frequency and amplitude. The method of stationary phase can thus be used to evaluate the far-field value of the integral in Eq. (59). Evaluation of Eq. (59) for each of the three oscillatory terms in the second line of Eq. (58) gives

$$p = \frac{\rho_0\gamma^2\ell^2}{4\pi^2}\frac{\Omega^{*2}}{c_0^2\lambda^{*2}}\left(\frac{\pi c_0}{\Omega^* r}\right)^{1/2}\cos\left[2\Omega^*\left(t-\frac{r}{c_0}\right)-2\theta-\frac{\pi}{4}\right] - \frac{a^2}{\lambda^{*2}}\frac{\rho_0\gamma^2\ell^2}{8\pi^2}\frac{\Omega^{*2}}{c_0^2\lambda^{*2}}\left(\frac{\pi c_0}{\Omega^* r}\right)^{1/2}\cos\left[2\Omega^*\left(t-\frac{r}{c_0}\right)-2\theta-\frac{\pi}{4}\right]. \tag{60}$$

In Eq. (60), Ω^* denotes Ω evaluated at the retarded time $t-r/c_0$, and λ^* denotes λ evaluated at $t-r/c_0$. In the inviscid point vortex limit $\nu\to 0$ and $a\to 0$, the first term on the right-hand side of Eq. (60) reduces to the previous result in Refs. 2 and 7. The second term on the right-hand side of Eq. (60), which comes from the second and third terms in the second line of Eq. (58), is a correction due to the finite vortex core size a.

To test the accuracy of our result for the far-field acoustic pressure, we compare our formula with a computational result presented in Ref. 8. In Ref. 8, the far-field acoustic pressure generated by a corotating Gaussian vortex pair is computed by a direct numerical simulation of the compressible Navier-Stokes equations. In our notation, case two of Ref. 8 corresponds to the following parameter values:

$$\mathrm{Re} = \frac{\gamma}{\nu} = 226\times 10^3, \quad M_0 = \frac{\frac{\gamma}{2\pi a}}{c_0} = 0.56, \quad \frac{a}{\ell} = 0.402. \tag{61}$$

Numerical results are given in Ref. 8 for the nondimensional pressure $p/\rho_0 c_0^2$ as a function of the nondimensional time tc_0/ℓ, at a distance $r=63\ell$ from the origin. This value of r corresponds to 3.6 acoustic wavelengths, or $k_0 r=22.6$, which is a reasonable far-field distance for a numerical computation. Results for $p/\rho_0 c_0^2$ are given in Ref. 8 for the time interval $40<tc_0/\ell<120$. For $80<tc_0/\ell<120$, the transients associated with the initialization of the computation have vanished, and the computed nondimensional pressure is very nearly sinusoidal. The amplitude of this sinusoidal pressure can be compared with that predicted by our Eq. (60).

TABLE I. Comparison of results for the amplitude of the nondimensional acoustic pressure produced by a corotating Gaussian vortex pair ($k_0 r=22.6$, $80<tc_0/\ell<120$).

	Nondimensional pressure amplitude	% error
Full numerical simulation (Ref. 8)	0.0014	⋯
Kirchhoff-Powell (Ref. 7)	0.0023	64
Present work	0.0017	21

For the numerical values given in Eq. (61), the dimensional viscous parameter $4\nu t/\lambda_0^2 = 4\nu t/\ell^2$ which appears in Eqs. (6) and (47), satisfies the inequality

$$\frac{4\nu t}{\ell^2} = \frac{8\pi}{\mathrm{Re}}\frac{tc_0}{\ell}\frac{a}{\ell}M_0 < 3\times 10^{-3}. \tag{62}$$

The term $4\nu t/\ell_0^2$ is thus negligible in Eqs. (6) and (47), and viscous effects are unimportant in Eq. (60). Therefore,

$$\lambda^* \approx \lambda_0 = \ell \tag{63}$$

and

$$\Omega^* \approx \frac{\gamma}{4\pi\ell^2}\left(1-\frac{a^2}{2\ell^2}\right). \tag{64}$$

The numerical value of the nondimensional pressure amplitude given by Eq. (60) is compared to the computational result of Ref. 8 in Table I. Table I also shows the Kirchhoff-Powell result, obtained by using the Kirchhoff point vortex theory in the source S defined by Eq. (52), which is equivalent to Eq. (60) with $\nu=a=0$. Möhring's equation for aeroacoustic sound, discussed in detail in Ref. 8, gives a result which is essentially the same as Kirchhoff-Powell value. It is seen from Table I that our result is substantially more accurate than previous theoretical aeroacoustic predictions.

V. CONCLUSION

Our result for the corotating Gaussian vortex pair shows that our vorticity expansion can be used in conjunction with the Lighthill equation to provide an analytic prediction of

vortex generated sound that is substantially more accurate than previous theoretical results. Viscosity is unimportant for the case that we have considered, so our improvement over previous theory must be attributed to the fact that our expansion (7) makes direct use of the vorticity equation (2) to derive the dynamics of the Gaussian vortices with finite core size. While we obviously do not capture all the fluid phenomena included in the numerical computation of Ref. 8, the result of our example suggests that our expansion is a generally useful tool for the prediction of sound generated by a distributed vorticity field.

ACKNOWLEDGMENTS

The authors wish to thank Professor C. E. Wayne for many valuable and enjoyable discussions. The research of David Uminsky is partially supported by NSF grant DMS-0405724.

APPENDIX A: PROOF THAT THE DIPOLE COEFFICIENT D_j IS CONSTANT IN TIME

According to Eq. (18), the time derivative of the dipole coefficient $D_j(t)$ is

$$\dot{D}_j = -\delta_{ij}\int\int d^2\mathbf{r}u_i\omega. \tag{A1}$$

To evaluate this integral, we introduce the stream function ψ, which is related to the velocity vector u_i and the vorticity ω by

$$u_i = \varepsilon_{ik}\frac{\partial\psi}{\partial x_k}, \quad \omega = -\frac{\partial^2\psi}{\partial x_m\partial x_m}. \tag{A2}$$

Thus,

$$\dot{D}_j = \varepsilon_{jk}\int\int d^2\mathbf{r}\frac{\partial\psi}{\partial x_k}\frac{\partial^2\psi}{\partial x_m\partial x_m} = -\varepsilon_{jk}\int\int d^2\mathbf{r}\frac{\partial^2\psi}{\partial x_k\partial x_m}\frac{\partial\psi}{\partial x_m}$$
$$= -\varepsilon_{jk}\int\int d^2\mathbf{r}\frac{\partial}{\partial x_k}\left(\frac{1}{2}\frac{\partial\psi}{\partial x_m}\frac{\partial\psi}{\partial x_m}\right) = 0. \tag{A3}$$

The integration by parts and the final equality in Eq. (A3) are valid because the velocity $u_i = \partial\psi/\partial x_i$ decays at least as fast as $1/r$ as $r\to\infty$.

APPENDIX B: PROOF THAT THE QUADRUPOLE TRACE Q_{kk} is constant in time

The time derivative of the quadrupole trace Q_{kk} is given by Eq. (20) as

$$\dot{Q}_{kk} = -\int\int d^2\mathbf{r}\left(\frac{r^2}{2}-\frac{\lambda^2}{2}\right)\frac{\partial}{\partial x_i}(u_i\omega) = \int\int d^2\mathbf{r}x_iu_i\omega. \tag{B1}$$

We again introduce the stream function ψ, and write

$$\psi = \hat{\psi} + \tilde{\psi}, \tag{B2}$$

where $\hat{\psi}$ is the stream function associated with the monopole term in the series (7) and $\tilde{\psi}$ is the stream function associated with all the remaining terms. We have

$$\hat{\psi} = C_1\ln r, \tag{B3}$$

and

$$\tilde{\psi}\to C_2\frac{f(\theta)}{r} \quad \text{as } r\to\infty, \tag{B4}$$

which is important for the integrations by parts that follow. Equation (B1) becomes

$$\dot{Q}_{kk} = -\varepsilon_{ij}\int\int d^2\mathbf{r}x_i\frac{\partial\hat{\psi}}{\partial x_j}\frac{\partial^2\hat{\psi}}{\partial x_m\partial x_m} - \varepsilon_{ij}\int\int d^2\mathbf{r}x_i\frac{\partial\hat{\psi}}{\partial x_j}\frac{\partial^4\tilde{\psi}}{\partial x_m\partial x_m} - \varepsilon_{ij}\int\int d^2\mathbf{r}x_i\frac{\partial\tilde{\psi}}{\partial x_j}\frac{\partial^2\hat{\psi}}{\partial x_m\partial x_m} - \varepsilon_{ij}\int\int d^2\mathbf{r}x_i\frac{\partial\tilde{\psi}}{\partial x_j}\frac{\partial^2\tilde{\psi}}{\partial x_m\partial x_m}. \tag{B5}$$

Since $\hat{\psi}$ is a function of r only, the first two integrands on the right-hand side of Eq. (B5) depend on the symmetric tensor x_ix_j; the integrals are thus symmetric in i and j, and the contraction with the antisymmetric tensor ε_{ij} implies that the first two integrals on the right-hand side of Eq. (B5) vanish.

The third integral on the right-hand side of Eq. (B5) is

$$-\varepsilon_{ij}\int\int d^2\mathbf{r}x_i\frac{\partial\tilde{\psi}}{\partial x_j}\frac{\partial^2\hat{\psi}}{\partial x_m\partial x_m} = \varepsilon_{ij}\int\int d^2\mathbf{r}\frac{\partial\tilde{\psi}}{\partial x_j}x_i\delta^{\lambda} = -\varepsilon_{ij}\int\int d^2\mathbf{r}\tilde{\psi}\frac{\partial}{\partial x_j}(x_i\delta^{\lambda})$$
$$= -\varepsilon_{ij}\int\int d^2\mathbf{r}\tilde{\psi}\left(\delta_{ij}\delta^{\lambda}+x_ix_j\frac{\delta^{\lambda'}}{r}\right), \tag{B6}$$

which is again zero due to the contraction of an antisymmetric tensor with two symmetric tensors.

The fourth integral on the right-hand side of Eq. (B5) is

$$-\varepsilon_{ij}\int\int d^2\mathbf{r}x_i\frac{\partial\tilde{\psi}}{\partial x_j}\frac{\partial^2\tilde{\psi}}{\partial x_m\partial x_m} = \varepsilon_{ij}\int\int d^2\mathbf{r}\tilde{\psi}\frac{\partial}{\partial x_j}\left(x_i\frac{\partial^2\tilde{\psi}}{\partial x_m\partial x_m}\right)$$
$$= \varepsilon_{ij}\int\int d^2\mathbf{r}\tilde{\psi}\left(\delta_{ij}\frac{\partial^2\tilde{\psi}}{\partial x_m\partial x_m}+x_i\frac{\partial^3\tilde{\psi}}{\partial x_j\partial x_m\partial x_m}\right)$$
$$= \varepsilon_{ij}\int\int d^2\mathbf{r}\tilde{\psi}x_i\frac{\partial^3\tilde{\psi}}{\partial x_j\partial x_m\partial x_m}$$
$$= \varepsilon_{ij}\int\int d^2\mathbf{r}\frac{\partial^2}{\partial x_m\partial x_m}(\tilde{\psi}x_i)\frac{\partial\tilde{\psi}}{\partial x_j} = \varepsilon_{ij}\int\int d^2\mathbf{r}\left(2\frac{\partial\tilde{\psi}}{\partial x_m}\delta_{im}+\frac{\partial^2\tilde{\psi}}{\partial x_m\partial x_m}x_i\right)\frac{\partial\tilde{\psi}}{\partial x_j} = 2\varepsilon_{mj}\int\int d^2\mathbf{r}\frac{\partial\tilde{\psi}}{\partial x_m}\frac{\partial\tilde{\psi}}{\partial x_j}$$

$$+\varepsilon_{ij}\int\int d^2\mathbf{r}x_i\frac{\partial^2\tilde{\psi}}{\partial x_m\partial x_m}\frac{\partial\tilde{\psi}}{\partial x_j}=\varepsilon_{ij}\int\int d^2\mathbf{r}x_i\frac{\partial^2\tilde{\psi}}{\partial x_m\partial x_m}\frac{\partial\tilde{\psi}}{\partial x_j}. \tag{B7}$$

The final line of Eq. (B7) states that the fourth integral on the right-hand side of Eq. (B5) is equal to the negative of itself, and is therefore zero. We have now shown that each of the four terms on the right-hand side of Eq. (B5) vanishes, and that Q_{kk} is constant in time.

[1]M. J. Lighthill, "On sound generated aerodynamically," Proc. R. Soc. London, Ser. A **211**, 564–586 (1952).

[2]M. S. Howe, *Theory of Vortex Sound* (Cambridge University Press, New York, 2003).

[3]H. Lamb, *Hydrodynamics* (Dover, New York, 1945), pp. 591–592.

[4]T. Gallay and C. E. Wayne, "Invariant manifolds and the long-time asymptotics of the Navier-Stokes and vorticity equations on $\mathbf{R}^2$," Arch. Ration. Mech. Anal. **163**, 209–258 (2002).

[5]H. Grad, "Note on *N*-dimensional hermite polynomials," Commun. Pure Appl. Math. **2**, 325–330 (1949).

[6]A. Sommerfeld, *Mechanics of Deformable Bodies* (Academic, New York, 1964), pp. 154–161.

[7]A. Powell, "Theory of vortex sound," J. Acoust. Soc. Am. **36**, 177–195 (1964).

[8]B. E. Mitchell, S. K. Lele, and P. Moin, "Direct computation of the sound from a compressible co-rotating vortex pair," J. Fluid Mech. **285**, 181–202 (1995).

Theoretical simulation of electroacoustic borehole logging in a fluid-saturated porous formation

Hengshan Hu[a] and Wei Guan
Department of Astronautics and Mechanics, Harbin Institute of Technology, Postbox 344, 92 West Dazhi Street, Harbin, 150001, China

Jerry M. Harris
Department of Geophysics, Stanford University, Stanford, California 94305

(Received 13 September 2006; revised 29 March 2007; accepted 9 April 2007)

Electroacoustic (E-A) logging describes the acoustic response to an electromagnetic (EM) source in a fluid-filled borehole surrounded by a porous medium. The E-A response is simulated by two different methods in this paper. In the coupled method, the EM field and the acoustic field are modeled using Pride's model, which couples Maxwell's equations and Biot's equations. In the uncoupled method, the EM field is uninfluenced by the converted acoustic field, resulting in separate acoustic formulation with an electrokinetic source term derived from the primary EM field. The difference of the transient full waveforms between the above two methods is remarkably small for all examples, thus confirming the validity of using the computationally simpler uncoupled method. It is shown from the simulated waveforms that an EM-accompanying acoustic field is coupled to the EM field and appears with an apparent phase velocity of the EM wave in the formation. Acoustic waves with the conventional acoustic velocities are also seen in the converted full waveforms. For the sandstone models used in this paper, when permeability is less than 1 Darcy, the E-A Stoneley wave amplitude increases with porosity, which is different from that in conventional acoustic-to-acoustic logging.
 [DOI: 10.1121/1.2735809]

PACS number(s): 43.30.Ky, 43.20.Bi, 43.40.Ph [RAS] Pages: 135–145

I. INTRODUCTION

In a fluid-saturated porous medium, acoustic (or seismic) and electromagnetic (EM) waves are coupled because of electrokinetic effect. The electrokinetic effect is related to the electric double layer (EDL) at the solid-fluid interface (Pride and Morgan, 1991) and the flow of pore fluid relative to the porous solid matrix. When an acoustic wave propagates through a fluid-saturated porous medium, relative fluid flow occurs (Biot, 1956, 1962), which in turn carries excess ions in the EDL, causing convection electric current and a streaming potential. This phenomenon is called acoustoelectric (A-E) conversion. Reversely, when an EM wave propagates, the electric field exerts forces on the ions in the EDL, which causes the relative flow between the fluid and the solid phases, producing pressure gradients and an acoustic wave. This phenomenon is known as electroacoustic (E-A) conversion.

Ivanov (1940) measured the electric signal caused by the electrokinetic effect. In order to explain the phenomenon observed by Ivanov, Frenkel (1944) established a theoretical model of poro-acoustics and attempted to study the relation between the relative flow induced by a seismic wave and the flow-induced electric field. He predicted that an electric field accompanies a compressional seismic wave in homogeneous porous medium. Nevertheless, he did not allow for the full set of Maxwell's equations so that he mistakenly concluded that shear waves do not induce EM disturbances. In the following 50 years, many experimental observations were conducted (such as Martner and Sparks, 1959; Broding *et al.*, 1963; and Long and Rivers, 1975), but few theoretical developments (e.g., Neev and Yeatts, 1989) were reported.

Substantial developments were achieved since the 1990s, due to fast development in signal recording and processing techniques and due to ever-increasing demand for earth resources. Pride (1994) derived from first principles the governing equations for coupled acoustic and EM fields in homogeneous porous media, which is a combination of Biot's equations of poroelasticity (Biot, 1956, 1962) and Maxwell's equations of electromagnetism. Pride and Haartsen (1996) obtained expressions for electroseismic plane waves in a homogeneous porous medium. Haartsen and Pride (1997) presented the numerical simulations of the seismoelectric fields excited by an acoustic point-source in the fluid-saturated stratified porous medium. There are two kinds of fields shown to exist: one is the stationary electric and magnetic fields accompanying seismic waves during their propagation, with no extent outside of the seismic pulses; the other is the independently propagating EM wave generated by seismic waves when traveling across medium interfaces. These important publications provided a theoretical framework to interpret the experimental observations. Along with the developments of theoretical studies, many field and laboratory studies as well as numerical simulations were reported. Thompson and Gist (1993), Butler *et al.* (1996), and Mikhailov *et al.* (1997), implemented the field studies about

[a]Electronic mail: wave_hu@yahoo.com

electrokinetic effect and demonstrated the above-mentioned two kinds of EM fields induced by a seismic source. While their studies revealed the potential use of wave-related electrokinetic effects in mapping subsurface fluid flow and in obtaining other useful information about the subsurface, the exploration depth is limited by the weak electric signals, resulting from the propagation attenuation from the source to the exploration target and back to the receivers on the ground surface. A-E logging and E-A logging were brought forward in the 1990s, which have much smaller source to target and target to receiver distances so that the exploration depth is not limited. Both the source and receiver are located in the borehole, and the formation to be measured is in the depth interval between the receiver and transmitter. In a series of publications, Zhu *et al.* (1999) and Zhu and Toksöz (2003, 2005) measured the induced EM signals of A-E logging by the use of scaled model wells in the laboratory. Hu (2000) and Hu and Liu (2002) numerically simulated the axisymmetric fields of A-E logging in a homogeneous fluid-saturated porous medium based on Pride's theory, and both the stationary and propagating EM fields were seen in their simulated waveforms.

Most of the above-mentioned studies are concerned with A-E conversion of the electrokinetic effect. Zhu *et al.* (1999) recorded distinct E-A Stoneley waves induced by an electrode in his scaled model wells. His experimental result showed the possibility of a new borehole logging technique that could explore more physical properties related to the fluid and its flow in the porous formations. Recently Thompson (2005) and Thompson *et al.* (2005) reported field experiments about E-A conversion. They summarized the results of E-A field tests in three locations and concluded that E-A conversion is a sensitive indicator of hydrocarbon saturation, which revealed potential applications of E-A conversion. Nevertheless, we have not seen published theoretical papers on E-A logging.

In this paper we simulate the acoustic field excited by a vertical electric dipole in the borehole due to electrokinetic effect in the porous formation. The major objective is to derive the E-A wave fields analytically and analyze the full waveforms. The paper is organized as below. In Sec. II, we provide two different methods for simulating E-A logging. In the first three subsections the acoustic field and EM field in the formation are seen as coupled and are formulated based on Pride's equations and boundary conditions. In the last subsection, the EM field is formulated separately by ignoring the influence of the electrokinetically induced acoustic field on the EM field itself, and the acoustic field is then formulated by taking the EM disturbances as a source term. In Sec. III, we calculate the acoustic as well as the electric waveforms for typical logging situations using expressions derived in Sec. II. Transient full waveforms are obtained by the above two methods and are compared. We continue to analyze the interrelationships between the waveforms and the porosity and permeability of the formation.

II. FORMULATION

A. Fields in the porous formation

The governing equations that describe the coupling between the acoustic and electric fields in homogeneous fluid-saturated porous medium were derived by Pride (1994) and can be expressed as follows, assuming $e^{-i\omega t}$ time dependence of all fields:

$$\nabla \times \mathbf{E} = i\omega \mathbf{B}, \tag{1}$$

$$\nabla \times \mathbf{H} = -i\omega \mathbf{D} + \mathbf{J}, \tag{2}$$

$$\mathbf{B} = \mu \mathbf{H}, \tag{3}$$

$$\mathbf{D} = \varepsilon \mathbf{E}, \tag{4}$$

$$\mathbf{J} = \sigma \mathbf{E} + L(-\nabla p + \omega^2 \rho_f \mathbf{u}), \tag{5}$$

$$-i\omega \mathbf{w} = L\mathbf{E} + (-\nabla p + \omega^2 \rho_f \mathbf{u})\kappa/\eta, \tag{6}$$

$$\boldsymbol{\nabla} \cdot \boldsymbol{\tau} = -\omega^2(\rho \mathbf{u} + \rho_f \mathbf{w}), \tag{7}$$

$$\boldsymbol{\tau} = (H - 2G)(\nabla \cdot \mathbf{u})\mathbf{I} + C(\nabla \cdot \mathbf{w})\mathbf{I} + G(\nabla \mathbf{u} + \nabla \mathbf{u}^{\mathbf{T}}), \tag{8}$$

$$-p = C\nabla \cdot \mathbf{u} + M\nabla \cdot \mathbf{w}, \tag{9}$$

where **E**, **D**, **J**, **B**, and **H** are the electric field, electric flux density, electric current density, magnetic flux density, and magnetic field, respectively, **u** is the displacement of the solid phase, **w** is the relative flow between the fluid and the solid phase, $\boldsymbol{\tau}$ is the bulk stress tensor, **I** is the identity tensor, p is the pore fluid pressure, ε is the permittivity of the formation, μ is the magnetic permeability of the formation and is assumed to equal to that of vacuum in this paper, ρ_f and η are density and viscosity of the pore fluid, respectively, ρ is the density of the formation, G is the shear modulus of the formation, H, C, and M are porous-medium moduli as defined by Biot (1962), κ is the dynamic permeability defined by Johnson *et al.* (1987), σ and L are the conductivity and electrokinetic coupling coefficient of the formation, respectively. Expressions for σ and L used in our calculation can be found in the paper by Pride (1994). In the low frequency limit, the expression for L is given by

$$L = -\frac{\phi}{\alpha_\infty}\frac{\varepsilon_f \varsigma}{\eta}\left(1 - 2\alpha_\infty \frac{\tilde{d}}{\Lambda}\right), \tag{10}$$

where ε_f is the permittivity of the fluid in the porous formation; α_∞ is the tortuosity; ς is called the zeta potential, which is the electric potential at the shear plane, the surface that separates the two layers, i.e., the adsorbed layer and the diffuse layer in the EDL; and Λ and $\tilde{d}$ are the weighted volume-to-surface radio and a length that is equal to or less than the Debye length, respectively, defined by Pride (1994). Electrokinetic coupling between acoustic fields and EM fields is reflected in Eqs. (5) and (6) through the terms with the coefficient L. If L is set to zero, the two equations decouple into Biot's and Maxwell's equations. The coefficient L is proportional to porosity in the low frequency limit [see Eq. (10)]

and is approximately proportional to porosity at higher frequencies.

In E-A logging the source is located in the borehole so that there are neither acoustic nor EM sources in the fluid-saturated porous formation. From Eqs. (1)–(9) one can derive equations in terms of **u**, **w**, and **E** as below:

$$(H-G)\nabla\nabla\cdot\mathbf{u}+G\nabla^2\mathbf{u}+\omega^2\rho\mathbf{u}+C\nabla\nabla\cdot\mathbf{w}+\omega^2\rho_f\mathbf{w}=0, \tag{11}$$

$$C\nabla\nabla\cdot\mathbf{u}+\omega^2\rho_f\mathbf{u}+M\nabla\nabla\cdot\mathbf{w}+\omega^2\bar{\rho}\mathbf{w}-i\omega\bar{\rho}L\mathbf{E}=0, \tag{12}$$

$$\nabla\nabla\cdot\mathbf{E}-\nabla^2\mathbf{E}-\omega^2\mu\bar{\varepsilon}\mathbf{E}+i\omega^3\mu\bar{\rho}L\mathbf{w}=0, \tag{13}$$

where

$$\bar{\varepsilon}=\varepsilon+i\sigma/\omega-\bar{\rho}L^2 \tag{14}$$

is the effective complex permittivity of the formation, and

$$\bar{\rho}=i\eta/(\omega\cdot\kappa) \tag{15}$$

is the effective density for relative flow.

According to the Helmholtz theorem, the fields **u**, **w**, and **E** in Eqs. (11)–(13) can be expressed in terms of potential functions. It can be shown through complex mathematical operations as in the paper by Pride and Haartsen (1996) that there are four different modes of waves in an infinite homogeneous fluid-saturated porous medium. Two of them are rotation-free waves, i.e., the fast compressional wave whose wave number is k_{pf}, and the slow compressional wave whose velocity is k_{ps}. The other two are divergence-free waves, i.e., the shear wave whose wave number is k_{sh}, and the EM wave whose wave number is k_{em}. The fields **u**, **w**, and **E** are coupled due to electrokinetic effect, and each of them can travel with the above four different wave numbers. The amplitudes and phases of the three fields are different. For example, **u**, **w**, and **E** can all travel with the fast compressional wave number, with $\mathbf{w}_{pf}=\alpha_{pf}\mathbf{u}_{pf}$ and $\mathbf{E}_{pf}=\beta_{pf}\mathbf{u}_{pf}$. The amplitude ratio and phase shift are reflected by α_{pf} between $\mathbf{w}_{pf}$ and $\mathbf{u}_{pf}$, and by β_{pf} between $\mathbf{E}_{pf}$ and $\mathbf{u}_{pf}$. The expressions for the wave numbers k_i and the factors α_i, β_i ($i=pf,ps,sh,em$) are given by Pride and Haartsen (1996) and Hu (2000).

Let us study the axisymmetric fields excited by a vertical electric dipole along the borehole axis. We adopt the cylindrical coordinate system (r, z, θ), with the z axis being the borehole axis, and the origin coinciding with the center of the dipole. In the porous formation the TM wave (the transverse magnetic wave, i.e., the magnetic field is perpendicular to the $r-z$ plane and the electric field is in the $r-z$ plane) is coupled with the P wave and the SV wave (the shear wave whose motion is in the $r-z$ plane) due to the electrokinetic effect. As a result, an axisymmetric EM field is associated with an axisymmetric acoustic field. From the above analysis, we can express **u**, **w**, and **E** by potential functions as below:

$$\mathbf{u}=A_{pf}\nabla\phi_{pf}+A_{ps}\nabla\phi_{ps}+A_{sh}\nabla\times(\Gamma_{sh}\mathbf{e}_\theta)+A_{em}\nabla\times(\Gamma_{em}\mathbf{e}_\theta), \tag{16}$$

$$\mathbf{w}=\alpha_{pf}A_{pf}\nabla\phi_{pf}+\alpha_{ps}A_{ps}\nabla\phi_{ps}+\alpha_{sh}A_{sh}\nabla\times(\Gamma_{sh}\mathbf{e}_\theta)+\alpha_{em}A_{em}\nabla\times(\Gamma_{sh}\mathbf{e}_\theta), \tag{17}$$

$$\mathbf{E}=\beta_{pf}A_{pf}\nabla\phi_{pf}+\beta_{ps}A_{ps}\nabla\phi_{ps}+\beta_{sh}A_{sh}\nabla\times(\Gamma_{sh}\mathbf{e}_\theta)+\beta_{em}A_{em}\nabla\times(\Gamma_{sh}\mathbf{e}_\theta), \tag{18}$$

where the potentials are

$$\phi_m=K_0(\eta_m r)e^{ikz}\quad(m=pf,ps), \tag{19}$$

$$\Gamma_n=K_1(\eta_n r)e^{ikz}\quad(n=sh,em), \tag{20}$$

where k is the axial wave number, η_m and η_n are the radial wave numbers, i.e., $\eta_m=\sqrt{k^2-k_m^2}$ and $\eta_n=\sqrt{k^2-k_n^2}$, I_n and K_n are modified Bessel functions of the nth order. The coefficients A_i ($i=pf,ps,sh,em$) can be determined by the boundary conditions on the wall, and $\mathbf{e}_\theta$ is the unit vector in the θ direction.

Once **u**, **w**, and **E** are known, one can obtain the stress $\boldsymbol{\tau}$ and the E-A pressure p from Eqs. (8) and (9). The magnetic field **H** can be derived from Faraday's law

$$\mathbf{H}=\frac{1}{i\omega\mu}\nabla\times\mathbf{E}. \tag{21}$$

B. Fields in the borehole fluid

To simplify our simulation, we use an oscillating electric dipole as the source in E-A logging. It may be a short line conductor in which an alternating current flows. We assume the length l of the conductor is very short compared with the wavelength ($l\ll\lambda$) and the amplitude I_0 of the current is uniform along the entire length l. Assuming $e^{-i\omega t}$ time dependence, the radiation field of the dipole can be represented by a magnetic vector potential **A** as (Guru and Hiziroğlu, 2004)

$$\mathbf{A}=\mathbf{e}_z\frac{\mu_b p_e}{4\pi}\frac{1}{R}e^{ik_{be}R}, \tag{22}$$

where **A** is related to the magnetic flux density **B** and the electric field **E** by

$$\mathbf{B}=\nabla\times\mathbf{A}, \tag{23}$$

$$\mathbf{E}=i\omega\left(\mathbf{A}+\frac{1}{k_{be}^2}\nabla\nabla\cdot\mathbf{A}\right), \tag{24}$$

where μ_b is the magnetic permeability of the borehole fluid, and is assumed to equal to the magnetic permeability of vacuum, so we rewrite μ_b as μ in what follows, $R=\sqrt{r^2+z^2}$, $p_e=I_0 l$, $\mathbf{e}_z$ is the unit vector in the z direction and where

$$k_{be}^2=\omega^2\mu\varepsilon_b+i\omega\mu\sigma_b, \tag{25}$$

k_{be} is the EM wave number of the borehole fluid, ε_b and σ_b are dielectric constant and conductivity of the borehole fluid, respectively.

According to the theory of Bessel functions, $(1/R)e^{ik_{be}R}$ can be represented as

$$\frac{1}{R}e^{ik_{be}R} = \frac{1}{\pi}\int_{-\infty}^{\infty} K_0(\eta_{be}r)e^{ikz}\,dk, \tag{26}$$

where $\eta_{be}=\sqrt{k^2-k_{be}^2}$. So we have

$$\mathbf{A} = \mathbf{e}_z\frac{\mu p_e}{4\pi^2}\int_{-\infty}^{\infty} K_0(\eta_{be}r)e^{ikz}\,dk. \tag{27}$$

Substituting Eq. (27) into Eq. (23) and using the constitutive laws, i.e., Eq. (3), we can get the solution for the magnetic field, of which the only nonzero component is

$$H_\theta^R = \frac{p_e}{4\pi^2}\int_{-\infty}^{\infty} \eta_{be}K_1(\eta_{be}r)e^{ikz}\,dk, \tag{28}$$

where the superscript "R" represents the radiation field. Similarly, substituting Eq. (27) into Eq. (24), we get the solutions for the electric field,

$$E_r^R = \frac{\mu p_e}{4\pi^2}\int_{-\infty}^{\infty} \omega\frac{k\eta_{be}}{k_{be}^2}K_1(\eta_{be}r)e^{ikz}\,dk, \tag{29}$$

$$E_z^R = \frac{\mu p_e}{4\pi^2}\int_{-\infty}^{\infty} i\omega\left(1-\frac{k^2}{k_{be}^2}\right)K_0(\eta_{be}r)e^{ikz}\,dk, \tag{30}$$

with the θ component of the electric field being zero.

The solution for the total EM field in the borehole is composed of the particular solution, which represents the radiation field from the dipole, i.e., Eqs. (28)–(30), and the solution of the associated homogeneous equations, which represents a source-free field that obeys the Helmholtz equation,

$$\nabla^2\mathbf{E} + k_{be}^2\mathbf{E} = 0. \tag{31}$$

Using the fact that the divergence of the electric field vanishes in the source-free region, the solution of Eq. (31) can be expressed as

$$E_r^O = \int_{-\infty}^{\infty} B_e(k,\omega)I_1(\eta_{be}r)e^{ikz}\,dk, \tag{32}$$

$$E_z^O = -\int_{-\infty}^{\infty} \frac{\eta_{be}B_e}{ik}I_0(\eta_{be}r)e^{ikz}\,dk, \tag{33}$$

where the superscript "O" represents the source-free field. The θ component of the electric field is zero because of axisymmetry. From Eq. (21) we can get the magnetic field, of which the only nonzero component is

$$H_\theta^O = \frac{1}{\omega\mu}\int_{-\infty}^{\infty}\left(k-\frac{\eta_{be}^2}{k}\right)B_eI_1(\eta_{be}r)e^{ikz}\,dk. \tag{34}$$

The total EM field in the borehole can be obtained by adding together the above particular solution and the solution to the homogenous equation, i.e.,

$$E_r = E_r^R + E_r^O, \quad E_z = E_z^R + E_z^O, \quad H_\theta = H_\theta^R + H_\theta^O. \tag{35}$$

The acoustic field is not coupled with the EM field in the borehole. It exists due to the boundary condition requirements at the borehole wall. The displacement and pressure in the borehole can be represented by a scalar potential ϕ_{bm} as

$$\mathbf{u} = \nabla\phi_{bm}, \tag{36}$$

$$p = \rho_b\omega^2\phi_{bm}, \tag{37}$$

where ρ_b is the density of the borehole fluid, and ϕ_{bm} obeys the homogeneous Helmholtz equation,

$$\nabla^2\phi_{bm} + k_{bm}^2\phi_{bm} = 0, \tag{38}$$

where

$$k_{bm} = \frac{\omega}{V_{bm}}\left(1+\frac{i}{2Q_{bm}}\right) \tag{39}$$

is the acoustic wave number of the borehole fluid, and V_{bm} and Q_{bm} are the acoustic velocity and quality factor in the borehole fluid, respectively. The solution of Eq. (38) in the frequency domain can be expressed as

$$\phi_{bm}(r,z,\omega) = \frac{1}{\pi}\int_{-\infty}^{\infty} A_m(k,\omega)I_0(\eta_{bm}r)e^{ikz}\,dk, \tag{40}$$

where $\eta_{bm}=\sqrt{k^2-k_{bm}^2}$. Substituting Eq. (40) into Eqs. (36) and (37), one obtains

$$p = \frac{\rho_b\omega^2}{\pi}\int_{-\infty}^{\infty} A_m(k,\omega)I_0(\eta_{bm}r)e^{ikz}\,dk, \tag{41}$$

$$u_r = \frac{1}{\pi}\int_{-\infty}^{\infty} A_m(k,\omega)\,\eta_m I_1(\eta_{bm}r)e^{ikz}\,dk, \tag{42}$$

$$u_z = \frac{1}{\pi}\int_{-\infty}^{\infty} A_m(k,\omega)I_0(\eta_{bm}r)ike^{ikz}\,dk. \tag{43}$$

C. Boundary conditions

The coefficients A_i $(i=pf,ps,sh,em)$ in Eqs. (16)–(18), B_e in Eqs. (32)–(34), and A_m in Eqs. (40)–(43) can be determined from the following boundary conditions at the borehole wall:

$$u_{r0} = u_{r1} + w_{r1}, \tag{44}$$

$$p_0 = p_1, \tag{45}$$

$$-p_0 = \tau_{rr1}, \tag{46}$$

$$0 = \tau_{rz1}, \tag{47}$$

$$E_{z0} = E_{z1}, \tag{48}$$

$$H_{\theta0} = H_{\theta1}, \tag{49}$$

where quantities with subscript "0" are fields in the borehole, and quantities with subscript "1" are fields in the formation. The above six boundary conditions lead to the following linear equations in the frequency–wave-number domain,

$$\mathbf{MA} = \mathbf{B}, \tag{50}$$

where

$$\mathbf{A} = \{A_m, B_e, A_{pf}, A_{ps}, A_{sh}, A_{em}\}^T, \tag{51}$$

$$\mathbf{B} = \left\{ 0,0,0,0, \frac{i\omega\mu p_e}{4\pi^2} \times \left(1 - \frac{k^2}{k_{be}^2}\right) K_0(\eta_{be} r), \frac{i\omega\mu p_e}{4\pi^2} \eta_{be} K_1(\eta_{be} r) \right\}^T, \quad (52)$$

and the elements of **M** are shown in Appendix A.

D. The uncoupled method

In the above derivation the acoustic field and the EM field are coupled in the formation; the process of obtaining the coupled fields is complicated. It may be much more complicated, or even impossible, to formulate the coupled fields for more complex geometries or more complex source distributions. If the formation is assumed to be heterogeneous, it is difficult to get analytical expressions for the fields. So, in this subsection, we adopt a different method to simulate E-A logging without solving directly the coupled fields in the formation. We called this method the uncoupled method in order to distinguish it from the coupled method in the above study. We hope this uncoupled method will be a check to the coupled method in some way.

In E-A logging, the EM wave excited by the source induces the acoustic field in the porous formation and the converted acoustic field influences back on the EM field because of the electrokinetic effect. Nevertheless, if the reverse influence is weak enough, we can assume it is negligible. We will validate the feasibility of this assumption by comparison of the coupled and uncoupled waveforms in Sec. 3. When the influence is negligible, the second term on the right hand of Eq. (5) becomes zero; the EM field in the formation can be obtained separately by solving Maxwell's equations. According to the theory of electrodynamics, electric field **E** in the source-free formation must obey the following equations,

$$\nabla^2 \mathbf{E} + k_{em}^2 \mathbf{E} = 0, \quad (53)$$

$$\nabla \cdot \mathbf{E} = 0, \quad (54)$$

where

$$k_{em}^2 = \omega^2 \mu\varepsilon + i\omega\mu\sigma. \quad (55)$$

From Eqs. (53) and (54) and using Eq. (21), we obtain expressions for the EM field in the frequency domain,

$$E_z = \int_{-\infty}^{\infty} A_e \frac{\eta_{em}}{ik} K_0(\eta_{em} r) e^{ikz}\, dk, \quad (56)$$

$$E_r = \int_{-\infty}^{\infty} A_e K_1(\eta_{em} r) e^{ikz}\, dk, \quad (57)$$

$$H_\theta = \int_{-\infty}^{\infty} A_e \frac{1}{\omega\mu} \left(k - \frac{\eta_{em}^2}{k}\right) K_1(\eta_{em} r) e^{ikz}\, dk. \quad (58)$$

where the unknown coefficient A_e can be determined by the boundary conditions at the borehole wall. Recalling that the EM field is not coupled with the acoustic field in the borehole fluid because there is no electrokinetic effect in it, Eq. (35) remains valid for the EM field in the borehole, but the coefficient B_e in Eqs. (32)–(34) is different now. We can obtain B_e as well as A_e in Eqs. (56)–(58) by imposing the boundary conditions for the EM field at the borehole wall. These boundary conditions are expressed as a set of linear equations in Appendix B.

Upon knowing the EM field, we can determine the converted acoustic field. The acoustic field still satisfies Eqs. (11) and (12). Notice that the last term on the left side of Eq. (12) depends on the known electric field **E** and can be considered as the source of the acoustic field. If **E** becomes zero, Eq. (12) turns to

$$C \nabla \nabla \cdot \mathbf{u} + \omega^2 \rho_f \mathbf{u} + M \nabla \nabla \cdot \mathbf{w} + \omega^2 \bar{\rho} \mathbf{w} = 0. \quad (59)$$

Equations (11) and (59) are the homogeneous Biot equations for acoustic waves in porous medium and can be solved as in the paper by Rosenbaum (1974). Now we expound on how to find a set of particular solutions for Eqs. (11) and (12). Let $\mathbf{u}^0$ and $\mathbf{w}^0$ be the set of particular solutions that are related to the electric field **E** as

$$\mathbf{u}^0 = T_1 \mathbf{E}, \quad \mathbf{w}^0 = T_2 \mathbf{E}, \quad (60)$$

where the coefficients T_1 and T_2 are be to determined as follows.

According to Eq. (54), we have

$$\nabla \cdot \mathbf{u}^0 = 0, \quad \nabla \cdot \mathbf{w}^0 = 0. \quad (61)$$

By substituting Eqs. (60) and (61) into Eqs. (11) and (12) we get

$$G T_1 \nabla^2 \mathbf{E} + (\omega^2 \rho T_1 + \omega^2 \rho_f T_2) \mathbf{E} = 0, \quad (62)$$

$$(\omega^2 \rho_f T_1 + \omega^2 \bar{\rho} T_2 - i\omega\bar{\rho} L) \mathbf{E} = 0. \quad (63)$$

A comparison between Eq. (53) and (62) leads to

$$(\omega^2 \rho T_1 + \omega^2 \rho_f T_2)/(G T_1) = k_{em}^2. \quad (64)$$

And, from Eq. (63), we have

$$\omega^2 \rho_f T_1 + \omega^2 \bar{\rho} T_2 - i\omega\bar{\rho} L = 0, \quad (65)$$

for a nontrivial solution of **E**.

Solving simultaneously Eqs. (64) and (65), one obtains the coefficients T_1 and T_2:

$$T_1 = \frac{iL}{\omega} \left(\frac{G k_{em}^2}{\omega^2 \rho_f} - \frac{\rho}{\rho_f} + \frac{\rho_f}{\bar{\rho}} \right)^{-1}, \quad (66)$$

$$T_2 = \frac{i\omega\bar{\rho} L - \omega^2 \rho_f T_1}{\omega^2 \bar{\rho}}. \quad (67)$$

By this time we have a set of particular solutions. And the solution for the converted acoustic field in the formation can be obtained by superposing the solution of the homogenous equation and the set of particular solutions.

Similar to the coupled method, we can solve for the unknown coefficients in the expressions for the acoustic fields in the formation and in the borehole by imposing boundary conditions at the borehole wall. These conditions lead to a set of linear equations as is given in Appendix C. The expressions (41)–(43) for the acoustic field in the borehole remain valid for the uncoupled method as long as the

TABLE I. Input parameters for simulating E-A logging. The relationships between ϕ and K_b and G_b are assumed to obey the experimental results of Vernik (1994). ε_0 is the permittivity of vacuum.

Parameter	Property	Value
a	Borehole radius (m)	0.1
ϕ	Porosity (%)	20
κ_0	Static Darcy permeability (Darcy)	1
K_b	Frame bulk modulus (GPa)	14.39
G_b	Frame shear modulus (GPa)	13.99
K_s	Solid bulk modulus (GPa)	35.70
K_f	Pore fluid bulk modulus (GPa)	2.25
ρ_s	Solid density (Kg/m^3)	2650
ρ_f, ρ_b	Pore fluid density, borehole fluid density (Kg/m^3)	1000
ε_s	Solid permittivity	$4\varepsilon_0$
$\varepsilon_f, \varepsilon_b$	Pore fluid permittivity, borehole fluid permittivity	$80\varepsilon_0$
C_f, C_b	Pore fluid salinity, borehole fluid salinity (mol/L)	0.01
η	Pore fluid viscosity (Pa·s)	10^{-3}
V_{bm}	Acoustic velocity in borehole fluid (m/s)	1500
Q_{bm}	Quality factor in the borehole fluid	100
T	Temperature (K)	298
α_∞	Tortuosity	3

coefficient A_m is determined according to Appendix C.

III. NUMERICAL SIMULATIONS

To simulate the E-A logging response, we calculated the E-A pressure p as well as the three components of the EM field, i.e., radial and axial components of the electric field E_r and E_z, and circumferential magnetic field H_θ. Although, in real logging tools, an acoustic receiver can not be placed at the same point where an EM field is sensed, it is assumed in our calculation that all four quantities can be measured at any point in the borehole. We first calculate the four fields in the frequency domain expressed by Eqs. (28)–(30) and (32)–(35) in the coupled method or Eqs. (56)–(58) in the uncoupled method and Eq. (41), and then Fourier transform them to get the time-domain response. For example, p in the borehole is

$$p(t) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} p(\omega) S_0(\omega) e^{-i\omega t}\, d\omega, \tag{68}$$

where $p(\omega)$ is the frequency-domain response obtained from Eq. (41), and $S_0(\omega)$ is the spectrum of the source pulse function $s_0(t)$,

$$s_0(t) = \begin{cases} \frac{1}{2}\left[1 + \cos\frac{2\pi}{T_c}\left(t - \frac{T_c}{2}\right)\right]\cos 2\pi f_0\left(t - \frac{T_c}{2}\right), & 0 \leqslant t \leqslant T_c \\ 0, & t < 0 \quad or \quad t > T_c \end{cases}, \tag{69}$$

where f_0 is the center frequency and T_c is the pulse width. In all of the numerical simulation examples below, we adopt an f_0 of 6 kHz and a T_c of 0.5 ms. The source intensity p_e of the electric dipole is assumed to be 1 A·m. The formation of our model is assumed to be consolidated sandstone, and the parameter relationships between the porosity ϕ and the frame moduli K_b and G_b are assumed to obey the experimental results of Vernik (1994). The input parameters of the model for simulating E-A logging are listed in Table I unless explicitly stated elsewhere.

We performed the simulations and obtained the waveforms by the two different methods introduced above. Then we calculated the difference waveforms when the same input parameters are used in the two methods and found the largest relative difference in amplitude is in the order of 10^{-4} for all examples in the paper. Such a small relative difference cannot be seen in the plots of waveforms. We think the difference partially results from the truncation error due to limited computer digits, and partially comes from the difference in conversion mechanism (the reverse influence is ignored in the uncoupled method). This conclusion of remarkably small difference is important because of lack of reachable experimental data to be used as a reference, and makes us confident in our calculations. It also proved that the uncoupled method in which the acoustic field is calculated after the EM field being evaluated separately is feasible and accurate. Thus we do not need to point out which waveforms are simulated by which method in the following examples.

Now let us study the properties of the converted acoustic field during E-A logging. Figure 1 shows the full waveforms of E-A pressure p on the borehole axis, when the distance to the source z varies from 3.0 to 4.5 m. Obviously, there are

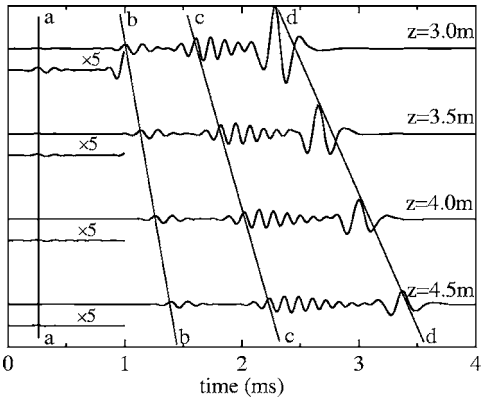


FIG. 1. Transient full waveforms of E-A pressure p on the borehole axis when the distance to the source z varies from 3 to 4.5 m. The waveforms are scaled to the wave magnitude at $z=3$ m. The waveforms in the first millisecond are magnified by five times.

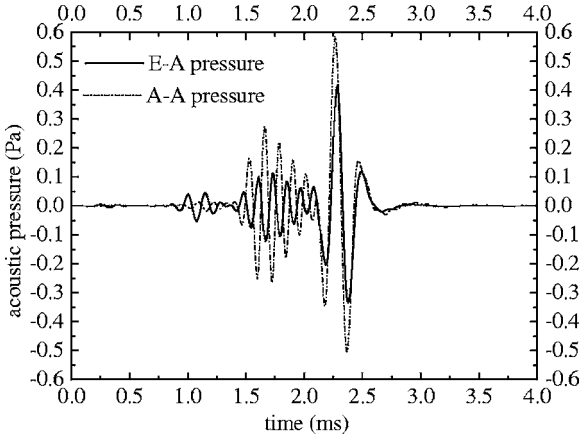


FIG. 2. Comparison between the waveforms of E-A pressure p and that in A-A logging. Both of the waveforms are calculated on the borehole axis at $z=3$ m.

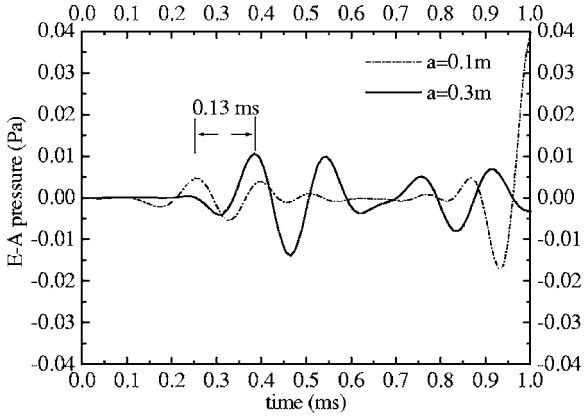


FIG. 3. Waveforms of E-A pressure p on the borehole axis at $z=3$ m when the borehole radii are 0.1 and 0.3 m. Only waves in the first millisecond are shown. The earliest arrival delays about 0.13 ms when the borehole radius changes from 0.1 to 0.3 m.

three different wave groups in the waveforms (shown by lines b-b, c-c, and d-d, respectively, in Fig. 1). In Fig. 2 we compare the waveform of E-A pressure p and that in acoustic-to-acoustic (A-A) logging. The waveform in A-A logging is calculated using Biot's theory, as done by Rosenbaum (1974). An acoustic point source as described in Tsang and Rader (1979) is used in the A-A logging simulation, with the peak pressure being 100Pa at a location of 0.01 m away from the point source. The same source function is used as that in E-A logging, with $f_0=6$ kHz, $T_c=0.5$ ms. Other required input parameters are the same as those in E-A logging and are given in Table I. It is seen that the three wave groups in the E-A logging waveform have the same wave velocities as those in A-A logging. From previous research on A-A logging, we conclude that the three wave groups are, in order of arrival time, the compressional wave group, the shear and pseudo-Rayleigh wave group, and the Stoneley wave group. These wave groups are originated at the borehole wall through conversion of EM energy to mechanical energy. Once generated, they propagate independently of the electric field.

When we magnify by five times the first millisecond of the waveforms in Fig. 1, we see a wave group (showed by line a-a in Fig. 1) that arrives earlier than the compressional wave group and reaches the four different receivers at almost the same time. In Fig. 3 we compare the first millisecond waveforms of E-A pressure p in boreholes with different radii. When the radius changes from 0.1 to 0.3 m, the earliest wave group delays 0.13 ms, which is the time for an acoustic wave in the borehole fluid to travel 0.2 m. We judge from this phenomenon that this wave group is a real physical signal rather than a spurious signal resulting from calculation. A further comparison is shown in Fig. 4 between the waveform of the electric field E_z and that of the E-A pressure p. The earliest acoustic disturbance, i.e., the wave group that arrives before the compressional wave, seems to accompany the electric field and to possess an apparent phase velocity of the EM wave in the formation. This phenomenon is consistent with an earlier theoretical prediction by Pride and Haartsen (1996). This wave group is induced by the conversion of the EM wave on the borehole wall at different depths and reflected back to the borehole fluid, which is named as an EM-accompanying acoustic group.

Now we investigate the effect of formation parameters on E-A logging. We focus our study on the effects of permeability and porosity in this paper. Shown in Fig. 5 are the waveforms of E-A pressure p for the permeability κ_0 being 0.01 and 10 Darcy. The amplitude of E-A pressure changes with permeability. Among all the wave groups, the E-A Stoneley wave group is the most sensitive to permeability. The E-A Stoneley wave amplitude decreases with the increase of permeability, which is similar to the case in A-A logging. From Figs. 5(a) and 5(b), the E-A Stoneley wave amplitude decreases 96.8% and 99.1% at the locations $z=3$ m and $z=4.5$ m, respectively, when permeability increases from 0.01 to 10 Darcy. The sensitivity to permeability increases with the distance to the source, since the acoustic propagation attenuation increases with permeability as well as the distance to the source.

Similar to the way of considering the effects of permeability above, we found that the E-A Stoneley wave group is also the most sensitive to porosity. In order to reveal possible relationships between the E-A Stoneley wave amplitude and

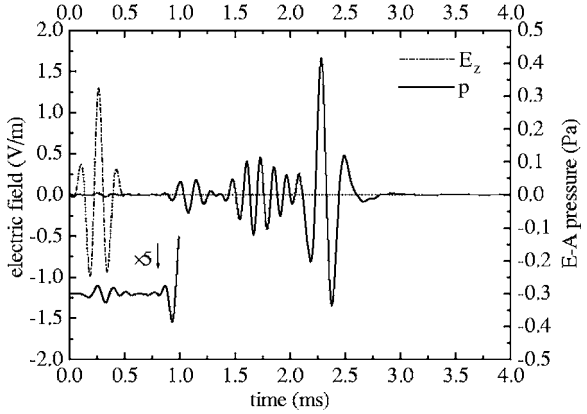


FIG. 4. Waveforms of E-A pressure p and electric field E_z on the borehole axis at $z=3$ m. Also shown in the figure is the front part of the waveform of E-A pressure p, which is magnified by five times for the comparison with the waveform of electric field E_z.

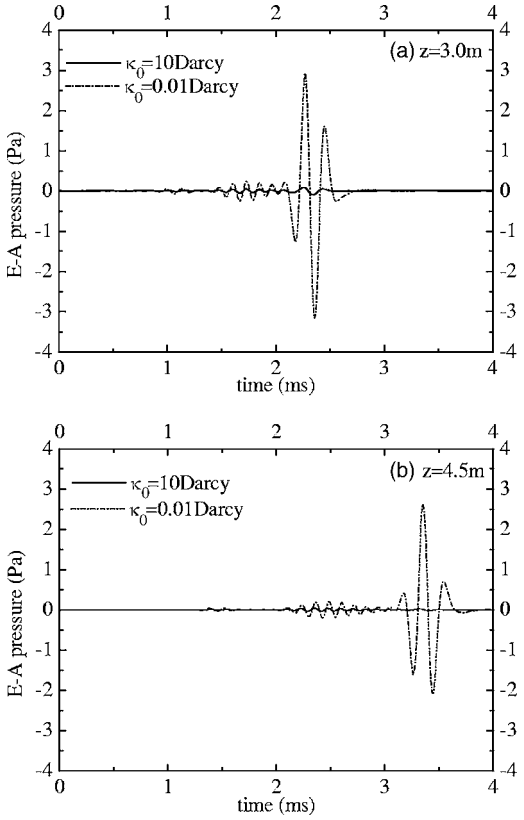


FIG. 5. Waveforms of E-A pressure p for permeabilities κ_0 of 10 and 0.01 Darcy. (a) Measured on the borehole axis at $z=3$ m. (b) Measured on the borehole axis at $z=4.5$ m.

permeability and porosity, and to discuss some different characteristics between E-A logging and A-A logging, calculations are performed under a series of combinations of permeability and porosity for E-A logging and A-A logging. The results are listed in Table II. To show the variations clearly, we plotted the results in Figs. 6 and 7 for E-A logging and A-A logging, respectively, according to the data in Table II. As shown in Fig. 6, when permeability is much lower than 1 Darcy, the E-A Stoneley wave amplitude is highly dependent on porosity as well as permeability: the higher the porosity, the larger the amplitude, and the higher the permeability, the smaller the amplitude. The effect of porosity on E-A Stoneley wave amplitude is different from that in A-A logging (see Fig. 7). This is not outside our expectation because two mechanisms together control the amplitude change of the E-A Stoneley wave. One is the E-A conversion ratio determined by the electrokinetic coupling coefficient L, and the other is the attenuation of the converted wave during its propagation in the porous medium. While the second mechanism is shared by A-A logging, the first mechanism belongs to E-A logging only. The A-A Stoneley wave amplitude decreases monotonically when porosity increases, because waves in a high porosity formation usually attenuate faster. In E-A logging, however, due to a larger L at a higher porosity, which is described by Eq. (10), and which means a stronger coupling between EM field and acoustic field, E-A pressure may increase with porosity. When permeability reaches 10 Darcy, the amplitude decreases with the increasing porosity (see Table II), revealing that the attenuation mechanism dominates amplitude variation. Thus there is a critical permeability where the amplitude of the Stoneley wave is invariant with porosity. From the data in Table II, it is calculated that E-A Stoneley wave amplitude decreases 68.6%, 79.5%, and 83.0% for porosity being 10%, 20%, and 25%, respectively, when permeability increases from 1 to 10 Darcy. But the A-A Stoneley wave amplitude de-

TABLE II. Stoneley wave amplitudes obtained from E-A logging and A-A logging with different porosities and permeabilities.

Porosity (%)	Permeability (Darcy)	Amplitude of Stoneley wave (Pa)	
		E-A logging	A-A logging
10	0.01	1.519	3.970
10	0.1	1.142	2.974
10	1.0	0.354	1.361
10	10	0.111	1.205
20	0.01	2.931	3.671
20	0.1	2.033	2.523
20	1.0	0.416	0.585
20	10	0.092	0.387
25	0.01	3.530	3.509
25	0.1	2.373	2.293
25	1.0	0.399	0.428
25	10	0.068	0.248

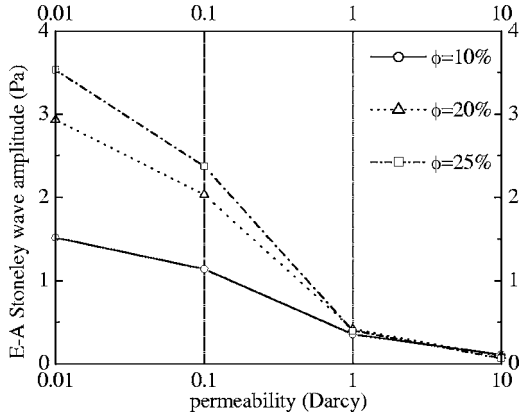


FIG. 6. Showing the amplitude variation of the E-A Stoneley wave with different porosities and permeabilities, according to the data in Table II.

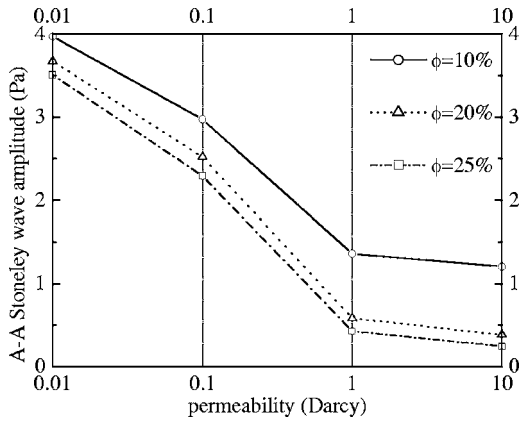


FIG. 7. Showing the amplitude variation of the A-A Stoneley wave with different porosities and permeabilities, according to the data in Table II.

creases 11.5%, 33.8%, and 42.0%, respectively, in the same situations. This reveals that E-A Stoneley wave amplitude drops more sharply in the high permeability range than that in conventional A-A logging. The A-A Stoneley wave amplitude is sensitive to porosity rather than permeability (see Fig. 7) in the high permeability range. In the paper by Wu *et al.* (1995), the error is large in estimating permeability from Stoneley wave attenuation when permeability is larger than 2 Darcy, because of the small change with permeability in the amplitude of the Stoneley wave in the high permeability range. The E-A Stoneley wave can be used instead for the inversion of permeability in highly permeable formations.

IV. CONCLUSIONS

A coupled method and an uncoupled method are put forward for simulating the axisymmetric fields of E-A logging in a fluid-saturated porous medium. A vertical electric dipole on the borehole axis is used as the exciting source. In the coupled method, the acoustic and EM fields in the formation are formulated by using Pride's governing equations. In the uncoupled method, the EM field is obtained by solving Maxwell's equations; the acoustic wave field is then obtained by solving linear inhomogeneous Biot equations, with the source term, due to the electrokinetic effect, originating from the EM disturbances. This uncoupled method is possible because the influence from the converted acoustic field on the EM field is negligible.

The transient full waveforms for the electric field and the converted acoustic field in the borehole are simulated by the two methods. In order to compare the waveforms obtained by the two methods, the difference waveforms are calculated when the same input parameters are used. It is found that the largest relative difference in amplitude is in the order of 10^{-4} for all examples in this paper, so that cannot be seen in the plots of waveforms. It proved that the uncoupled method is correct and feasible. The idea behind the uncoupled method can be extended to evaluate E-A conversion response in more complex geometric situations in which the analytical solution is unobtainable. For example, if the fluid-saturated formation outside the borehole is stratified horizontally, there will be no analytical solution to either the EM field or the acoustic field. But we can solve for the EM field first by a well established finite-element or finite-difference method, and then solve for the inhomogeneous acoustic field by the same numerical technique.

The full waveform of the converted acoustic field in E-A logging contains four wave groups: EM-accompanying wave group, compressional wave group, shear and pseudo-Rayleigh wave group, and Stoneley wave group. The EM-accompanying wave group accompanies the EM field and has an apparent phase velocity of the EM wave, but is a mechanical disturbance in itself. The other three wave groups can propagate independently of the EM field and are generated when the EM wave emitted by the electric dipole travels across the borehole wall.

The amplitude of E-A Stoneley wave changes with permeability and porosity. It decreases with permeability, similar to conventional A-A logging. It decreases with porosity when permeability is very high (several Darcies or higher), but increases with porosity in the common sediment rock permeability range. This is because the amplitude of the E-A Stoneley wave is indirectly affected by porosity in two different ways. It is affected by porosity through the electrokinetic coupling coefficient L as well as through propagation attenuation. The coefficient L increases with porosity, while propagation attenuation increases with porosity. When permeability is small, the variation of the E-A Stoneley wave amplitude is dominated by L, resulting in the increase of the E-A Stoneley wave amplitude with porosity. Nevertheless, when permeability is high enough, the attenuation mechanism dominates, and the E-A Stoneley wave amplitude decreases with increasing porosity. There is a critical permeability where the E-A Stoneley wave amplitude does not change with porosity. When permeability reaches or exceeds 1 Darcy, the sensitivity of the E-A Stoneley wave amplitude to permeability is higher than that in A-A logging.

ACKNOWLEDGMENTS

The first author would like to express his gratitude to Professor Kexie Wang at Jilin University of China for his help on analysis of related wave phenomena, and to Dr. Zhenya Zhu at Massachusetts Institute of Technology for his lectures on seismoelectric experiments. This work is supported by the National Natural Science Foundation of China (Grant No. 10272038). This work was completed while the first author was visiting the Wave Physics Laboratory at Stanford University.

APPENDIX A: EXPRESSIONS OF THE ELEMENTS OF *M* IN EQ. (50)

$$m_{11} = \eta_{bm} I_1(\eta_{bm} a)/\pi,$$

$$m_{21} = m_{31} = \rho_b \omega^2 I_0(\eta_{bm} a)/\pi,$$

$$m_{41} = m_{51} = m_{61} = m_{12} = m_{22} = m_{32} = 0,$$

$$m_{42} = m_{63} = m_{64} = m_{25} = m_{26} = 0,$$

$$m_{13} = \eta_{pf}(1 + \alpha_{pf}) K_1(\eta_{pf} a),$$

$$m_{14} = \eta_{ps}(1 + \alpha_{ps}) K_1(\eta_{ps} a),$$

$$m_{15} = ik(1 + \alpha_{sh}) K_1(\eta_{sh} a),$$

$$m_{16} = ik(1 + \alpha_{em}) K_1(\eta_{em} a),$$

$$m_{23} = -k_{pf}^2 (C + M\alpha_{pf}) K_0(\eta_{pf} a),$$

$$m_{24} = -k_{ps}^2 (C + M\alpha_{ps}) K_0(\eta_{ps} a),$$

$$m_{33} = \frac{2G\eta_{pf} K_1(\eta_{pf} a)}{a} + [(2G - H - C\alpha_{pf}) k_{pf}^2 + 2G\eta_{pf}^2] K_0(\eta_{pf} a),$$

$$m_{34} = \frac{2G\eta_{ps}K_1(\eta_{ps}a)}{a} + [(2G - H - C\alpha)k_{ps}^2 + 2G\eta_{ps}^2]K_0(\eta_{ps}a),$$

$$m_{35} = 2ikG[\eta_{sh}K_0(\eta_{sh}a) + K_1(\eta_{sh}a)/a],$$

$$m_{36} = 2ikG[\eta_{em}K_0(\eta_{em}a) + K_1(\eta_{em}a)/a],$$

$$m_{43} = -2ik\eta_{pf}K_1(\eta_{pf}a),$$

$$m_{44} = -2ik\eta_{ps}K_1(\eta_{ps}a),$$

$$m_{45} = (k^2 + \eta_{sh}^2)K_1(\eta_{sh}a),$$

$$m_{46} = (k^2 + \eta_{em}^2)K_1(\eta_{em}a),$$

$$m_{52} = \eta_{be}I_0(\eta_{be}a)/ik,$$

$$m_{53} = ik\beta_{pf}K_0(\eta_{pf}a),$$

$$m_{54} = ik\beta_{ps}K_0(\eta_{ps}a),$$

$$m_{55} = -\beta_{sh}\eta_{sh}K_0(\eta_{sh}a),$$

$$m_{56} = -\beta_{em}\eta_{em}K_0(\eta_{em}a),$$

$$m_{62} = i(\eta_{be}^2/k - k)I_1(\eta_{be}a),$$

$$m_{65} = \beta_{sh}k_{sh}^2K_1(\eta_{sh}a). \quad \text{(A1)}$$

$$m_{66} = \beta_{em}k_{em}^2K_1(\eta_{em}a). \quad \text{(A1)}$$

APPENDIX B: EXPRESSIONS OF THE LINEAR EQUATIONS FOR EM FIELD IN THE UNCOUPLED METHOD

To determine the unknown coefficients A_e in Eqs. (56)–(58) and B_e in Eqs. (32)–(34), we impose the boundary conditions at the borehole wall, i.e., continuity of the axial component of the electric field and continuity of the circumferential component of the magnetic field. These conditions lead to the following linear equations:

$$\begin{bmatrix} m_{11}m_{12} \\ m_{21}m_{22} \end{bmatrix} \cdot \begin{bmatrix} A_e \\ B_e \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}, \quad \text{(B1)}$$

where

$$m_{11} = \frac{\eta_{em}}{ik}K_0(\eta_{em}a),$$

$$m_{12} = \frac{\eta_{be}}{ik}I_0(\eta_{be}a),$$

$$m_{21} = \frac{1}{\omega\mu}\left(k - \frac{\eta_{em}^2}{k}\right)K_1(\eta_{em}a),$$

$$m_{22} = -\frac{1}{\omega\mu}\left(k - \frac{\eta_{be}^2}{k}\right)I_1(\eta_{be}a),$$

$$y_1 = \frac{\mu p_e}{4\pi^2}i\omega\left(1 - \frac{k^2}{k_{be}^2}\right)K_0(\eta_{be}a),$$

$$y_2 = \frac{p_e}{4\pi^2}\eta_{be}K_1(\eta_{be}a). \quad \text{(B2)}$$

APPENDIX C: EXPRESSIONS OF THE LINEAR EQUATIONS FOR ACOUSTIC FIELD IN THE UNCOUPLED METHOD

Here we establish the linear equations to solve the unknown coefficients in the expressions for the acoustic fields in the formation and in the borehole in the uncoupled method by employing the acoustic field boundary conditions at the borehole wall. These boundary conditions are the same as Eqs. (44)–(47). The linear equations are expressed below in the frequency-wavenumber domain:

$$\mathbf{MA} = \mathbf{B}, \quad \text{(C1)}$$

where

$$\mathbf{A} = \{A_m, A_{pf}, A_{ps}, A_{sh}\}^T, \quad \text{(C2)}$$

$$\mathbf{B} = \{b_1, b_2, b_3, b_4\}^T, \quad \text{(C3)}$$

and the elements of **M** and **B** are given by

$$m_{11} = \frac{\eta_{bm}I_1(\eta_{bm}a)}{\rho_b\omega^2},$$

$$m_{21} = m_{31} = I_0(\eta_{bm}a),$$

$$m_{12} = \eta_{pf}(1 + \alpha_{pf})K_1(\eta_{pf}a),$$

$$m_{13} = \eta_{ps}(1 + \alpha_{ps})K_1(\eta_{ps}a),$$

$$m_{14} = ik(1 + \alpha_{sh})K_1(\eta_{sh}a),$$

$$m_{41} = m_{24} = 0,$$

$$m_{22} = -k_{pf}^2(C + M\alpha_{pf})K_0(\eta_{pf}a),$$

$$m_{23} = -k_{ps}^2(C + M\alpha_{ps})K_0(\eta_{ps}a),$$

$$m_{32} = 2G\eta_{pf}\frac{K_1(\eta_{pf}a)}{a} + [2Gk^2 - (H + \alpha_{pf}C)k_{pf}^2]K_0(\eta_{pf}a),$$

$$m_{33} = 2G\eta_{ps}\frac{K_1(\eta_{ps}a)}{a} + [2Gk^2 - (H + \alpha_{ps}C)k_{ps}^2]K_0(\eta_{ps}a),$$

$$m_{34} = 2Gik\left[\eta_sK_0(\eta_{sh}a) + \frac{K_1(\eta_{sh}a)}{a}\right],$$

$$m_{42} = -2ik\eta_{pf}K_1(\eta_{pf}a),$$

$$m_{43} = -2ik\eta_{ps}K_1(\eta_{ps}a),$$

$$m_{44} = (k^2 + \eta_{sh}^2)K_1(\eta_{sh}a),$$

$$b_1 = (T_1 + T_2)A_e K_1(\eta_{em} a),$$

$$b_2 = 0,$$

$$b_3 = 2GT_1 A_e \left(\eta_{em} K_0(\eta_{em} a) + \frac{1}{a} K_1(\eta_{em} a) \right),$$

$$b_4 = \frac{k^2 + \eta_{em}^2}{ik} T_1 A_e K_1(\eta_{em} a). \quad \text{(C4)}$$

Biot, M. A. (**1956**). "Theory of propagation of elastic waves in a fluid-saturated porous solid. I. low-frequency range," J. Acoust. Soc. Am. **28**, 168–178.

Biot, M. A. (**1962**). "Mechanics of deformation and acoustic propagation in porous media," J. Appl. Phys. **33**, 1482–1498.

Broding, R., Buchanan, S., and Hearn, D. (**1963**). "Field experiments on the electroseismic effect," IEEE Trans. Geosci. Electron. **GE-1**, 23–31.

Butler, K. E., Russel, R. D., Kepic, A. W., and Maxwell, M. (**1996**). "Measurement of the seismoelectric response from a shallow boundary," Geophysics **61**, 1769–1778.

Frenkel, J. (**1944**). "On the theory of seismic and seismoelectric phenomena in a moist soil," (in Russian), J. Phys. (Moscow) **8**(4), 230–241.

Guru, B. S., and Hiziroğlu, H. R. (**2004**). *Electromagnetic Field Theory Fundamentals*, 2nd ed. (Cambridge U.P., London), p. 553.

Haartsen, M. W., and Pride, S. R. (**1997**). "Electroseismic waves from point sources in layered media," J. Geophys. Res. **102**(B11), 24745–24769.

Hu, H. (**2000**). Ph.D. thesis, Jilin University, China, (in Chinese).

Hu, H., and Liu, J. (**2002**). "Simulation of the converted electric field during acoustoelectric logging," 72nd Annual International Meeting, Soc. Exploration Geophysics, Expanded Abstracts, Vol. **21**, pp. 348–351.

Ivanov, A. G. (**1940**). "The electroseismic effect of the second kind," Bull. Acad. Sci. USSR, Geophys. Ser. **5**, 699–727 (in Russian).

Johnson, D. L., Koplik, J., and Dashen, R. (**1987**). "Theory of dynamic permeability and tortuosity in fluid-saturated porous media," J. Fluid Mech. **176**, 379–402.

Long, L. T., and Rivers, W. K. (**1975**). "Field measurement of the electroseismic response," Geophysics **40**, 233–425.

Martner, S. T., and Sparks, N. R. (**1959**). "The electroseismic effect," Geophysics **24**, 297–308.

Mikhailov, O. V., Haartsen, M. W., and Toksöz, M. N. (**1997**). "Electroseimic investigation of the shallow subsurface: Field measurements and numerical modeling," Geophysics **62**, 97–105.

Neev, J., and Yeatts, F. R. (**1989**). "Electrokinetic effects in fluid-saturated poroelastic media," Phys. Rev. B **40**, 9135–9141.

Pride, S. R., and Morgan, F. (**1991**). "Electrokinetic dissipation induced by seismic waves," Geophysics **56**, 914–925.

Pride, S. R. (**1994**). "Governing equations for the coupled electromagnetics and acoustics of porous media," Phys. Rev. B **50**, 15678–15696.

Pride, S. R., and Haartsen, M. W. (**1996**). "Electroseismic wave properties," J. Acoust. Soc. Am. **100**, 1301–1315.

Rosenbaum, J. H. (**1974**). "Synthetic microseismograms: logging in porous formations," Geophysics **39**, 14–32.

Thompson, A. H. (**2005**). "Electromagnetic-to-seismic conversion: successful developments suggest viable applications in exploration and production," 75th Annual International Meeting, Soc. Exploration Geophysics, Expanded Abstracts, Vol. **24**, pp. 554–556.

Thompson, A. H., Hornbostel, S., Burns, J., Murray, T., Raschke, R., Wride, J., McCammon, P., Sumner, J., Haake, G., Bixby, M., Ross, W., White, B., Zhou, M., and Peczak, P. (**2005**). "Field tests of electroseismic hydrocarbon detection," 75th Annual International Meeting, Soc. Exploration Geophysics, Expanded Abstracts, Vol. **24**, pp. 565–568.

Thompson, A. H., and Gist, G. A. (**1993**). "Geophysical applications of electrokinetic conversion," The Leading Edge **12**, 1169–1173.

Tsang, L., and Rader, D. (**1979**). "Numerical evaluation of the transient acoustic waveform due to a point source in a fluid-filled borehold," Geophysics **44**, 1706–1720.

Vernik, L. (**1994**). "Predicting lithology and transport properties from acoustic velocities based on petrophysical classification of siliclastics," Geophysics **59**, 420–427.

Wu, X., Wang, K., Guo, L., Yu, S., and Dong, Q. (**1995**). "Inversion of permeability from the full waveform acoustic logging data," Acta Geophys. Sin. Suppl. I **38**, 224–231 (in Chinese).

Zhu, Z., Haartsen, M. W., and Toksöz, M. N. (**1999**). "Experimental studies of electrokinetic conversions in fluid-saturated borehole models," Geophysics **64**, 1349–1356.

Zhu, Z., and Toksöz, M. N. (**2003**). "Crosshole seismoelectric measurements in borehole models," Geophysics **68**, 1519–1524.

Zhu, Z., and Toksöz, M. N. (**2005**). "Seismoelectric and seismomagnetic measurements in fractured borehole models," Geophysics **70**, F45–F51.

Spectral properties of the interference head wave

Jee Woong Choi and Peter H. Dahl
Applied Physics Laboratory, University of Washington, 1013 NE 40th Street, Seattle, Washington 98105-6698

(Received 5 December 2006; revised 17 April 2007; accepted 29 April 2007)

The interference head wave propagating through a sediment with a linear sound speed gradient is studied as a function of the parameter ζ, which is itself a function of acoustic frequency f, sediment sound speed and its gradient, and range. For $\zeta < 1$ the amplitude spectrum of the interference head waves goes as $|S(f)|/f$, where $S(f)$ is the source spectrum. For increasing ζ beyond unity a more complicated modulation of $S(f)$ ensues, which is explained by a channel transfer function $H_1(f)$, constructed analytically from a summation of terms involving zeroth-order refracted waves (referred to as a ray approach). For $\zeta \gtrsim 2$ this summation compares well with a wave theory result for the interference head wave involving a fluid-fluid boundary. The amplitude spectrum of the interference head wave in the absence of sediment attenuation is $|S(f)| \cdot |H_1(f)|$ and it is essential to know these to obtain an estimate of sediment attenuation from field observations. Examples of $|S(f)| \cdot |H_1(f)|$ are presented for which $H_1(f)$ is computed directly using the ray approach and indirectly using the parabolic wave equation. A brief discussion on the application of these results towards the inversion of sediment attenuation is given. © *2007 Acoustical Society of America.*
[DOI: 10.1121/1.2743156]

PACS number(s): 43.30.Ma, 43.30.Cq, 43.30.Pc [RAS] Pages: 146–150

I. INTRODUCTION

A precursor arrival, defined as a signal that propagates primarily through the sediment-borne path and arrives prior to any water-borne arrival, is a valuable measure of seabed properties. Some properties associated with precursor arrivals are discussed by Choi and Dahl[1] and are used in geoacoustic inversion of field measurements.[2] One such precursor arrival is the interference head wave, which represents the transition between a first-order head wave existing in an isospeed seabed and a zeroth-order refracted wave [or C_n waves $(n=0,1,\ldots)$] generated in the presence of a positive sound-speed gradient in the seabed (Fig. 1); it arises from the superposition of the C_n waves, and thus has special properties distinct from those of the first-order head wave and the refracted wave. An important characteristic of the interference head wave is the change in dominant frequency from that of the source amplitude spectrum $|S(f)|$, where f is a frequency. This property is discussed briefly by Červený and Ravindra[3] and is further studied using a parabolic equation (PE) simulation by Choi and Dahl,[1] who show, for example, that the maximum positive change in dominant frequency occurs when the nondimensional parameter ζ is ~ 2.3, where

$$\zeta = L_2(a/c_{1s})^{2/3}/\lambda_1^{1/3}, \tag{1}$$

and where c_{1s} and λ_1 are the sound speed and acoustic wavelength just below the water-sediment interface, respectively, a and L_2 are the positive sediment sound-speed gradient and the horizontal range (Fig. 1) corresponding to the first-order head wave generated within a (hypothetical) seabed of the constant sound speed c_{1s}, respectively. Furthermore, ζ can be expressed by $(32f\Delta T_{0\_1})^{1/3}$, where ΔT_{0_1} (equal to $a^2L_2^3/32c_{1s}^3$) is the arrival time difference between the C_0 wave and C_1 wave.[2]

In this paper, the dominant frequency change phenomenon of the interference head wave is investigated using a ray theory approach, and a better understanding of the relation among it and the parameter ζ and source signal bandwidth is reached. The relation is confirmed with PE simulation, and its significance in terms of estimating sediment attenuation from precursor amplitude spectra is discussed. A brief review of the ray-based approach of Červený and Ravindra for describing the interference head wave is given in Sec. II, including a comparison with a wave-based approach. Section III presents the key result of this paper—a means to understand the change in the spectrum of the interference head wave as a function of the parameter ζ. The result builds on the ray-based approach, and a simple analytic description for the channel transfer function of the interference head wave is given, with results compared to those obtained from PE simulation. A summary and short discussion on the effect of sediment attenuation, including application of these results to sediment attenuation inversion, are given in Sec. IV.

II. THE INTERFERENCE HEAD WAVES AS VIEWED BY RAY AND WAVE THEORIES

For a case in which the sediment sound speed at the water-sediment interface c_{1s} is greater than the water sound speed, and sediment sound speed increases linearly with depth [solid line in Fig. 1(a)], the precursor arrival consists of C_n waves $(n=0,1,\ldots)$. Furthermore, when the arrival time difference between the C_0 wave and C_1 wave, ΔT_{0_1}, is less than the characteristic time duration δt of the incident wave, the C_n waves interfere, producing an interference head wave.[1–3]

 0001-4966/2007/122(1)/146/5/$23.00

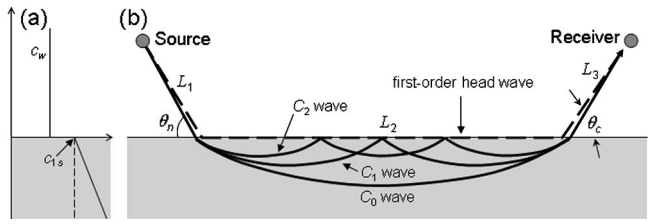


FIG. 1. (a) Sound-speed profile in seabed showing constant sound speed, zero gradient (dashed line), and nonzero positive gradient (solid line). (b) Equivalent ray path for a first-order head wave corresponding to a constant sediment sound speed (dashed line) and multiple ray paths, C_n waves ($n=0,1,\ldots$), constituting a zeroth-order refracted wave, which corresponds to a positive gradient for sound speed (solid line). Note that grazing angle θ_n varies slightly with C_n wave, and is greater than the critical angle θ_c.

Following Červený and Ravindra,[3] the interference head wave can be interpreted as a superposition of C_n waves ($n=0,1,\ldots$), which they define as a C wave,

$$P_C = \sum_{n=0}^{\infty} P_{C_n}. \tag{2}$$

An asymptotic expression for the pressure associated with each C_n wave is expressed in terms of the first-order head wave, P_H, as given by

$$P_{C_n} \sim \frac{P_H \pi \zeta^3}{(n+1)^2} \exp\left[-i(n-1)\frac{\pi}{2} + \frac{i\pi\zeta^3}{12(n+1)^2}\right], \tag{3}$$

where the first-order head wave is given by[3–5]

$$P_H = \frac{i2c_w S(f)}{kmc_1[1-(c_w/c_1)^2]r^{1/2}L_2^{3/2}} \exp[ik(L_1+L_3)+ik_1L_2]. \tag{4}$$

In the above, k and k_1, are wave numbers in the upper and lower media, respectively, c_w and c_1 are the water sound speed and iso-speed sediment sound speed (here, $c_1=c_{1s}$), respectively, m is the ratio of lower and upper medium densities, and $S(f)$ is the source spectrum of unit Pa m (or kg/s^2). The variables L_1 and L_3 are the slant ranges of the first-order head wave path from the source to water-sediment interface and from the interface to receiver, respectively (Fig. 1). Note that Eq. (4) applies only to a fluid-fluid interface, which is a reasonable assumption for unconsolidated sediment. Furthermore, for a sound-speed ratio (c_w/c_1) of less than 0.577 that is characteristic of unconsolidated sediments, then the corresponding term for a first-order head wave involving shear wave propagation in the sediment is negligible.[3] Červený and Ravindra[3] refer to the analysis giving rise to Eq. (3) as a "ray approach" because its derivation assumes a degree of smallness in the transmitted grazing angle of the C_0 wave, insofar as this angle should represent a small increase from that of the first-order head wave, which is 0°. We find this derivation specifically requires that the parameter ar/c_{1s} be less than unity, where r is the source-receiver range.[6]

From Eqs. (3) and (4), the amplitude spectrum of each C_n wave is proportional to $|S(f)|$, whereas that of the first-order head wave goes as $|S(f)|/f$. On the other hand, the amplitude of the interference head wave varies with the parameter ζ owing to the superposition of the C_n waves ($n=0,1,\ldots$), which is discussed subsequently. The amplitude for each C_n wave as predicted by Eq. (3) becomes less accurate for small ζ, as here the grazing angle approaches the critical angle and the zeroth-order term in the ray series vanishes.[3] Still, the ray approach for describing the interference head wave is very useful because of its efficiency and conceptual clarity, and Eq. (3) can be further manipulated to provide information on the spectral properties of precursor arrivals.

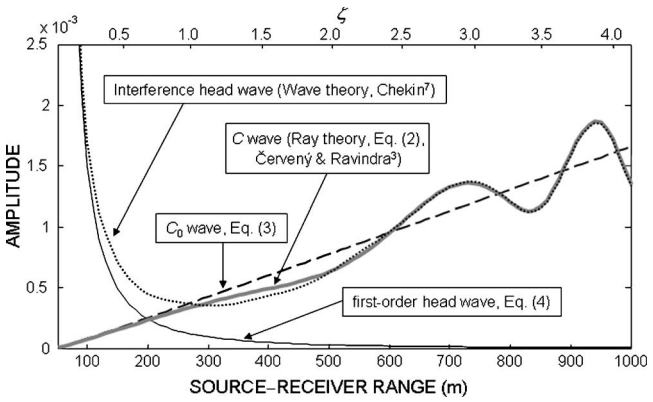


FIG. 2. Comparison of the amplitudes (absolute value) of C_0 wave, interference head wave, and first-order head wave, as a function of range (lower axis) and parameter ζ (upper axis). Note that the first-order head wave, Eq. (4), is not itself a function of parameter ζ; here, $S(f)$ is set to 1.

An approach based on wave theory was derived by Chekin,[7] although the resulting expression is cumbersome to evaluate. Figure 2 shows an approximate evaluation of Chekin's result, compared with the ray approach [Eqs. (2)–(4)]. The amplitudes of the interference head wave and the first-order head wave as a function of range and parameter ζ are computed using the following geoacoustic parameters: $c_w=1508$ m/s, $c_{1s}=1689$ m/s, water density (ρ)$=1033$ kg/m^3, sediment density (ρ_1)$=2\rho$, $a=1.6$ s^{-1}, and $f=150$ Hz. For the first-order head wave a constant sound speed in the seabed is used, which is equivalent to c_{1s}. Both source and receiver heights are set to 10 m above the bottom (water depth is not considered in this calculation). In the case of the first-order head wave, the amplitude decreases rapidly with the range going as $1/r^{1/2}L_2^{3/2}$, whereas the amplitude of the C_0 wave increases linearly with a range because it is proportional to $L_2^{3/2}/r^{1/2}$ when ar/c_{1s} is less than unity. The amplitude curve of the C wave predicted from Eq. (2) varies about the C_0-wave amplitude, computed by Eq. (3). However, for $\Delta T_{0\_1} > \delta t$, the C_0 wave becomes separated from the subsequent C_n waves ($n=1,2,\ldots$);[1] this separation occurs at higher ζ (or as range increases). Chekin's wave theory result for the interference head wave amplitude is shown by the dotted line in Fig. 2. For $\zeta<1$, the interference head wave amplitude approaches that of the first-order head wave; as ζ increases, the amplitude tends toward that of zeroth order, and this is consistent with the properties of the interference head wave (Sec. I). The C wave, however, fails to predict the amplitude going to that of the first-order head wave in the case of a small ζ. For $\zeta \geq 2$, the amplitude of the C wave is nearly identical to that of the interference head wave as predicted by the wave-theory approach.

III. THE SPECTRUM OF THE INTERFERENCE HEAD WAVE AND ITS CHANGE IN DOMINANT FREQUENCY

The spectrum of the interference head wave can undergo a significant change in dominant frequency, which is essential to understand if the spectrum is to be used for geoacoustic inversion studies. Červený and Ravindra[3] remark on this property, and show graphically a case in which the dominant frequency of the interference head wave initially increases with range, and then falls back to that of the original source spectrum. Choi and Dahl[1] investigate this property empirically using PE simulation. They found for the parameter $f_c/B \geqslant 1$, where f_c is the center frequency and B is the half-power bandwidth of the source spectrum, that a small change, Δf, in dominant frequency for the interference head wave, when expressed as the parameter $f_c \Delta f / B^2$, is well described by a single curve that is a function of ζ. A maximum change occurs at $\zeta \approx 2.3$.

In this section the frequency change phenomenon of the interference head wave is explained using the ray approach (Sec. II). It is straightforward to recast Eq. (2) as a function of frequency, which becomes

$$P_C(f) \sim \frac{i c_w^2 a^2 L_2^{3/2} S(f)}{m(1-c_w^2/c_{1s}^2) c_{1s}^4 r^{1/2}} \sum_{n=0}^{\infty} (n+1)^{-2} \times \exp\left[\frac{i\pi K f}{(n+1)^2} - \frac{i\pi(n-1)}{2} \right], \quad (5)$$

where

$$K = \frac{L_2^3 a^2}{12 c_{1s}^3}. \quad (6)$$

Note that, strictly speaking, Eq. (5) applies for $\zeta \gtrsim 2$ owing to its derivation from the ray theory approach. The key property of Eq. (5) lies in the summation term, which we define as a channel transfer function $H_1(f)$ associated with the channel composed of C_n waves ($n=0,1,\ldots$) (Fig. 1).

Figure 3 shows the modulus of $H_1(f)$ as a function of frequency and parameter ζ computed using a different number of terms in Eq. (5), and based on the same geoacoustic parameters and source-receiver geometry as those of Fig. 2 for the source-receiver range of 600 m. The contribution of the C_n waves ($n=1,2,\ldots$) to the overall sum rapidly decreases with increasing n because their amplitudes are inversely proportional to $(n+1)^2$. This is evident in the partial sums (C_0+C_1, and C_1+ up to C_4), wherein the salient features of $H_1(f)$ are effectively determined by summation of the first four terms beyond the C_0 wave. As a practical matter, $H_1(f)$ is well approximated by the first ten terms of the summation in Eq. (5).

The dominant frequency change of the interference head wave is determined by $H_1(f)$ that itself originates from the superposition of C_n waves ($n=0,1,\ldots$) as illustrated in Fig. 3. Figure 4 shows the interference head wave amplitude spectrum predicted from the ray approach (i.e., $|S(f)|$ multiplied by $|H_1(f)|$) compared with that obtained via modeling using the range-dependent acoustic model (RAM)[8] PE code. A Hanning window centered at 150 Hz with a half-power bandwidth B equal to 110 Hz ($f_c/B \approx 1.4$) is used as $|S(f)|$, which serves as the PE source amplitude spectrum. The geoacoustic parameters and source-receiver geometry are the same as those used in the previous examples, and water

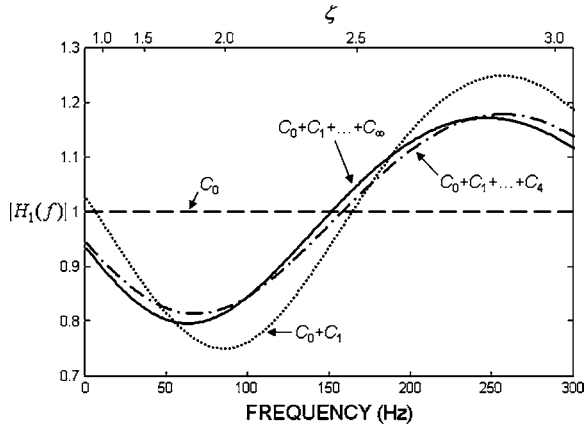


FIG. 3. The modulus of the channel transfer function, $H_1(f)$, computed using a different number of terms in the summation in Eq. (5), as a function of frequency (lower axis, linear scale) and parameter ζ (upper axis, nonlinear scale). The modulus of $H_1(f)$ for the C_0 wave alone ($n=0$) is independent of frequency and is represented by the dashed horizontal line. The addition of terms beyond $n=0$ represents the superposition of the C_0 wave and the subsequent C_n waves, and resulting modulus of $H_1(f)$ varies about that for the C_0 wave alone. (For the calculation to infinity, $n=100$ is used.)

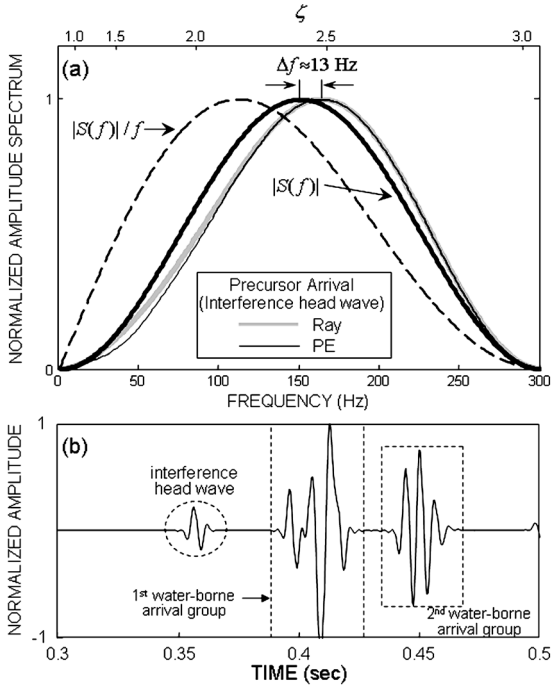


FIG. 4. (a) Comparison of the interference head wave amplitude spectrum (thick gray line) predicted from the ray theory approach with that obtained from the PE simulation (thin solid line, and falling nearly entirely within gray line). The thick solid line is amplitude spectrum of the source, and, for reference, the dashed line represents the amplitude spectrum of a first-order head wave. (b) The PE simulated time series (within circle) from which the spectrum in (a) is estimated. The two arrival groups arriving after the interference head wave are the water-borne arrivals. The first group consists of the arrivals that have one or fewer interactions with the sea surface; these include the direct, bottom, sea surface, bottom-surface, surface-bottom, and bottom-surface-bottom paths. The second group consists of the arrivals that have interacted with the sea surface twice.

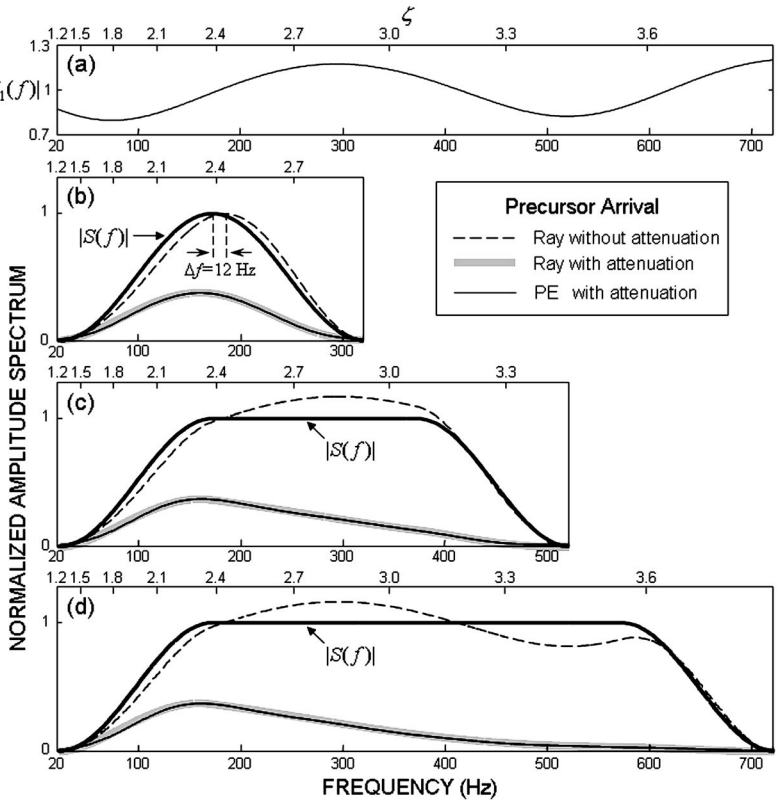


FIG. 5. (a) The modulus of the channel transfer function, $H_1(f)$, as a function of frequency (lower axis, linear scale) and parameter ζ (upper axis, nonlinear scale) for a geoacoustic model corresponding to a site in the Yellow Sea (see the text). (b) Amplitude spectrum for a precursor arrival based on a source amplitude spectrum, $|S(f)|$, parameterized by $B=110$ Hz and $f_c=175$ Hz ($f_c/B\approx1.6$). The amplitude spectrum is computed using the ray approach (with and without sediment attenuation based on $\alpha=0.08$ dB/m/kHz) and via PE simulation (with sediment attenuation). Note: the PE-derived spectrum is normalized to the maximum value of that predicted from the ray approach with sediment attenuation. (c) and (d) display the same as (b) but with source amplitude spectrum parameterized by (c) $B=310$ Hz and $f_c=275$ Hz ($f_c/B\approx0.9$), and (d) $B=510$ Hz and $f_c=375$ Hz ($f_c/B\approx0.7$).

depth is 80 m. Because the time duration δt of the incident wave (based on $1/B$) and ΔT_{0_1} are approximately 9.1 and 2.9 ms, respectively, the precursor arrival is a true interference head wave. For the case of $\Delta T_{0\_1}>\delta t$, the first arrival is the C_0 wave and its amplitude spectrum follows $|S(f)|$ because the channel transfer function of the C_0 wave (dashed line in Fig. 3) is independent of frequency. For $\zeta\gtrsim2$, the amplitude spectrum of the interference head wave predicted from the ray approach is in precise agreement with that obtained from PE simulation [for which the time series is identified in Fig. 4(b)] and change in dominant frequency is about 13 Hz.

IV. SUMMARY AND DISCUSSION

The properties of the interference head wave associated with a linear and positive sediment sound-speed gradient are governed in large part by the parameter ζ: as ζ approaches zero, the amplitude spectrum of the interference head wave becomes close to that of a first-order head wave, which is $|S(f)|/f$, where $S(f)$ is the source spectrum. As ζ increases, the amplitude spectrum goes as $|S(f)|$, with some degree of modulation due to the superposition of the C_n waves ($n=0,1,\ldots$). We show that the dominant frequency change of the interference head wave can be explained by the channel transfer function $H_1(f)$ representing the effect of the superposition of the C_n waves.

For the case in which the source spectrum is broad (e.g., $f_c/B<1$), such as with an explosive source, the idea of a change in dominant frequency no longer applies, but the modulation effect still does. In this case, the reduced time duration δt of the incident wave will likely be such that the interference head wave is replaced by noninterfering C_n waves. To illustrate both the change in dominant frequency and the modulation effect, and also introduce the effect of sediment attenuation, a geoacoustic model derived from inversion of precursor arrivals measured in the Yellow Sea[2] is used, wherein $c_{1s}=1573$ m/s, $a=1.1$ s^{-1}, $\rho_1=1.85$ g/cm^3, and the sediment attenuation coefficient $\alpha=0.08$ dB/m/kHz (Fig. 5). The source and receiver heights are 21 and 9 m, respectively, above the bottom, and the source-receiver range is 700 m, and the aforementioned condition of $ar/c_{1s}<1$ is satisfied.

Figure 5(a) shows the modulus of the channel transfer function $|H_1(f)|$ as functions of both frequency and parameter ζ for the Yellow Sea geoacoustic example. Three types of source spectra are considered here: (1) $B=110$ Hz, $f_c=175$ Hz ($f_c/B\approx1.6$); (2) $B=310$ Hz, $f_c=275$ Hz ($f_c/B\approx0.9$); and (3) $B=510$ Hz, $f_c=375$ Hz ($f_c/B\approx0.7$), which are shown in Figs. 5(b)–5(d), respectively. For the first case, $\delta t=9.1$ ms and $\Delta T_{0\_1}=2.6$ ms, and thus the precursor arrival is an interference head wave. The change in dominant frequency of approximately 12 Hz is estimated by the amplitude spectrum of the interference head wave, predicted from $|S(f)|$ multiplied by $|H_1(f)|$. For the second case, the precursor arrival is still an interference head wave because δt(3.2 ms) exceeds ΔT_{0_1}; for the third case, however, δt is about 2.0 ms, and thus the precursor arrival consists of noninterfering C_n waves. For these two cases with $f_c/B<1$ the main effect of a change in dominant frequency no longer applies as it did in the first case with $f_c/B>1$ with $\Delta f=12$ Hz. Nevertheless, the spectra are still determined by modulation with $H_1(f)$ and are in a satisfactory agreement with those predicted from PE simulation. Note that, although our examples use flat and broad spectra in Figs. 5(c) and

5(d), a spectrum more typical of an explosive source, i.e., both broad and a slightly more defined spectral peak, will be modulated by $H_1(f)$ in the same manner.

To account for attenuation in the sediment, the amplitude spectrum is divided by α', where $20\log_{10}\alpha' = \alpha\hat{L}_2 f/1000$, $\hat{L}_2$ is the path length of the C_0 wave in the sediment (approximately L_2, see Ref. 2) and $\alpha = 0.08$ dB/m/kHz. The ray and PE-derived spectra for all three cases in Fig. 5 are in very close agreement for $\zeta \gtrsim 2$, and it is difficult to see any mismatch for smaller ζ as well (even with a decibel scale). Lastly, even when the precursor arrival consists of the time-separated C_0 wave and C_n waves ($n=1,2,\ldots$), as in the third case, a true observation of the separation between a C_0 wave and the subsequent precursor arrivals is likely to be seldom achievable owing to the attenuation in the sediment. (More discussion on this point is given in Sec. V of Ref. 2.)

Details concerning the inversion of precursor arrivals measured in the Yellow Sea are in Ref. 2; however, we emphasize here that to invert for sediment attenuation using precursor arrival spectra, it is essential to account first for the spectral change owing to the modulation effect as discussed in this paper. This modulation is set by $H_1(f)$ [if the parameter $\zeta \ll 1$, then the modulation simplifies to $S(f)/f$, equivalent to that of a first-order head wave]. For example, in Ref. 2 the objective function used to invert for sediment attenuation is defined as the difference between the measured precursor arrival spectrum and the following model:

$$M \cdot |S(f)| \cdot |H_1(f)|/\alpha', \tag{7}$$

where M is an arbitrary offset. Note that in Ref. 2, $H_1(f)$ was derived from PE simulation; however, it is more convenient to derive it from the ray approach discussed in this work, with results of PE- and ray-approaches being identical as evidenced by Figs. 4 and 5.

As noted in Sec. III, Červený and Ravindra[3] and Choi and Dahl[1] related properties of the interference head wave to the parameter ζ, and analysis of this parameter was essential to the work by Dahl and Choi[2] in their study on classifying precursor arrivals in the Yellow Sea, and utilizing them in geoacoustic inversion. In view of the importance of the parameter ζ in these kinds of studies, giving it a name is appropriate. We propose calling ζ the *Chekin parameter*, to recognize the 1965 work by Chekin[7] from which the version of ζ as presented here can be derived.

ACKNOWLEDGMENT

This work was funded by the Office of Naval Research Code 321, Ocean Acoustics.

[1] J. W. Choi and P. H. Dahl, "First-order and zeroth-order head waves, their sequences, and implications for geoacoustic inversion," J. Acoust. Soc. Am. **119**, 3660–3668 (2006).

[2] P. H. Dahl and J. W. Choi, "Precursor arrivals in the Yellow Sea, their distinction from first-order head waves, and their geoacoustic inversion," J. Acoust. Soc. Am. **120**, 3525–3533 (2006).

[3] V. Červený and R. Ravindra, *Theory of Seismic Head Waves* (University of Toronto Press, Toronto, 1971), Chap. 6.

[4] G. V. Frisk, *Ocean and Seabed Acoustics: A Theory of Wave Propagation* (PTR Prentice-Hall, Englewood Cliffs, NJ, 1994), Chap. 4.

[5] L. M. Brekhovskikh, *Waves in Layered Media*, 2nd ed. (Academic, New York, 1980), Chap. 39.

[6] The criterion of $ar/c_{1s} < 1$ originates from an approximate expression for the transmitted grazing angle θ_t in the sediment, which is subsequently used to derive Eq. (3). The source-receiver range for the C_n waves is given by $r = (\mathrm{SH}+\mathrm{RH})\cot\theta_n + (2c_{1s}/a)(n+1)\tan\theta_t$, where SH and RH are source height and receiver height above the interface, respectively. We find that for $ar/c_{1s} < \min[1, \sim 3\sin\theta_c]$, the angle θ_t can be approximated as follows:

$$\theta_t \sim a\sin\{(aL_2)/[2c_{1s}(n+1)][1-(L_2a^2)/[8c_{1s}^2(n+1)^2] \times (L_2 - [\cos\theta_c(\mathrm{SH}+\mathrm{RH})]/(\sin^3\theta_c))]\},$$

which is an essential approximation to proceed to Eq. (3) involving the parameter ζ. Generally, for the case in which sediment-borne refracted C_n waves are well separated from water-borne arrivals, the criterion becomes $ar/c_{1s} < 1$.

[7] B. S. Chekin, "The effect on a head wave of small inhomogeneities in a refracting medium," Izv., Acad. Sci., USSR, Earth Phys. Ser. **3**, 143–147 (1965).

[8] M. D. Collins, "A split-step Padé solution for the parabolic equation method." J. Acoust. Soc. Am. **93**, 1736–1742 (1993).

Recreational boating traffic: A chronic source of anthropogenic noise in the Wilmington, North Carolina Intracoastal Waterway

Genevieve Haviland-Howell,[a] Adam S. Frankel,[b] Christopher M. Powell,[c] Alessandro Bocconcelli,[d] Russell L. Herman, and Laela S. Sayigh
University of North Carolina Wilmington, 601 South College Road, Wilmington, North Carolina 28401

(Received 6 June 2006; revised 29 January 2007; accepted 28 February 2007)

The majority of attention on the impact of anthropogenic noise on marine mammals has focused on low-frequency episodic activities. Persistent sources of mid-frequency noise pollution are less well studied. To address this data gap, the contribution of 25 physical, biological and anthropogenic factors to the ambient noise levels in the Wilmington, North Carolina Intracoastal Waterway were analyzed using a principal components analysis and least squares regression. The total number of recreational vessels passing through the waterway per hour is the factor that had the single greatest influence on environmental noise levels. During times of high boat traffic, anthropogenic noise is continuous rather than episodic, and occurs at frequencies that are biologically relevant to bottlenose dolphins. As a daily part of resident bottlenose dolphins' acoustic environment, recreational boating traffic may represent a chronic source of acoustic harassment. © *2007 Acoustical Society of America.* [DOI: 10.1121/1.2717766]

PACS number(s): 43.30.Nb, 43.50.Rq, 43.80.Nd, 43.28.Hr [BSF] Pages: 151–160

I. INTRODUCTION

As concern grows in both the scientific community and the public arena regarding the impact of anthropogenic noise on marine mammals, attention has focused on activities such as military sonar exercises (Frantzis, 1998; Tyack, 1998; Miller *et al.*, 2000; United States Department of the Navy, 1998), oil drilling, dredging and icebreaking activities (Richardson *et al.*, 1990, 1995; Erbe and Farmer, 1998) and acoustic thermometry (Munk *et al.*, 1994; Au *et al.*, 1997; Frankel and Clark, 1998, 2000, 2002). Such events generally occur at low frequencies over relatively short time periods, but can produce signals of high intensity and thus have a seemingly great potential to cause physiological damage and/or behavioral disruptions to marine mammals.

Although the U.S. Navy's peacetime use of low-frequency active sonar was heavily restricted after the mass Bahamian *Ziphius* strandings of March, 2000, it was mid-frequency sonar that was actually implicated in that event (Evans and England, 2001). In May 2003, researchers videotaped the reactions of resident *Orcinus orca* in Juan De Fuca and Haro Straits while simultaneously recording the *USS Shoup*, a Navy ship conducting mid-frequency ($\approx$3 kHz) sonar exercises. The whales stopped feeding and gathered close together at the surface of the water for the duration of the exercise (Balcomb, 2004). An abnormally high number of harbor porpoises were later stranded, and although necropsies performed by NMFS did not find evidence of acoustically mediated damage, over 70% of specimens were in moderate to advanced stages of decomposition (NOAA/NMFS Final Report, 2004).

There is a clear need to understand the extent to which marine mammals may be affected by mid-frequency noise. Sound intensity, duration and frequency are all critical parameters of noise and different combinations of these parameters result in noise occurring at "biologically relevant" levels to various species; levels that have the potential to interfere with behaviors essential to life and reproduction (National Research Council, 2000). Loud sounds and/or sounds lasting over long time periods can only be relevant if they are occurring in a frequency range that the receiver can perceive. This report presents the results of a study designed to quantify the total environmental noise of a section of the Wilmington, NC Intracoastal Waterway (ICW), and to identify the dominant components of that noise. This area of the ICW supports a relatively closed population of bottlenose dolphins sighted year round (Koster, 2002); thus, particular attention was given to noise within the range of bottlenose dolphin hearing.

II. MATERIALS AND METHODS

A. Data collection

Audio data were collected continuously from 21 Jun 01–9 Sep 01 with a calibrated HTI 94-SSQ hydrophone (with a sensitivity of −170.1 dB re 1 V/μPa) mounted approximately 1 m off the bottom of the ICW and 1 m off the University of North Carolina Wilmington's Center for Marine Science (CMS) pier. The pier extends into the ICW, which at this location is roughly 100 m wide, with an approximately 30-m-wide channel ranging in depth from

[a]Author to whom correspondence should be addressed; electronic mail: ghowell@seabird.com; current affiliation: Sea-Bird Electronics, Inc., 136th Place NE, Bellevue, WA 98005.
[b]Current affiliation: Marine Acoustics, Inc., 706 Giddings Avenue, Suite 1C, Annapolis, MD 21401.
[c]Current affiliation: Old Dominion University, Center for Coastal Physical Oceanography, Norfolk, VA 23529.
[d]Current affiliation: Woods Hole Oceanographic Institute, Department of Applied Physics and Ocean Engineering, Woods Hole, MA 02543.

TABLE I. Potential factors influencing received sound pressure levels in the Wilmington ICW (P=physical, A=anthropogenic, T=temporal).

Factor	Type	Description
Mean hourly wind speed	P	Meters / second
Total hourly precipitation	P	Millimeters / hour
Tidal state	P	% of time per hour high tide
Total number of boats per hour	A	Number passing per hour
Small outboard motorboat	A	Length ≤6 m; outboard engine clearly visible at back of vessel
Small inboard motorboat	A	Length ≤6 m; flat aft panel of boat clearly visible
Small motorboat, engine unknown	A	Length ≤6 m; no engine visible at back, no clear view of aft panel
Medium outboard motorboat	A	Length 6–12 m; outboard engine clearly visible at back of vessel
Medium inboard motorboat	A	Length 6–12 m; flat aft panel of boat clearly visible
Medium motorboat, engine unknown	A	Length 6–12 m; no engine visible at back, no clear view of aft panel
Medium twin engine outboard boat	A	Length 6–12 m; two distinct engines visible at back of vessel
Large inboard vessel	A	Length ≥12 m; flat aft panel of boat clearly visible
Medium sailboat, under sail	A	Length ≤9 m with at least one sail hoisted
Medium sailboat, under motor	A	Length ≤9 m with at no sails hoisted
Large sailboat, under sail	A	Length >9 m with at least one sail hoisted
Large sailboat, under motor	A	Length >9 m with no sail hoisted
Canoes/kayaks	A	Small vessels under manual power
Industrial	A	Tugboats, barges, dredge ships
Jet skis	A	Personal watercraft
Unidentifiable	A	Any vessel that could not be placed into any of the above categories
R/V Cape Fear	A	Percentage of time per hour present
Seawater pumping	A	Percentage of time per hour on
Hour of the day	T	0600–2000; 0100–0300
Day of the week	T	1–8; 8=holidays
Day of the year	T	1–365

3 to 7 m (depending on tidal state). The hydrophone output was amplified by a Shure FP11 preamplifier, low-pass filtered at 30 kHz using a Frequency Devices filter (model 900-C9L8B), and then digitized at a sampling rate of 75 kHz by the field computer, a Dell Optiplex desktop equipped with a 12 bit National Instruments PCI-MIO 16E-4 analog-to-digital converter and a SCB-68 shielded input/output connector block. The field computer, preamplifier, filter, and analog-to-digital shielded input/output connector block were secured in a weatherproof container at the end of the Center for Marine Science pier. A data acquisition program on the field computer recorded audio data continuously and saved the wave forms as 30 s AIFF files on the field computer. Data were then transmitted to the CMS lab by fiber optic cable, processed and archived. Mean received sound pressure levels (RLs) were calculated for each second of raw wave form data in each of 13 variable width frequency bands. Hourly mean RLs were calculated and the hourly means were used for further analysis.

Potential factors affecting environmental noise in this area were recorded throughout the study period (Table I). Physical noise sources included wind, precipitation and tidal state. The North Carolina National Estuarine Research Reserve provided mean hourly wind speed and precipitation data which were collected from a site on Masonboro Island, approximately 2 km north of the Center for Marine Science (Fig. 1). High tide was considered to be ±3 h of the high tide time at Masonboro Inlet, a channel located approximately 4 km north of the study site. Tidal state was recorded as a percentage of time per hour during which it was high tide.

The number of vessels passing through the study area was measured from video data recorded from 0630–2030. Recordings were made with a surveillance camera mounted in a weatherproof housing to a light fixture at the end of the pier. Time stamps in the video and audio filenames were used to synchronize data files.

Video data of the ICW were annotated by recording the total number of boats passing per hour and then classifying those boats into one of 17 categories based on size and engine type. Small vessels were boats less than approximately 6 m, medium vessels ranged from 6 to 12 m and large vessels were longer than 12 m. Engine types included inboard, outboard and "engine unknown" (when the size of a boat could be determined, but not the engine type). Temporal factors were also tracked in order to determine if time of day,

FIG. 1. (Color online) Map of study area (image from Google Earth, ©2007 Europa Technologies, Image ©2007 New Hanover County, NC).

day of the week, or day of the year had any effect on mean received sound pressure levels; these factors were likely to be highly correlated with boat traffic.

Also noted was the percentage of time per hour seawater pumps were on, which were located in close proximity to the hydrophone. These pumps transferred seawater from the ICW to the CMS building. The presence or absence of the R/V Cape Fear, a 21 m research vessel that docked at the CMS pier, was also noted in terms of the percentage of time each hour this vessel was present. This vessel's proximity to the hydrophone meant it could have added to overall noise levels when its engines were turned on or decreased levels when its engines were turned off by shielding the hydrophone from noise sources offshore of the vessel.

Finally, while no independent confirmation of the presence of fish, shrimp, or other biological sound sources was possible, nighttime (0100–0300) audio data were considered to be representative of base line biological noise in this area. This time period was chosen to maximize the likelihood that no boats would be passing and no anthropogenic factors other than the presence of the R/V Cape Fear would be affecting noise levels. Each nighttime hour was randomly spot checked for boat noise before being added to the data set.

B. Statistical procedures

Outliers were first removed from the data set; these were defined as those hours in which the mean RLs recorded were outside four standard deviations of the overall mean. This resulted in the removal of 47 h of data, eight of which were greater than the mean and 39 of which were less than the mean. In each case when the RL recorded was higher than four standard deviations of the mean, unusual biological noise was responsible; a single fish or group of fish was vocalizing very near the hydrophone during those instances. A specific cause could not be determined for RLs lower than four standard deviations of the mean, despite spot checking those hours for equipment failure. Neither the extremely high nor low sound levels were considered representative samples of the population, however, and were therefore discarded (Zar, 1999).

Mean RLs were calculated from pressure values (μPa) and were then converted to decibels for statistical testing. In order to determine if mean broadband (0–37.5 kHz) RLs were higher during the day than at night, a Kruskal-Wallis test was used because of differences in sample sizes between day and night (Zar, 1999).

To determine if daytime mean RLs were higher than nighttime mean RLs within each of the 13 frequency bands, a multivariate analysis of variance (MANOVA) was used. With a large daytime sample size, this test was appropriate because it is a single test that reduces Type I error (Zar, 1999). Although Bonferroni corrections can also be employed to reduce Type I error associated with multiple testing of the same data set, Type II error can increase unchecked

when variables are correlated (Perneger, 1998). Therefore a single parametric test, which is robust to deviations from normality (Zar, 1999), was used for this particular data set. Differences between frequency bands were determined using *post hoc* ANOVA tests.

For the remainder of the analyses, daytime and nighttime hours were tested separately. To determine if mean RLs differed between individual hours of the day or individual hours of the night, ANOVA testing followed by *post hoc* Tukey-Kramer tests were used. Finally, the effect that day of the week had on daytime and nighttime mean RLs was analyzed using an ANOVA test, followed by *post hoc* Tukey-Kramer tests when necessary.

Once the characteristics of mean received sound pressure level in this area had been described, analyses were constructed to determine which factors contributed most highly to those patterns. Twenty-five factors that could potentially contribute to ambient were examined. This large set of interrelated variables was linearly transformed into a set of 25 statistically uncorrelated linear combinations of variables, or principal components (PCs). The PCs were ranked according to the amount of variation each explained in the original data set. Those explaining approximately 85% of the variability in the original data set (in this case, the first 15) were used for further testing.

A least squares regression model using these 15 PCs as independent variables and daytime mean RLs the dependent variable was run to determine which PCs predict mean RLs. The PCs that were significantly related to mean RLs were then combined, weighted by how much of the overall model each one explained. These weights were determined from the regression model; the sum of the squares for each principal component was divided by the sum of the squares for the whole model. Each eigenvector of each of the significant principal components was multiplied by its corresponding percentage. Thus, the first principal component's eigenvectors were multiplied by the largest percentage, since the first principal component explained the most variation. This process allowed each weighted principal component to have eigenvectors that could be compared to each other directly. Finally, the weighted eigenvectors from each principal component with a significant relationship to RL were summed for each of the 25 factors.

This process yielded a single vector, termed the Noise Index, which summarized the combined results of these tests. The principal components created, not the individual factors that potentially affected ambient noise, were statistically compared directly to mean RLs. The Noise Index helped to determine which factors were the strongest contributors to mean RLs, which were then subjected to additional tests.

Descriptive statistics were used to summarize the total number of boats passing on any given day, and the total number of boats passing during each hour of any given day. Analysis of variance was used to determine if the total number of boats recorded per hour (log transformed) varied significantly with respect to hour of the day and/or day of the week. *Post hoc* Tukey-Kramer testing identified the hours and/or days that had significantly more or less boat traffic than others.

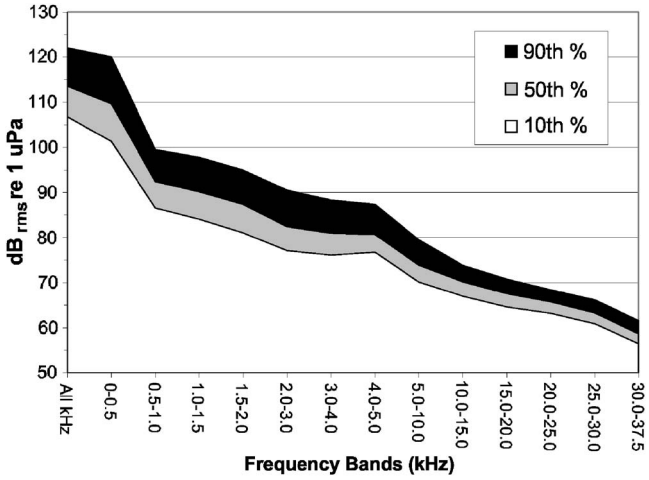


FIG. 2. Environmental daytime noise level profile for the Wilmington, North Carolina ICW (June–September 2001).

Additionally, hourly totals of environmental noise levels, wind speed and boat traffic were compared with ANOVAs during times when fishing tournaments were being held in Wilmington waters to those times when no fishing tournaments were taking place.

The spectral characteristics and received sound pressure levels of some vessel types were examined by generating composite power spectra [fast Fourier transform (FFT) =4096, no overlap, Hanning window]. These spectra were preferentially generated from times when only a single boat was passing the hydrophone, the R/V Cape Fear was absent, seawater pumps were off, the mean hourly wind speed was <3 m/s, and it was high tide (as previously defined). These were the times that had the highest signal-to-noise ratio, but these criteria were not met for all vessel types. For each vessel type that did meet these criteria the spectral values from each sample ($n=2-7$) were averaged to form that vessel type's composite spectrum.

To address the potential impact of recreational boating traffic on bottlenose dolphins, data from photoidentification survey efforts at the University of North Carolina at Wilmington were analyzed. Chi-square tests were used to compare the observed and expected total number of dolphins in the ICW on weekends versus weekdays for the time period of the present study. Expected values were determined by dividing the number of weekday/weekend days surveyed by the total number of days surveyed and then multiplying this value by the total number of dolphins seen. Thus, if dolphins were observed in the ICW equally on weekends and weekdays, it was expected that the number of dolphins seen would closely mirror the amount of survey effort on weekends versus weekdays.

III. RESULTS

A. Daytime environmental noise profile

Quantile daytime RLs averaged across the entire data set were plotted to form a profile of daytime noise levels over the time period of this study (Fig. 2). When averaged hourly across all summed frequency bands, the mean RL for the

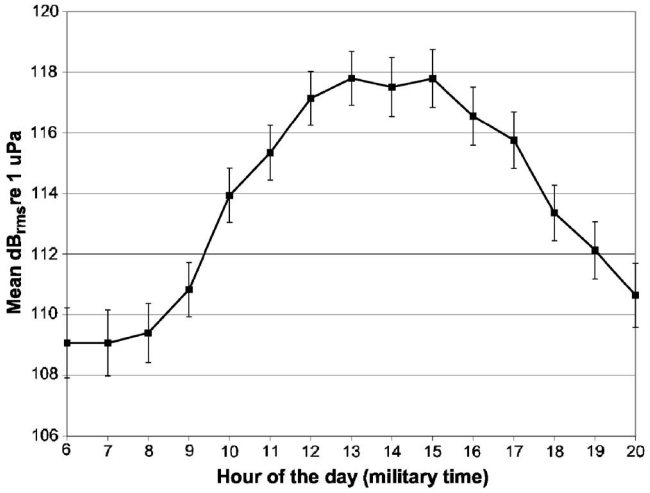


FIG. 3. Mean received sound pressure levels by hour of the day; standard error bars are shown for each data point.

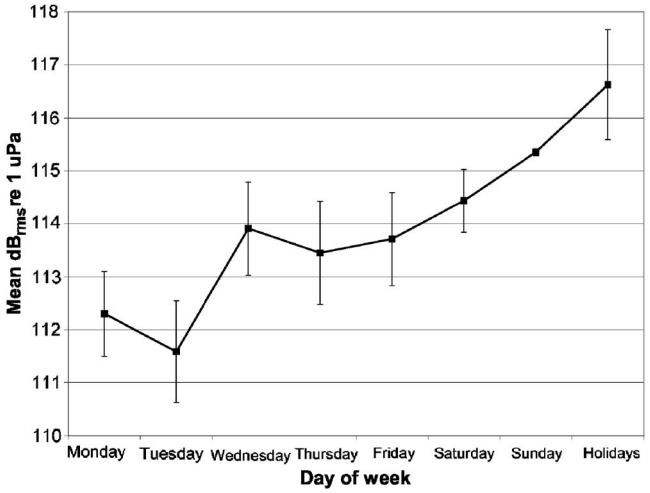


FIG. 4. Mean received sound pressure level by day of the week; standard error bars are shown for each data point.

study area was 116 dB_{rms} re 1 μPa (SE=87.5 dB). Ninety percent of RL values were between 107 and 122 dB_{rms} re 1 μPa and the maximum value for any given hour was 127 dB_{rms} re 1 μPa.

The mean value of 116 dB_{rms} re 1 μPa was exceeded during 129 h, or 34.4% of the total time surveyed. Of that time, mean RLs were between 120 and 127 dB_{rms} re 1 μPa during a total of 63 h, or 16.8% of the total survey time. Hourly mean RLs exceeded 116 dB_{rms} re 1 μPa for some portion of 27 out of 30 days surveyed, and occurred during two or more consecutive hours on 19 of those 27 days. Ten consecutive hours of hourly RLs exceeding 116 dB_{rms} re 1 μPa were recorded on 1 Jul 01, which was the maximum for any 14 h survey day.

B. Temporal characteristics of the acoustic environment in the ICW

Daytime (0630–2030) mean RLs were higher than nighttime (0100–0300) mean RLs when averaged over 0–37.5 kHz ($\chi^2=101.90$, $df=1$, $p<0.0001$), and within each frequency band except the two highest: 25.0–30.0 kHz and 30.0–37.5 kHz. This relationship was further confirmed by parametric MANOVA testing ($F=18.21$, $df=12$, error $df=404$, $p<0.0001$), which showed significantly higher noise levels during daytime hours (mean=115.9 dB_{rms} re 1 μPa, SE±87.5, $n=375$ h) than nighttime hours (mean =102.0 dB_{rms} re 1 μPa, SE±80.8, $n=42$ h).

Nighttime mean RLs did not vary significantly with hour of the day or day of the week, however, daytime mean RLs varied with both. In general, total environmental noise increased steadily through the day, peaking around 1300–1500 h and then falling off in the evening hours ($F=11.46$, $df=14$, error $df=360$, $p<0.0001$, Fig. 3). Further, total environmental noise tended to increase from relatively low values during the beginning of the week to significantly higher values on the weekends and the holidays of 4 Jul 01 and 3 Sep 01 ($F=3.20$, $df=7$, error $df=367$, $p=0.0026$, Fig. 4).

C. Major sources of environmental noise

The relative effects of anthropogenic, temporal and physical sound sources on environmental noise levels in the ICW at this location were determined using a principal components analysis, as previously described. The first 15 principal components accounted for 86.8% of the variability. The overall regression model was statistically significant ($r^2=0.5313$, $F=27.13$, $df=15$, error $df=359$, $p<0.0001$). Nine principal components were significantly related to mean RLs at $p<0.01$, however principal components 1, 12 and 13 were found to predict mean received sound pressure levels at $p<0.0001$ and thus were used to calculate the Noise Index (Table II).

The resulting Noise Index, with factors influencing noise on the x axis and summed, weighted eigenvectors on the y axis, is plotted in Fig. 5. Total number of boats recorded per hour was the factor with the highest magnitude, meaning this factor was the largest contributor to the three principal components most significantly related to daytime mean RLs. Of the 11 factors with the highest magnitudes, all but one were anthropogenic noise specific to recreational boating traffic; the exception was mean hourly wind speed.

Despite the complex interactions occurring between factors contributing to overall noise levels, linear fit regressions revealed that the log of total boats observed per hour was significantly related to RLs averaged across 0–37.5 kHz ($r^2=0.37$, $F=216.62$, $df=1$, error $df=371$, $p<0.0001$), as was mean hourly wind speed ($r^2=0.35$, $F=221.23$, $df=1$, error $df=415$, $p<0.0001$).

A total of 12,562 boats passed the CMS pier during 36 14 h days during June–September 2001, with the maximum for any given day being 847 (4 Jul 01). Boat traffic varied significantly with hour of the day ($F=17.66$, $df=14$, error $df=358$, $p<0.0001$). Relatively few boats were observed during the early morning hours; boat activity generally increased throughout the day, peaked between 1400 and 1600 h and declined in the late afternoon (Fig. 6). Boat traffic also varied significantly with day of the week, with significantly more boat traffic on Saturdays, Sundays and holi-

TABLE II. Generating the Noise Index (all values rounded): (A) Regression results and weighting calculations; (B) eigenvector chart (prime values indicate weighted principal components).

(A)

	SS	*F* Ratio	*p*	% of model
Whole model	4.17E+13	27.1273	<0.0001	100.00%
Prin. Comp. 1	2.94E+13	286.6537	<0.0001	70.45%
Prin. Comp. 12	1.83E+12	17.8163	<0.0001	4.38%
Prin. Comp. 13	3.31E+12	32.2919	<0.0001	7.94%

(B)

Factors	1	1′	12	12′	13	13′	Summed
Hour of the day	0.0879	0.0619	0.3003	0.0131	−0.2852	−0.0226	0.0525
Day of the year	0.0092	0.0065	−0.0411	−0.0018	0.5640	0.0448	0.0495
Boat totals per hour	0.3866	0.2723	−0.0667	−0.0029	0.0022	0.0002	0.2696
Small outboard motorboat	0.3254	0.2292	−0.0145	−0.0006	−0.0781	−0.0062	0.2224
Small inboard motorboat	0.2847	0.2006	0.0545	0.0024	−0.0258	−0.0020	0.2009
Small motorboat, eng. unk.	0.2748	0.1936	−0.1135	−0.0050	0.0527	0.0042	0.1928
Medium outboard motorboat	0.3035	0.2138	−0.1066	−0.0047	−0.0082	−0.0007	0.2085
Medium inboard motorboat	0.3083	0.2172	0.0290	0.0013	−0.0252	−0.0020	0.2165
Medium motorboat, eng. unk.	0.2618	0.1844	−0.1453	−0.0064	0.0679	0.0054	0.1835
Large inboard motorboat	0.2550	0.1797	0.0529	0.0023	0.1336	0.0106	0.1926
Medium twin outboard	0.2540	0.1789	−0.1808	−0.0079	−0.0013	−0.0001	0.1709
Medium sailboat (motor)	0.0560	0.0395	0.1640	0.0072	0.3430	0.0272	0.0739
Medium sailboat (sail)	0.1365	0.0961	−0.1034	−0.0045	−0.2259	−0.0179	0.0737
Large sailboat (motor)	0.0600	0.0422	0.0947	0.0041	−0.1980	−0.0157	0.0307
Large sailboat (sail)	0.0802	0.0565	0.1354	0.0059	−0.1402	−0.0111	0.0513
Canoes /kayaks	0.0439	0.0309	0.4644	0.0203	−0.0177	−0.0014	0.0499
Industrial	0.0383	0.0270	0.1127	0.0049	0.1209	0.0096	0.0415
Jet skis	0.3128	0.2204	−0.0943	−0.0041	−0.0261	−0.0021	0.2142
Unidentifiable	0.1420	0.1000	−0.1402	−0.0061	0.0601	0.0048	0.0987
% hour seawater pumps on	−0.0033	−0.0023	−0.0090	−0.0004	−0.1052	−0.0084	−0.0111
% hour Cape Fear present	0.0628	0.0442	0.0738	0.0032	0.4670	0.0371	0.0845
Day of the week	0.0567	0.0399	−0.0604	−0.0026	−0.2496	−0.0198	0.0175
% hour±3 h high tide	−0.0615	−0.0433	−0.2180	−0.0095	0.1075	0.0085	−0.0443
Total hourly precipitation	−0.0335	−0.0236	−0.2890	−0.0127	0.0687	0.0055	−0.0308
Mean hourly wind speed	0.1722	0.1213	0.5972	0.0261	0.1157	0.0092	0.1566

days (4 Jul 01, and 3 Sep 01) than on other days of the week ($F=17.09$, $df=7$, error $df=365$, $p<0.0001$, Fig. 7).

In the Wilmington area during the summer months, wind tends to peak in velocity during the afternoon hours and fall off at night ($F=11.26$, $df=17$, error $df=399$, $p<0.0001$). Mean hourly wind speeds also varied significantly with day of the week ($F=7.93$, $df=7$, error $df=409$, $p<0.0001$) with holidays having significantly more wind than Mondays, Tuesdays or Thursdays. Sundays also had significantly more wind than Tuesdays.

Two fishing tournaments were held in Wilmington waters over three days used in this study: 30 June 01, 1 July 01, and 14 July 01. Mean RLs were significantly higher on fishing tournament days ($F=6.67$, $df=1$, error $df=373$, $p=0.0102$). The log of total boats recorded per hour was also significantly higher on days with fishing tournaments ($F=18.59$, $df=1$, error $df=371$, $p<0.0001$) while mean hourly wind speed was not ($F=0.10$, $df=1$, error $df=373$, $p=0.7520$).

D. Power spectra for boat types

Composite power spectra (FFT=4096, no overlap, Hanning window) for six boat types were used to calculate the maximum dB_{rms} re 1 μPa received sound levels for those vessel types. These values were plotted along with the percentage of total traffic comprised by each vessel type (Fig. 8). Small outboard motorboats have the highest maximum received dB_{rms} re 1 μPa levels and comprise the highest percentage traffic in this area.

E. Bottlenose dolphin use of the ICW

A total of 113 dolphins were sighted in the ICW between Carolina Beach Inlet and Masonboro Inlet from June through September of 2001 (the time period of the current study). There was a significant difference between the number of dolphins observed and expected on weekends versus weekdays ($\chi^2=6.28$, $df=1$, $p=0.0122$, $n=85$, Fig. 9). Fewer dolphins were observed in the ICW than expected on weekends while more dolphins were observed than expected during weekdays.

IV. CONCLUSION

A. Factors affecting environmental noise

Total number of boats, and nine other boat-related factors were identified as the most important contributors to

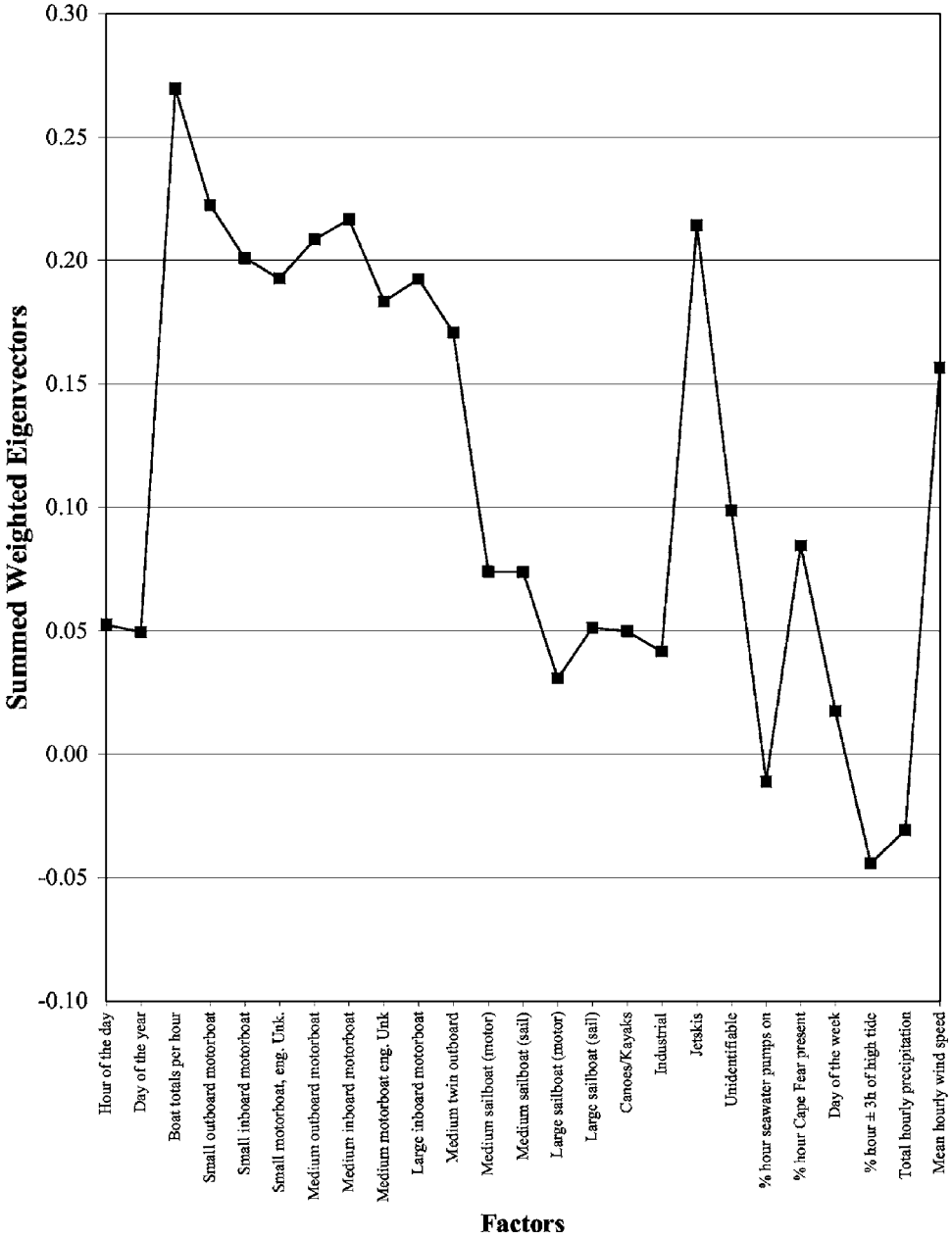


FIG. 5. Noise Index plot.

mean RLs. Daytime mean RLs measured on an hourly basis closely matched the hourly boat total trend, which peaked between 1200 and 1600 h. A combined principal components analysis and linear regression confirmed this strong relationship; total number of boats observed per hour had a significant relationship to received sound pressure levels averaged per hour, despite the complex interactions between number of boats and other potential factors influencing sound. The log of total number of boats observed per hour accounted for 36.9% of the overall variability in received sound levels.

The most important natural source of noise was mean hourly wind speed; however, principal components analysis ranked total number of boats and several boat categories as more influential than mean hourly wind speed. Mean RLs were nearly 2.5 dB higher on fishing tournament days, meaning those days had sound levels 70% higher than days without fishing tournaments. During fishing tournaments, boat traffic was significantly higher than on days without fishing tournaments, but mean hourly wind speed was not. Further, linear regressions showed that the amount of boat traffic explained slightly more variability in daytime mean RLs (36.9%) than did mean hourly wind speed (34.8%). Therefore, data from the present study suggest that anthropogenic noise from recreational boating traffic is the greatest contributor to environmental noise levels in this area.

B. Recreational boating traffic in the ICW

Boat traffic was quantified continuously on a daily basis over a period of several months to provide an accurate representation of this anthropogenic noise source in an environment utilized by bottlenose dolphins. The Atlantic Intracoastal Waterway is an important component of the habitat of bottlenose dolphins in the Wilmington area (Koster, 2002).

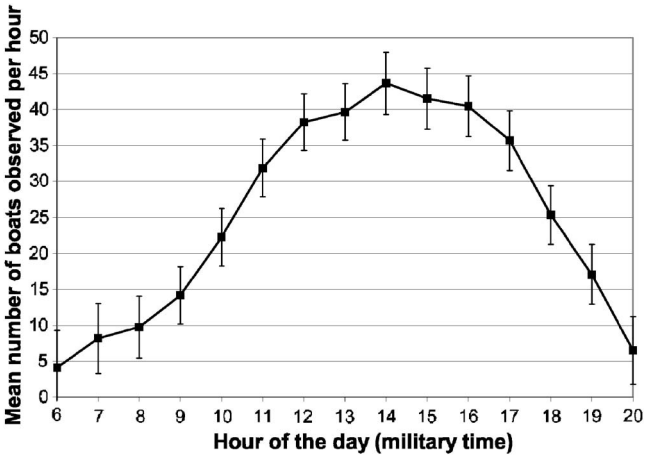


FIG. 6. Mean number of boats observed averaged per daytime hour; standard error bars are shown for each data point.

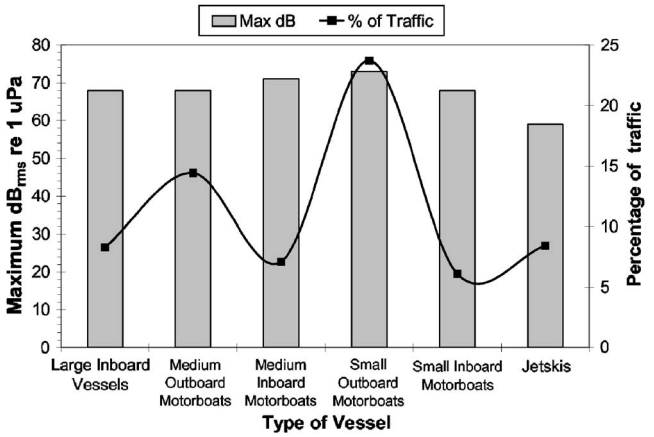


FIG. 8. Percent occurrence of various vessel types and maximum received sound levels of each type.

The amount of vessel traffic seen during this study was unexpected. The maximum amount of vessel traffic for any given day (847 boats) was recorded on 4 Jul 01, however, this number is a conservative estimate of the true number of vessels out that day. All daytime recordings stopped at 2030, but waterfront fireworks were scheduled to begin at 2100; thus, it is likely that a great deal more traffic was actually present than was recorded.

Large amounts of vessel traffic also occurred on days other than the 4th of July holiday. The maximum number of boats recorded per hour (118) occurred on 14 Jul 2001. In fact, hourly means exceeded 60 boats (one boat per minute) on 56 different occasions, often over consecutive hours. At this frequency, boat traffic essentially stops being an incidental occurrence in the ICW and becomes a continuous presence.

Although some traffic could be attributed to seasonal visitors transiting craft through the area, the predominance of small outboard motorboats suggests the ICW is being utilized most often by area residents or weekend visitors to Wilmington. Not surprisingly, boat traffic was heavier on weekends than on weekdays and peaked in the mid-to-late afternoon hours each day. Boat traffic in the ICW was also higher when fishing tournaments were taking place.

Van Parijs and Corkeron (2001) found that acoustic communication in Pacific humpback dolphins (*Sousa chinensis*) was significantly affected by boat traffic averaging only one boat per hour. With boat traffic averaging 36 boats per hour during weekends in the Wilmington ICW, it seems likely that vessel traffic may affect bottlenose dolphin behavior and acoustic communication in this area as well. Moreover, since these recordings were made at the edge of the ICW, the values reported here may be lower than dolphins may actually hear, since dolphins could be in closer proximity to the boats. Given that the ICW is only about 30 m wide in this area, it is highly likely that dolphins were often closer to boats than to the hydrophone.

C. Potential for harassment of bottlenose dolphins

Daytime sound levels were higher than nighttime, or natural ambient noise levels, and above dolphin auditory threshold levels from 1 to 25 kHz (Fig. 10). Of particular concern are the levels recorded during the daytime hours between 5 and 25 kHz, where the primary energy of social

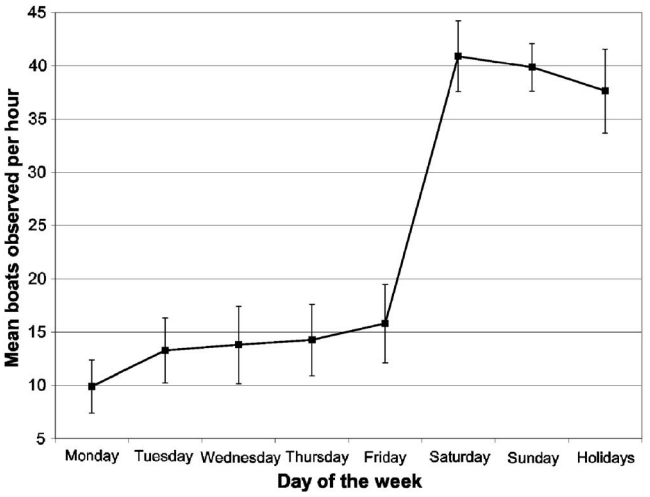


FIG. 7. Mean number of hourly boat totals observed, averaged per day of the week; standard error bars are shown for each data point.

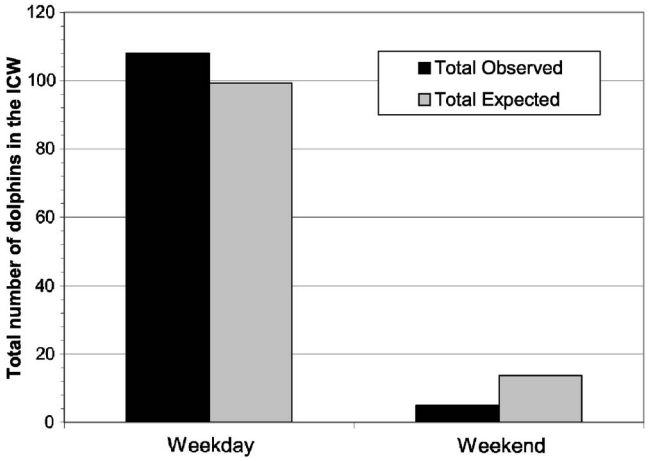


FIG. 9. Total number of dolphins observed and expected on weekdays versus weekends (based on survey effort) in the Wilmington ICW between Masonboro Inlet and Carolina Beach Inlet from June 2001 to September 2001 ($\chi^2=6.28$, $p=0.0122$).

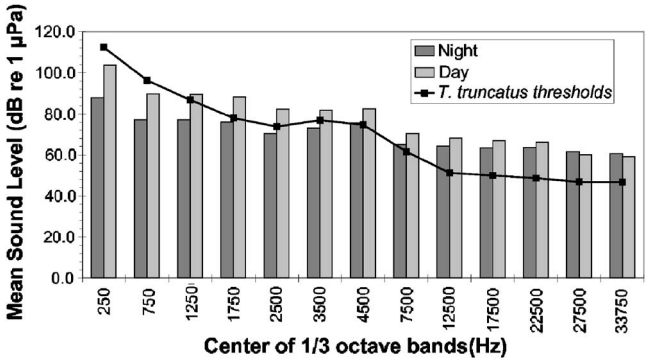


FIG. 10. Mean RLs for daytime and nighttime, 0–37.5 kHz, and *Tursiops truncatus* auditory threshold levels (extracted from Johnson (1967) as presented in Richardson *et al.* (1995).

whistles occurs. Anthropogenic noise is thus occurring at frequencies that are biologically relevant to bottlenose dolphins and may have the potential to cause harassment as it has been broadly defined under the Marine Mammal Protection Act.

Additionally, mean hourly RLs exceeded 116 dB_{rms} re 1 μPa nearly every day surveyed, indicating bottlenose dolphins in the ICW near Wilmington, North Carolina could be at risk for noise exposure on a daily basis. High mean RLs were often recorded over consecutive hours, making high sound levels the rule in this area during the summer, not the exception.

Thus, bottlenose dolphins utilizing the ICW during times of heavy boat traffic may have a more difficult time communicating than when the ICW is relatively free of anthropogenic noise inputs. Additionally, anthropogenic noise could pose an impediment to feeding by obscuring fish vocalizations. Gut content analysis indicates that Atlantic bottlenose dolphins feed mainly on soniferous fishes (Barros and Wells, 1998) and a recent study by Gannon *et al.* (2005) confirmed that dolphins use passive listening during foraging. Noise from recreational boats could make the passive localization of prey more difficult, as our observations from this study indicate the primary energy of most fish vocalizations falls below 1 kHz.

Preliminary data suggest that dolphins may be avoiding the ICW during times of increased boat traffic. After adjusting for survey effort, bottlenose dolphins were seen less often than expected on weekends (times of heavy boat traffic) than on weekdays throughout the study period. However, it is not known if boat traffic was responsible for this change in dolphin distribution.

D. Future research

The present study showed that an area regularly utilized by bottlenose dolphins was subject to large amounts of boat traffic on a nearly continuous basis during the summer months. This raises questions regarding both the potential short and long-term impacts of such traffic. Future research is needed to determine if bottlenose dolphins are exhibiting avoidance behavior patterns in response to boat traffic, and if so, whether or not that those behaviors are influenced by the accompanying increased noise levels. In order to investigate matters of this nature, there is a need to adequately define what may constitute acoustic harassment of marine mammals, in terms of frequency ranges, sound levels, and length of exposure to sounds of varying intensity.

ACKNOWLEDGMENTS

We would like to thank the Information Systems and Operations Management Department of the University of North Carolina Wilmington for providing the grant which made this work possible. Additionally, we thank Carole Yoder from CMS for providing pumping times and Steven Ross and Tara Nye from the North Carolina National Estuarine Research Reserve for providing wind speed and precipitation data.

Au, W. W. L., Nachtigall, P. E., and Pawloski, J. L. (**1997**). "Acoustic effects of the ATOC signal (75 Hz, 95 dB) on dolphins and whales," J. Acoust. Soc. Am. **101**, 2973–2977.

Balcomb, K. (**2004**). Personal communication. Manager and Data Analysts at North Carolina Wildlife Resources Commission, Raleigh, NC 27602.

Barros, N. B., and Wells, R. S. (**1998**). "Prey and feeding patterns of resident bottlenose dolphins (*Tursiops truncatus*) in Sarasota Bay, Florida," J. Mammal. **79**(3), 1045–1059.

Erbe, C., and Farmer, D. M. (**1998**). "Masked hearing thresholds of a beluga whale (*Delphinapterus leucas*) in icebreaker noise," Deep-Sea Res., Part II **45**, 1373–1388.

Evans, D. L., and England, G. R. (eds.) (**2001**). Joint interim report on Bahamas marine mammal stranding event of 14–16 March 2000. U.S. Department of Commerce (NOAA)/U.S. Navy, 1–61.

Frankel, A. A., and Clark, C. W. (**2002**). "ATOC and other factors affecting distribution and abundance of humpback whales (*Megaptera novaeangliae*) off the north shore of Kauai," Marine Mammal Sci. **18**, 644–662.

Frankel, A. S., and Clark, C. W. (**2000**). "Behavioral responses of humpback whales (*Megaptera novaeangliae*) to full-scale ATOC signals," J. Acoust. Soc. Am. **108**, 1930–1937.

Frankel, A. S., and Clark, C. W. (**1998**). "Results of low-frequency playback of M-sequence noise to humpback whales, *Megaptera novaeangliae*, in Hawaii," Can. J. Zool. **76**, 521–535.

Frantzis, A. (**1998**). "Does acoustic testing strand whales?," Nature (London) **392**, 29.

Gannon, D. P., Barros, N. B., Nowacek, D. P., Read, A. J., Wapels, 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.

Koster, D. (**2002**). "Residency and association patterns of bottlenose dolphins near Wilmington, North Carolina," M.S. Thesis, University of North Carolina, Wilmington, NC 28403.

Miller, P. J. O., Biassoni, N., Samuels, A., and Tyack, P. L. (**2000**). "Whale songs lengthen in response to sonar," Nature (London) **405**, 903.

Munk, W. H., Spindel, R. C., Baggeroer, A., and Birdsall, T. G. (**1994**). "The Heard Island feasibility test," J. Acoust. Soc. Am. **96**, 2330–2342.

National Oceanic and Atmospheric Administration and National Marine Fisheries Services (**2004**). Final report multi-disciplinary investigation of harbor porpoises (*Phocoena phocoena*) stranded in Washington State from 2 May—2 June 2003 coinciding with the mid-range sonar exercises of the *U.S.S. Shoup*.

National Research Council. (**2000**). *Marine Mammals and Low-Frequency Sound: Progress Since 1994* (National Academy Press, Washington, D.C.). p. 67.

Perneger, T. V. (**1998**). "What's wrong with Bonferroni adjustments." Br. Med. J. (Clin Res. Ed) **316**, 1236–1238.

Richardson, W. J., Greene, Jr., C. R., Malme, C. I., and Thomson, D. H. (**1995**). *Marine Mammals and Noise* (Academic, San Diego).

Richardson, W. J., Würsig, B., and Greene, Jr., C. R. (**1990**). "Reactions of bowhead whales, *Balaena mysticetus*, to drilling and dredging noise in the Canadian Beaufort Sea," Mar. Environ. Res. **29**, 135–160.

Tyack, P. L. (**1998**). "The LFA sonar experiments and humpback whales in Hawaii," Whalewatcher, Fall/Winter, 3–11.

United States Department of the Navy. (**1998**). *Final Environmental Impact*

Statement: Shock Testing the Seawolf Submarine, (Department of the Navy—lead agency, National Marine Fisheries Service (cooperating agency), pp. Appendix E1–E30.
Van Parijs, S. M., and Corkeron, P. J. (**2001**). "Boat traffic affects the acoustic behavior of Pacific humpback dolphins, *Sousa chinensis*," J. Mar. Biol. Assoc. U.K. **81**, 553–538.
Zar, J. H. (**1999**) *Biostatistical Analysis*, 4th ed. (Prentice–Hall, Upper Saddle River, NJ), pp. 82, 86, 185.

Uncertainty estimation in seismo-acoustic reflection travel time inversion

Jan Dettmer[a] and Stan E. Dosso
School of Earth and Ocean Sciences, University of Victoria, Victoria, British Columbia V8W 3P6, Canada

Charles W. Holland
The Pennsylvania State University, Applied Research Laboratory, State College, Pennsylvania 16804-0030

(Received 7 February 2007; accepted 12 April 2007)

This paper develops a nonlinear Bayesian inversion for high-resolution seabed reflection travel time data including rigorous uncertainty estimation and examination of statistical assumptions. Travel time data are picked on seismo-acoustic traces and inverted for a layered sediment sound-velocity model. Particular attention is paid to picking errors which are often biased, correlated, and nonstationary. Non-Toeplitz data covariance matrices are estimated and included in the inversion along with unknown travel time offset (bias) parameters to account for these errors. Simulated experiments show that neglecting error covariances and biases can cause misleading inversion results with unrealistically high confidence. The inversion samples the posterior probability density and provides a solution in terms of one- and two-dimensional marginal probability densities, correlations, and credibility intervals. Statistical assumptions are examined through the data residuals with rigorous statistical tests. The method is applied to shallow-water data collected on the Malta Plateau during the SCARAB98 experiment.
 [DOI: 10.1121/1.2736514]

PACS number(s): 43.30.Pc, 43.60.Pt [AIT] Pages: 161–176

I. INTRODUCTION

The shallow-water sound field is generally dominated by bottom-interacting paths, and is therefore influenced strongly by seabed geoacoustic properties. Hence, the spatial variability of seabed properties at fine scales is of key interest to understand shallow-water sound propagation (lack of knowledge of physical seabed parameters often represents the limiting factor in ocean applications[1]). However, direct sampling of vertical and lateral variability via coring can be prohibitively laborious and expensive. Geoacoustic inversion, which infers seabed properties from acoustic measurements in the water column, represents a promising approach.

Holland and Osler[2] developed an experimental technique to collect high-resolution seabed reflectivity data in shallow water using a bottom-moored hydrophone and ship-towed impulsive source, with a seafloor experimental footprint of ~100 m. Because of this local scale, the effects of spatial and temporal variability in the water column and seabed are greatly reduced compared to long-range acoustic measurements (such as matched field inversion). The experiment yields seismo-acoustic data that can be evaluated in the time domain (reflection travel times) and/or in the frequency domain (reflection coefficients). Holland and Osler[2] used Bryan's method[3,4] to estimate layer sound velocities and thicknesses from the seismo-acoustic data. This paper develops a Bayesian ray-tracing inversion of reflection travel-time data picked from multiple seabed layers (reflectors) to construct a model of the sediment sound-velocity profile, with an emphasis on rigorous quantitative uncertainty estimation.

Bayesian inversion formulates an inverse problem in terms of the posterior probability density (PPD) of the model parameters,[5–7] incorporating both data and prior information. The solution is quantified in terms of properties of the PPD representing parameter estimates, parameter uncertainties (e.g., marginal distributions, credibility intervals), and parameter inter-relationships (correlations) which can be computed using numerical methods for nonlinear inverse problems.[8–11]

Quantitative uncertainty estimation is of key importance for meaningful geoacoustic inversion results. This requires not only a nonlinear inversion approach, but also rigorous estimation of the data error statistics. In particular, ignoring data error correlations (covariance) and biases can lead to underestimation of geoacoustic uncertainties and biased, misleading parameter estimates.[12] The approach applied here is based on nonparametric estimation of data error covariance matrices from residual analysis, including effects of non-stationary errors, and on rigorous *a posteriori* statistical tests to examine these estimates and the underlying assumptions.[11–14] Biased data errors are accounted for by explicitly including unknown travel time offset parameters for each reflector in the inversion. The inversion is applied to acoustic data collected during the SCARAB98 experiment on the Malta Plateau (Site 2), Mediterranean Sea. The upper ~20 m of the sediment are considered in the inversion.

The remainder of this paper is organized as follows. Section II considers the Bayesian formulation of the inverse problem. Data misfit functions and methods for treating cor-

[a]Electronic mail: jand@uvic.ca

related and nonstationary errors are developed and a scheme for examining the statistical assumptions is described. Section III develops the Bayesian travel-time inversion and conducts a detailed synthetic study to examine and account for typical errors in picking arrival times including correlated, nonstationary, and biased errors. Section IV applies the travel time inversion to the Malta Plateau data. Section V summarizes and discusses the results.

II. INVERSION TECHNIQUE

A. Bayesian inversion

The following provides a brief overview of the Bayesian formulation used here for the geoacoustic inversion; more complete treatments of Bayesian theory can be found elsewhere.[5,15] Bayes' rule can be written

$$P(\mathbf{m}|\mathbf{d}) = \frac{\mathcal{L}(\mathbf{d}|\mathbf{m})P(\mathbf{m})}{P(\mathbf{d})}, \tag{1}$$

where $\mathbf{m} \in \mathbb{R}^M$ and $\mathbf{d} \in \mathbb{R}^N$ are random variables that represent the model parameters and data, respectively. $P(\mathbf{m}|\mathbf{d})$ is the PPD, $\mathcal{L}(\mathbf{d}|\mathbf{m})$ is the likelihood function, $P(\mathbf{m})$ is the model prior distribution, and $P(\mathbf{d})$ is the data prior (a constant factor once the data are measured). The likelihood function can generally be expressed as $\mathcal{L}(\mathbf{m}) \propto \exp(-E(\mathbf{m}))$ where $E(\mathbf{m})$ is an appropriate data error function (considered later). Equation (1) then becomes

$$P(\mathbf{m}|\mathbf{d}) = \frac{\exp(-\phi(\mathbf{m}))}{\int_{\mathcal{M}} \exp(-\phi(\mathbf{m}'))d\mathbf{m}'}, \tag{2}$$

where the integration is over the model space $\mathcal{M} \subset \mathbb{R}^M$ and $\phi(\mathbf{m}) = E(\mathbf{m}) - \log_e P(\mathbf{m})$ is the generalized misfit. The PPD represents the full solution to the inverse problem in the Bayesian formulation. However, due to the PPD's multidimensional nature, interpretation is nontrivial and properties such as the maximum *a posteriori* (MAP) estimate $\hat{\mathbf{m}}$, the mean model $\bar{\mathbf{m}}$, the model covariance matrix $\mathbf{C}^{(\mathbf{m})}$, and marginal probability distributions $P(m_i|\mathbf{d})$, must be calculated to provide parameter estimates, uncertainties, and interrelationships:

$$\hat{\mathbf{m}} = \mathrm{Arg}_{\max} P(\mathbf{m}|\mathbf{d}), \tag{3}$$

$$\bar{\mathbf{m}} = \int_{\mathcal{M}} \mathbf{m}' P(\mathbf{m}'|\mathbf{d})d\mathbf{m}', \tag{4}$$

$$\mathbf{C}^{(\mathbf{m})} = \int_{\mathcal{M}} (\mathbf{m}' - \bar{\mathbf{m}})(\mathbf{m}' - \bar{\mathbf{m}})^T P(\mathbf{m}'|\mathbf{d})d\mathbf{m}', \tag{5}$$

$$P(m_i|\mathbf{d}) = \int_{\mathcal{M}} \delta(m_i' - m_i) P(\mathbf{m}'|\mathbf{d})d\mathbf{m}', \tag{6}$$

where δ denotes the Dirac delta function. Higher-dimensional marginal distributions can be defined similar to Eq. (6). Uncertainties of parameter estimates can also be quantified in terms of highest probability density (HPD) credibility intervals. The β% HPD interval is defined as the interval of minimum width that contains β% of the area of the marginal probability distribution. Interrelations of model parameters can be quantified by the model correlation matrix $R_{ij} = C_{ij}/(C_{ii}C_{jj})^{1/2}$. While analytic solutions to Eq. (3)–(6) exist for linear inverse problems, solutions for nonlinear problems must be found numerically.

MAP estimates, Eq. (3), can be found by numerical minimization of $\phi(\mathbf{m})$, such as adaptive simplex simulated annealing, an efficient hybrid optimization algorithm that combines the local downhill-simplex method with a very fast simulated annealing global search.[8] The integrals in Eqs. (5) and (6) are solved by applying the Markov-chain Monte Carlo method of Metropolis Gibbs sampling to sample $\phi(\mathbf{m})$.[9,16,17]

B. Likelihood function and data uncertainties

The following defines data misfit functions (as part of the generalized misfit) based on the likelihood function. Further, a framework to rigorously treat uncorrelated and correlated data errors is developed, including data covariance matrix estimates and quantitative statistical tests to validate statistical assumptions *a posteriori*.

Formulating the likelihood function $\mathcal{L}(\mathbf{m})$ requires specifying the data uncertainty distribution, including both measurement errors and theory errors. In general, the above-outlined Bayesian inversion can be applied with arbitrary uncertainty distributions. In practice, however, the lack of specific knowledge of the error distribution often suggests a mathematically simple distribution be assumed. In particular, Gaussian distributions are commonly considered, with their statistical parameters estimated from the data.

For N observed data $\mathbf{d}_i$ at each of L reflectors with unbiased Gaussian-distributed random errors, the likelihood function is given by

$$\mathcal{L}(\mathbf{m}) \propto \prod_{i=1}^{L} \exp\left(-\frac{1}{2}(\mathbf{d}_i - \mathbf{d}_i(\mathbf{m}))^{\mathsf{T}}(\mathbf{C}_i^{(\mathbf{d})})^{-1}(\mathbf{d}_i - \mathbf{d}_i(\mathbf{m}))\right), \tag{7}$$

where $\mathbf{d}_i(\mathbf{m})$ are the replica data computed for model $\mathbf{m}$ and $\mathbf{C}_i^{(\mathbf{d})}$ are the L $N \times N$ data covariance matrices. The data misfit function is given by the negative log likelihood

$$E(\mathbf{m}) = \sum_{i=1}^{L} \frac{1}{2}(\mathbf{d}_i - \mathbf{d}_i(\mathbf{m}))^{\mathsf{T}}(\mathbf{C}_i^{(\mathbf{d})})^{-1}(\mathbf{d}_i - \mathbf{d}_i(\mathbf{m})). \tag{8}$$

In many cases, the covariance matrices $\mathbf{C}_i^{(\mathbf{d})}$ are not known. Depending on the data error distribution, different approaches can be applied to estimate the covariance matrices. If the data do not have significantly correlated errors, the covariance matrices can be approximated as diagonal, $\mathbf{C}_i^{(\mathbf{d})} = \sigma_i^2\mathbf{I}$, where $\mathbf{I}$ is the identity matrix. The likelihood function then becomes

$$\mathcal{L}(\mathbf{m}) \propto \prod_{i=1}^{L} \exp\left(-\sum_{j=1}^{N} \frac{(d_{ij} - d_{ij}(\mathbf{m}))^2}{2\sigma_i^2} \right), \tag{9}$$

where d_{ij} represents the jth datum for the ith reflector. A maximum likelihood (ML) estimate for the σ_i can be found by maximizing $\mathcal{L}(\mathbf{m})$ over σ_i and $\mathbf{m}$ to yield

$$\tilde{\sigma}_i = \left(\frac{1}{N} \sum_{j=1}^{N} (d_{ij} - d_{ij}(\tilde{\mathbf{m}}))^2 \right)^{1/2}, \tag{10}$$

where the ML estimate $\tilde{\mathbf{m}}$ is found by minimizing the misfit obtained by substituting Eq. (10) into Eq. (9) to yield[18]

$$E(\mathbf{m}) = \sum_{i=1}^{L} \left(\frac{N}{2} \log_e \left[\sum_{j=1}^{N} (d_{ij} - d_{ij}(\mathbf{m}))^2 \right] \right). \tag{11}$$

To date, the simplifying assumption of uncorrelated data errors is usually made in geoacoustic inversion. In practice, however, strong error correlations are often present due to theory errors.[11,12] It is important to determine whether or not significant error correlations are present in Bayesian inversion, since neglecting these can result in underestimating geoacoustic uncertainties. One approach to treat correlated errors is to downsample the data below the correlation length of the errors, with an obvious loss of information. Another approach, applied in this paper, is to analyze data residuals and incorporate error correlations into the inversion without loss of data information. For this approach, a data covariance matrix is estimated from the data residuals using a nonparametric approach (which does not assume a specific correlation form) and is applied in the likelihood function of the Bayesian inversion.

The covariance matrices $\mathbf{C}_i^{(\mathbf{d})}$ can be estimated as follows. Initially assuming uncorrelated errors of unknown standard deviation, Eq. (11) is minimized to yield an estimate $\tilde{\mathbf{m}}^{(1)}$. The data covariance matrix for each reflector can then be estimated from the autocovariance of the data residuals $\mathbf{r}_i^{(1)} = \mathbf{d}_i - \mathbf{d}_i(\tilde{\mathbf{m}}^{(1)})$.[11] The jth element of the autocovariance function is given by

$$c_{ij} = \frac{1}{N} \sum_{k=0}^{N-j-1} (r_{ij+k} - \bar{r}_i)(r_{ik} - \bar{r}_i), \tag{12}$$

where $\bar{r}_i$ is the mean of the residuals. Every term c_{ij} is then used to fill the jth diagonal of the initial data covariance matrix estimate $\tilde{\mathbf{C}}_{(i,1)}^{(\mathbf{d})}$ producing a Toeplitz or banded matrix. This covariance matrix can then be applied in a subsequent inversion for a MAP estimate $\tilde{\mathbf{m}}^{(2)}$ by minimizing Eq. (8). If the covariance matrix computed for $\tilde{\mathbf{m}}^{(2)}$ using Eq. (12) is sufficiently close to $\hat{\mathbf{C}}_{(i,1)}^{(\mathbf{d})}$, the procedure can be considered to have converged. If not, further iterations can be performed.

The assumption of stationary data errors is not always justified. In particular, for travel time data the arrival-time pick uncertainties (standard deviation) often vary slowly along a reflector (illustrated in Sec. III). This nonstationarity can be taken into account when estimating the data covariance matrix as follows. The standard deviation as a function of range for each reflector is calculated as a running rms average over Q data

$$\sigma_j = \sqrt{\frac{1}{Q} \sum_{k=j-Q/2}^{j+Q/2} r_k^2}. \tag{13}$$

These standard deviations are then used to scale the data residuals so that

$$n_j = \frac{r_j}{\sigma_j}. \tag{14}$$

The n_i are then used in place of r_i to compute Toeplitz data covariance matrices $\tilde{\mathbf{C}}_i^{(\mathbf{d})}$ according to Eq. (12). To account for the nonstationary residuals, $\tilde{\mathbf{C}}_i^{(\mathbf{d})}$ is scaled to yield the final non-Toeplitz data covariance matrix

$$C_{jl}^{(\mathbf{d})} = \tilde{C}_{jl} \sigma_j \sigma_l, \tag{15}$$

that include nonstationary effects.

C. Statistical tests

Formulating the likelihood function for Bayesian inversion requires making assumptions about the data error statistics (e.g., Gaussian distributed with estimated covariance). To provide confidence in the inversion results, the underlying statistical assumptions should be examined *a posteriori*.[11,12] Since the data residuals can be considered a realization of the data errors, statistical tests can be applied to the residuals. The effect of taking covariances into account in the inversion can be examined by testing standardized and raw residuals. Standardized residuals are scaled according to $\tilde{\mathbf{r}}_i = \mathbf{L}_i^{-1} \mathbf{r}_i$, where $\mathbf{L}_i$ is the lower triangular matrix of the Cholesky decomposition (matrix square root)

$$\mathbf{C}_i^{(\mathbf{d})} = \mathbf{L}_i \mathbf{L}_i^{\mathsf{T}}. \tag{16}$$

Raw residuals are scaled only by their standard deviation. Tests can be applied to examine both the form of the residual distribution (e.g., Gaussian) and the randomness of the residuals. The form of the distribution can be qualitatively examined by considering histograms of data residuals. A quantitative examination can be carried out using the Kolmogorov-Smirnov (KS) test.[19,20] Depending on the sample size, tables with critical values for different levels of significance can be calculated. Lilliefors[21] used a Monte Carlo calculation to estimate the critical values for the case of estimating mean and covariance from the sample and for different sample sizes. For this paper, the method was adapted by calculating critical values for fixed mean and variance estimated from the sample. The Monte Carlo calculation was performed for the appropriate sample size and for 5000 realizations of Gaussian distributed samples. Randomness of the data residuals can be examined using the runs test.[20] For both tests, a level of significance of $p = 0.05$ is generally accepted as little or no evidence against the initial assumption of randomly distributed Gaussian data errors with covariance $\mathbf{C}^{(\mathbf{d})}$.

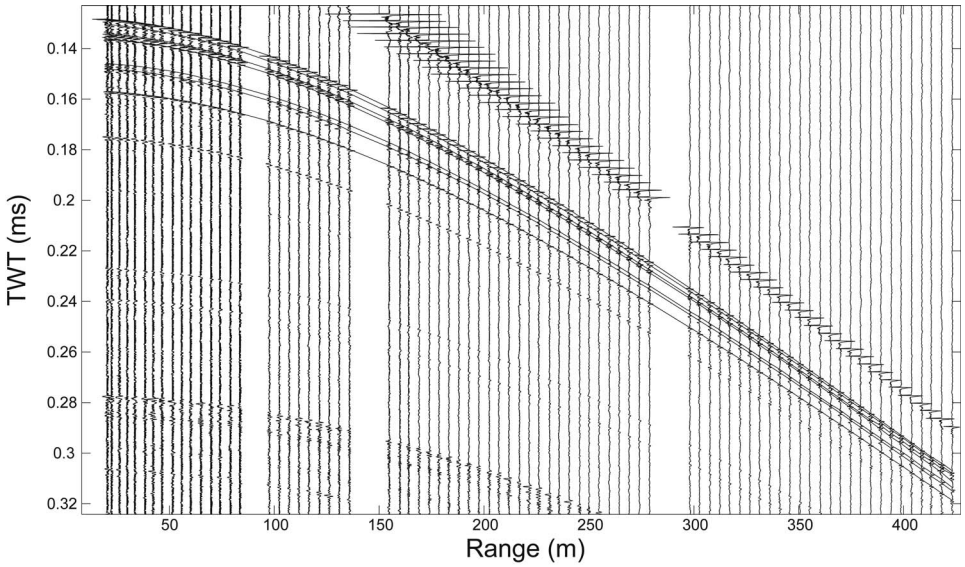


FIG. 1. Acoustic traces (in two-way travel time, TWT) and picked travel times for six sediment layers in the Malta Plateau data. Missing traces are due to interruptions when data are written to disk.

III. TRAVEL TIME INVERSION

A. Picking travel times

The experiment developed by Holland and Osler[2] involves a single hydrophone mounted to the seafloor and a ship towed impulsive source that transmits at fixed range increments. The experiment is described more fully in Sec. IV. This results in seismo-acoustic travel time data as shown in Fig. 1. The data include a number of clearly defined seafloor and subbottom reflections, which suggest a high information content in the time domain. This section develops a fully nonlinear travel time inversion that is based on picked travel times for multiple reflectors. The forward model uses ray tracing in horizontally stratified homogeneous layers, resulting in straight rays within layers with refraction and reflection at layer boundaries. The ray tracing algorithm used here determines eigenrays that connect source and receiver efficiently using the method of images and Newton's method.[22]

The inversion is based on travel time data which are determined by picking the first-break of the reflected arrivals on recorded seismo-acoustic traces, such as those in Fig. 1. The picking is based on a matched filter

$$t_{\text{pick}} = \text{Arg}_{\max}(A_{xy}), \tag{17}$$

where A_{xy} is the cross-correlation function of vectors x and y. Vector x consists of the acoustic trace for a certain range and vector y represents the fixed-length wavelet that is matched. Initially, y is taken to be the arrival wavelet starting at a hand picked point on the first trace. The wavelet is automatically updated while proceeding through the traces (i.e., the wavelet beginning at the arrival time picked for trace i is used to correlate with trace $i+1$). The matched filter picks are then edited by hand to ensure stable and consistent results. The picking can be improved iteratively by comparing replica data for a MAP inversion model to the seismic traces and refining the picks in areas where the replica seems to fit better than the initial pick. Since the goal of this inversion is to quantify information content, the focus here is on treating the errors in picking rigorously. In particular, subjective criteria, such as picking arrivals along perfect hyperbolae are avoided. This assures that the uncertainty estimates are meaningful and actually relate to the true travel times (not an interpreted travel time model with unquantifiable theory error).

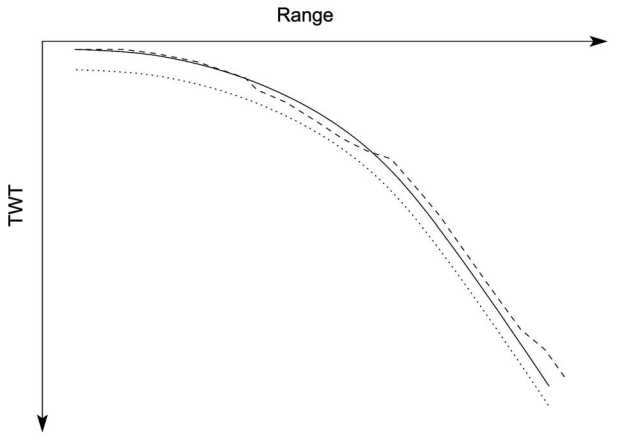


FIG. 2. Illustration of constant bias and correlated error in picking. The dotted line represents the true first break and the solid line a feature that appears to be the first break. The dashed line illustrates the type of picks that would likely be made by an interpreter when correlating features across traces. The picks show biased, strongly correlated errors.

There are several challenges involved with picking travel times. First, the actual start of the signal is not always obvious due to overlap of different arrivals, particularly at long ranges and due to complicated seabed structure where wave forms of different events are superimposed. In general, it is easier to correlate events from trace to trace and consistently pick the same signal feature across range than it is to definitively pick the first break. This often results in a bias (offset) of the arrival time across many traces. Since even small errors in picking can have a significant effect on the recovered model, strong biases in the PPD can result from these picking errors. To address this in the inversion, addi-

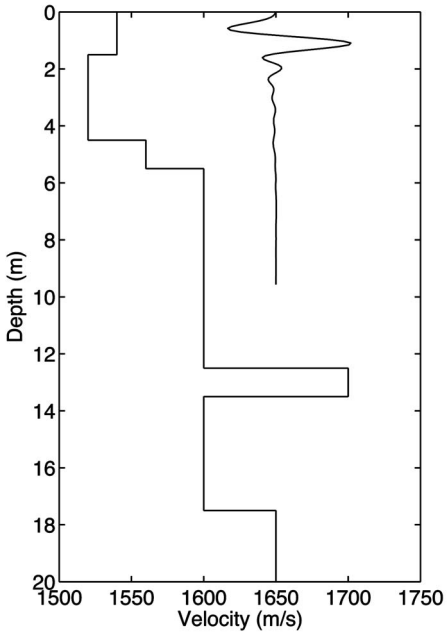


FIG. 3. Sound speed profile with six layers over a half-space that is considered for the synthetic data. Also shown is the source pulse converted to depth for a sound velocity of 1511 m/s.

tional nuisance parameters can be included to account for an unknown picking bias for each reflector to obtain a generalization of Eq. (11),

$$E(\mathbf{m}) = \sum_{i=1}^{L} \frac{N}{2} \log_e \left(\sum_{j=1}^{N} (d_{ij} + a_i - d_{ij}(\mathbf{m}))^2 \right), \tag{18}$$

where a_i is the unknown offset (bias) parameter for reflector i.

The picking of travel times typically becomes more difficult and less accurate with increasing range, since amplitudes decrease and the arrivals for different interfaces occur closer together increasing overlap. Hence, it is possible to consistently pick travel times for a number of traces and then lose a particular feature and begin picking on another, similar, feature, as illustrated in Fig. 2. The result is strongly correlated data errors, which can be taken into account by estimating covariance matrices from the data residuals for each reflector. Further, as arrivals run into each other and amplitude decreases at large ranges, it becomes more difficult to identify the correct feature, resulting in increased picking uncertainty with range. Note that the severity of these issues differs for different interfaces, and depends on the layer thickness. The thicker a layer is, the less likely the corresponding reflected arrivals are to overlap with other arrivals. To address the above-presented issues, nonstationary data covariance matrix estimates can be calculated according to Sec. II B.

Accounting for both the covariance matrix estimate and the picking offset nuisance parameters, the data misfit can then be written as

$$E(\mathbf{m}) = \sum_{i=1}^{L} \frac{1}{2} (\mathbf{d}_i + a_i - \mathbf{d}_i(\hat{\mathbf{m}}))^{\mathsf{T}} (\mathbf{C}_i^{(\mathbf{d})})^{-1} (\mathbf{d}_i + a_i - \mathbf{d}_i(\hat{\mathbf{m}})), \tag{19}$$

where L is the number of reflectors, and $\mathbf{C}_i^{(\mathbf{d})}$ are estimated data covariance matrices.

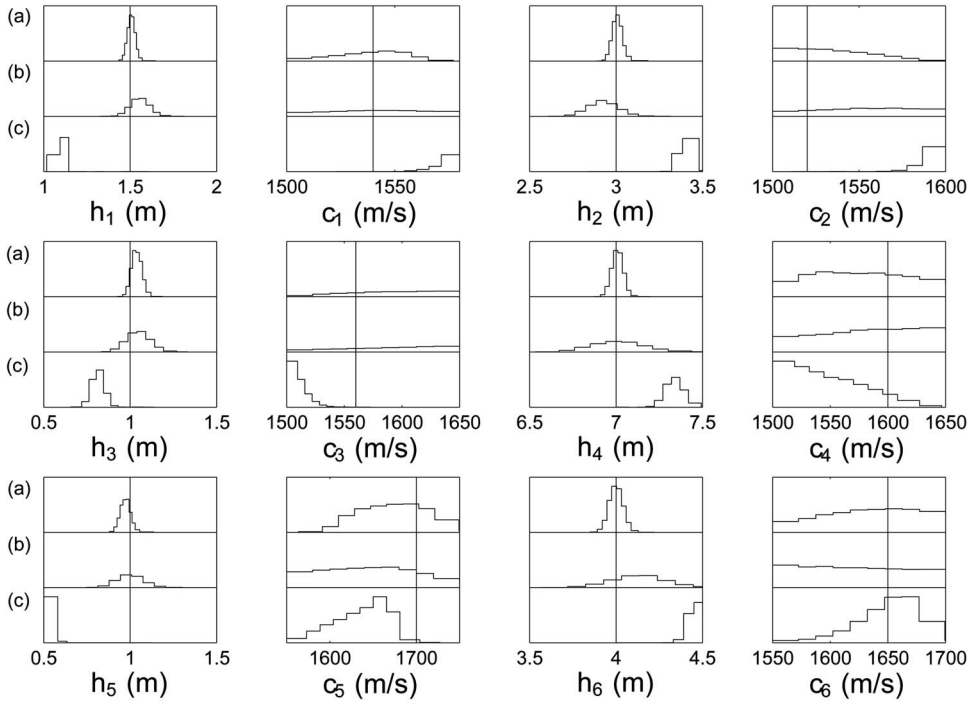


FIG. 4. Recovered marginal probabilities for data simulated by ray tracing (a) without biased errors, (b) with biased errors and additional unknown parameters accounting for biases, and (c) with biased errors that are ignored in the inversion. True parameter values are given as solid lines. Case (b) shows the most realistic results, while (a) indicates overly confident results and (c) is strongly biased.

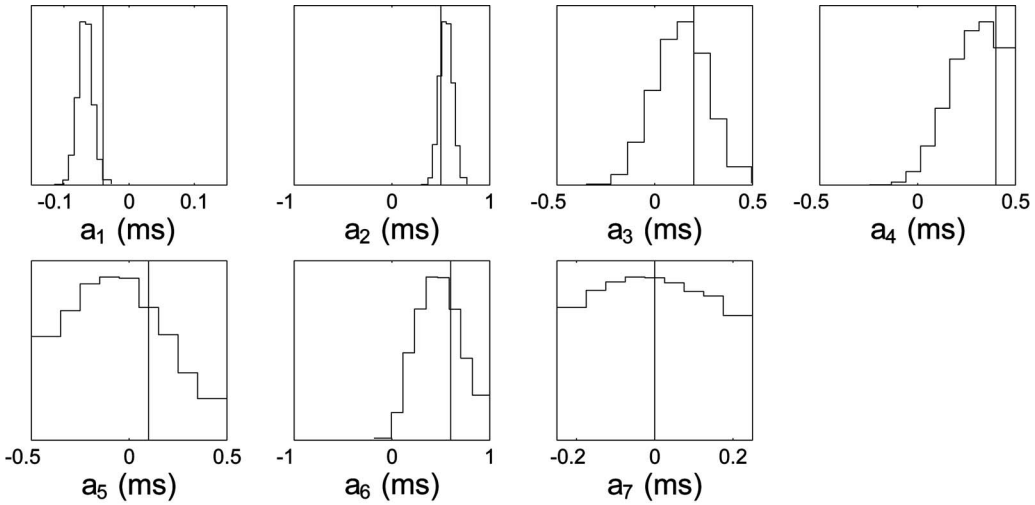


FIG. 5. Marginal probability distributions of the recovered picking (bias) offset parameters that were added to the simulated ray-tracing data. True values are given as solid black lines. The marginals indicate sensitivity of the inversion to the bias parameters and yield reasonable estimates.

B. Data simulated with ray tracing

To investigate the resolution for an idealized travel time inversion problem with picking errors that are unbiased and uncorrelated, a synthetic travel time data set was generated by ray tracing and zero-mean uncorrelated Gaussian errors with standard deviation of $\sigma=0.1$ ms were added. The true model consists of six sediment layers (i.e., seven reflectors) shown in Fig. 3 and a constant sound-velocity water column of 150 m depth. The model was chosen with several thin layers (1 m) to represent a challenging test case. The source was placed at 0.35 m depth below the water surface and the receiver at 122 m depth.

The inversion results for unbiased data errors are compared to results for similar data but with realistic constant time offsets (biases) of 0–0.6 ms added to each reflector hyperbola. For these data, two inversions are performed. In one inversion, the picking offsets are taken into account by including unknown offset (bias) parameters a_i for each reflector in the inversion. In the other, the data are inverted without including the bias parameters. Figure 4 shows the resulting marginal distributions for all three cases. Comparing marginal distributions for the unbiased data [marginals labeled (a) in Fig. 4] to those for biased data with bias parameters included in the inversion [marginals labeled (b)] indicates that the presence of data biases significantly reduces geoacoustic information content (widens marginals). However, layer thicknesses are still well resolved, although layer velocity resolution is relatively poor (for both cases). Figure 4 also shows that neglecting picking biases [marginals labeled (c)] can result in geoacoustic parameter estimates that are strongly biased with unreasonably tight credibility intervals that misrepresent the actual information content of the picked data.

Figure 5 shows the marginal distributions for the recov-

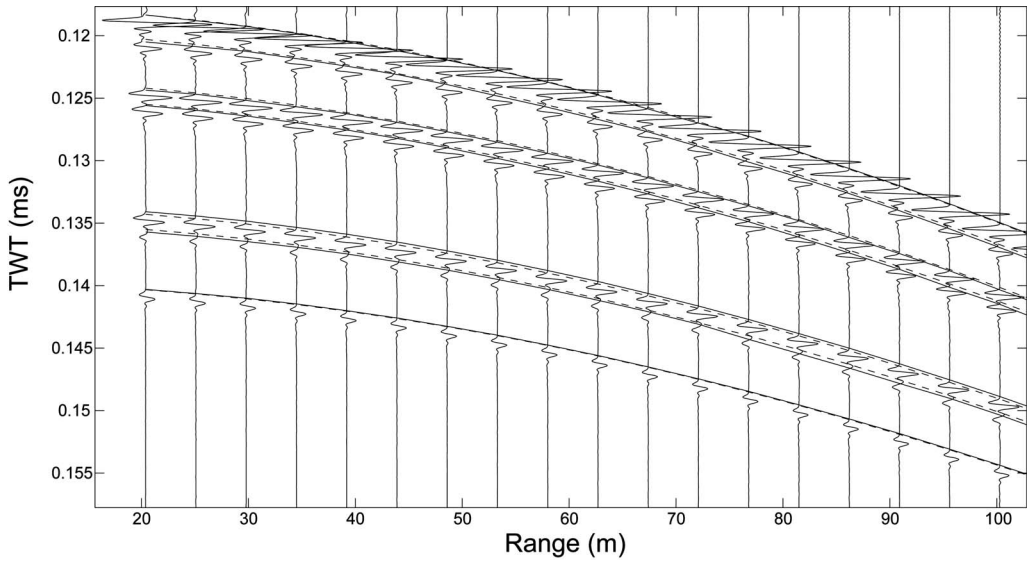


FIG. 6. Simulated acoustic traces and picked travel times for six sediment layers (solid lines). The dashed lines indicate the hyperbola for the true model as computed by ray tracing. Note that in some cases there is an offset between the picked travel times and true solution representing a travel time picking bias. Only a crop of ranges from 20 to 100 m is shown for clarity.

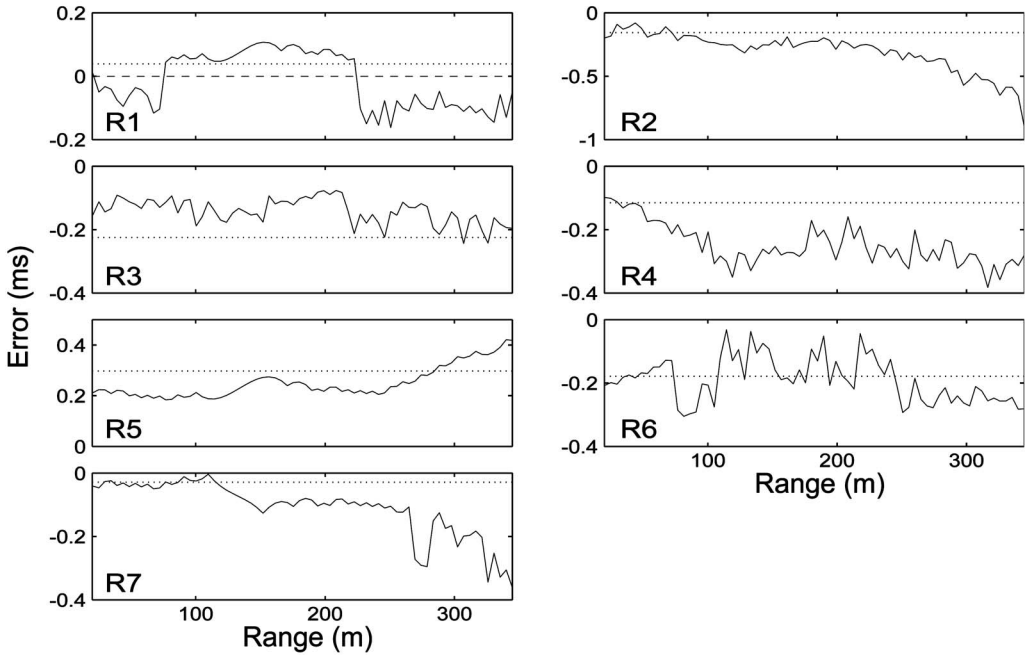


FIG. 7. Picking errors (difference between true travel times and picked travel time data) for the seven reflectors. Dotted lines indicate the time offsets recovered by the inversion. The recovered offsets agree well with the picking errors to first order. Zero is given as dashed lines.

ered time offset (bias) unknowns. The prior bounds for the last offset parameter a_7 were chosen to be narrower than the others to simulate higher confidence in the travel time picks, which is a common feature for some reflectors in measured data. All offsets are resolved reasonably well, indicating the data have sufficient information to resolve biases as well as geoacoustic parameters.

C. Data from synthetic seismo-acoustic traces

To consider more realistic correlated picking errors, a full-wave-field model (OASES[23]) was used to compute the Green's function for the environmental model of Sec. III B. The Green's functions were convolved with the source wavelet shown in Fig. 3 to produce realistic seismo-acoustic traces. The wavelet was taken from experimental data by averaging four traces recorded at short ranges and then low-pass filtering with a 3000 Hz cut-off. The simulated acoustic traces are shown in Fig. 6. Picking arrival times from these traces results in a simulated data set that includes effects of the resolution limitations due to the finite pulse-length which is comparable to several of the model layer thicknesses (see Fig. 3). This leads to realistic picking errors, including biases

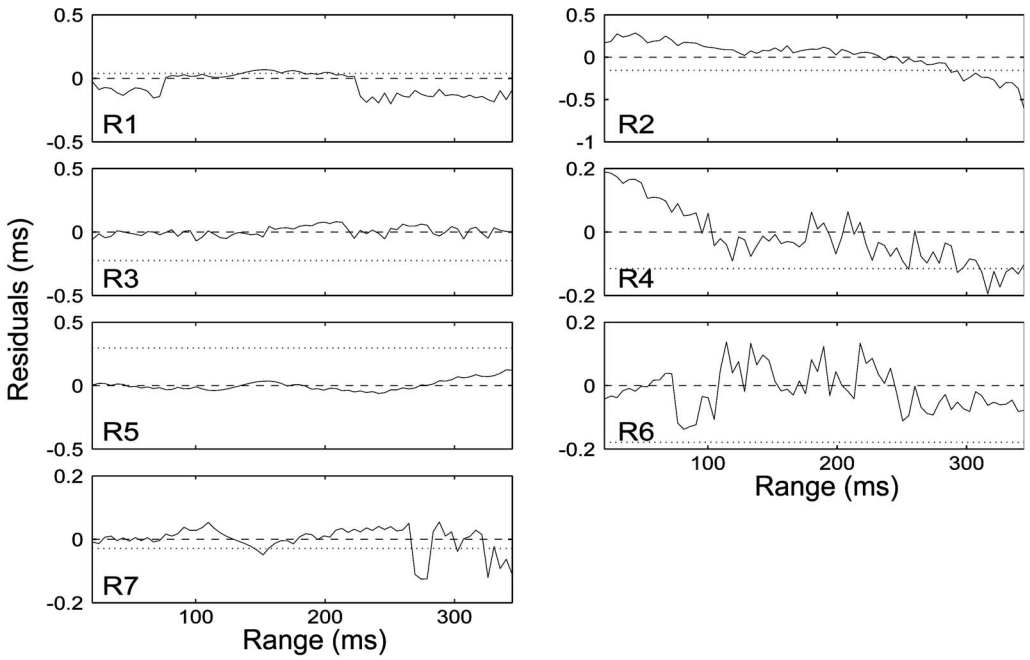


FIG. 8. The data residuals for picked and MAP replica data for seven reflectors. The dotted line indicates the recovered time offset parameters, zero is given as dashed lines.

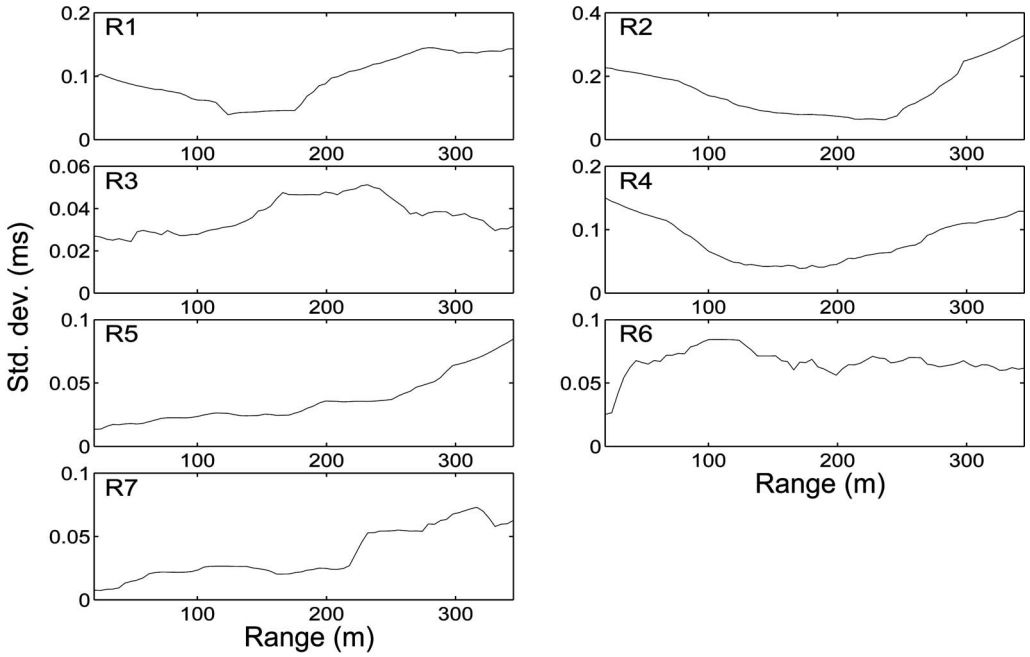


FIG. 9. Standard deviation estimates averaged with 20 point running rms filter. Significant changes in standard deviation across range indicate nonstationarity of the residuals. Note that the magnitude of the errors is different from reflector to reflector.

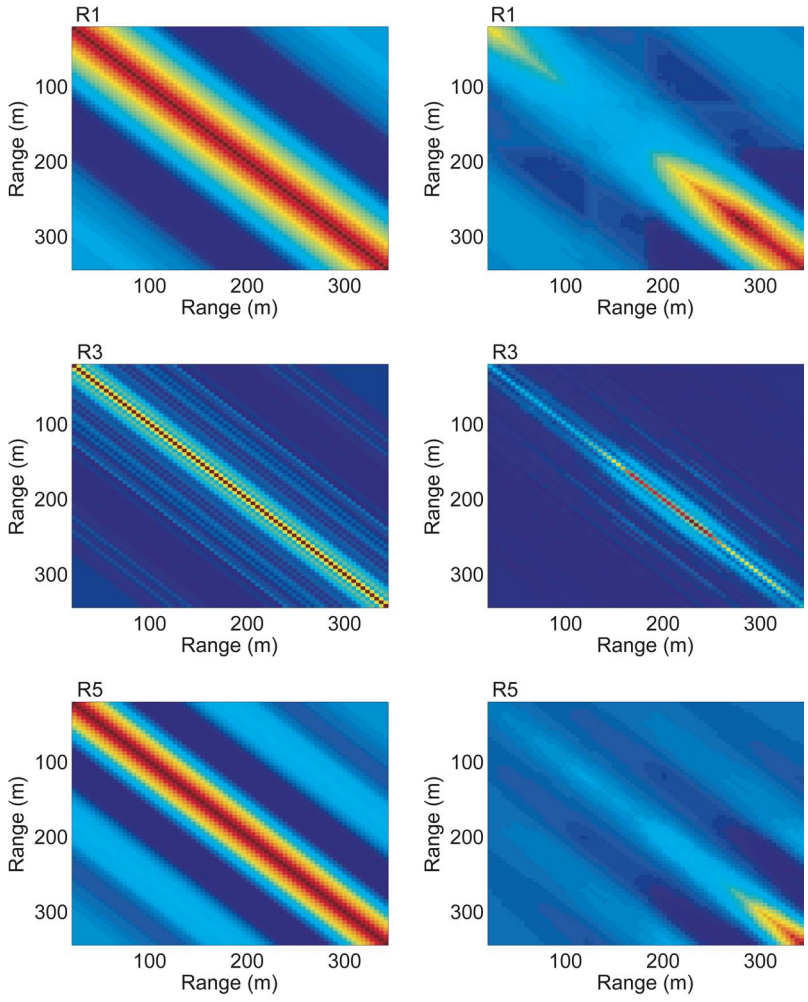


FIG. 10. (Color online) Covariance matrix estimates for travel time inversion before (left column) and after correcting for nonstationarity (right column) for reflectors R1, R3, and R5. Correcting for nonstationarity results in non-Toeplitz matrices.

TABLE I. Numerical values for true parameters and inversion results (MAP model and 95% HPD credibility intervals) for picked travel time data. Picking was based on simulated seismo-acoustic traces.

	Thickness (m)			Sound velocity (m/s)			
True	Prior	MAP	95% HPD	True	Prior	MAP	95% HPD
1.5	0.10–8.00	1.34	1.24–1.73	1540	1500–1580	1501	1500–1558
3.0	0.10–8.00	3.19	2.67–3.24	1520	1500–1600	1516	1500–1575
1.0	0.10–8.00	1.10	1.00–1.39	1560	1500–1650	1613	1542–1648
7.0	0.10–8.00	6.84	6.74–7.01	1600	1500–1650	1608	1516–1645
1.0	0.10–8.00	0.85	0.79–1.04	1700	1550–1750	1614	1550–1684
4.0	0.10–8.00	4.19	3.90–4.19	1600	1550–1700	1569	1554–1684

and nonstationary correlated data errors. Travel times were picked from these traces as outlined in Sec. III A. Figure 6 compares the picked travel times to those computed with ray tracing. For several reflectors, the picked travel times show an offset or bias as discussed previously.

The picking bias is illustrated more clearly by plotting the differences between the true and picked data, which represents the picking error, as shown in Fig. 7. Note that the errors are strongly biased for some reflectors and show strong serial correlations (i.e., are not randomly distributed about zero). The resulting MAP estimates for the time offset parameters for each reflector (included in the inversion) are also shown in Fig. 7 and fit the mean picking error reasonably well for most reflectors. The effect of the error correlations and biases are significantly reduced by including time offsets in the inversion. Although the time offsets do not account for all correlations, the error appears to be accounted for to first order. Of course, the picking errors are not available when considering measured data since the true (error-free) data are not known. However, the data residuals (difference between MAP replica and picked data) are available, and are shown in Fig. 8.

Even with constant offset parameters included in the inversion, the data residuals (Fig. 8) indicate strongly correlated data errors. Further, the residuals are nonstationary, in that the magnitude of the standard deviation changes significantly across range, as illustrated by a running rms average over 20 points (Fig. 9). Examples of the nonstationary covariance matrices, computed as outlined in Sec. II B, are shown in Fig. 10. The left side shows unscaled Toeplitz covariance matrices for reflectors R1, R3, and R5. The strong off-diagonal terms represent the correlated errors. The right side shows the data covariance matrices for the same reflectors but scaled to account for the nonstationary errors which results in a non-Toeplitz form.

After applying Gibbs sampling inversion using the covariance matrix estimates (prior bounds given in Table I), the information recovered from the PPD is given as marginal

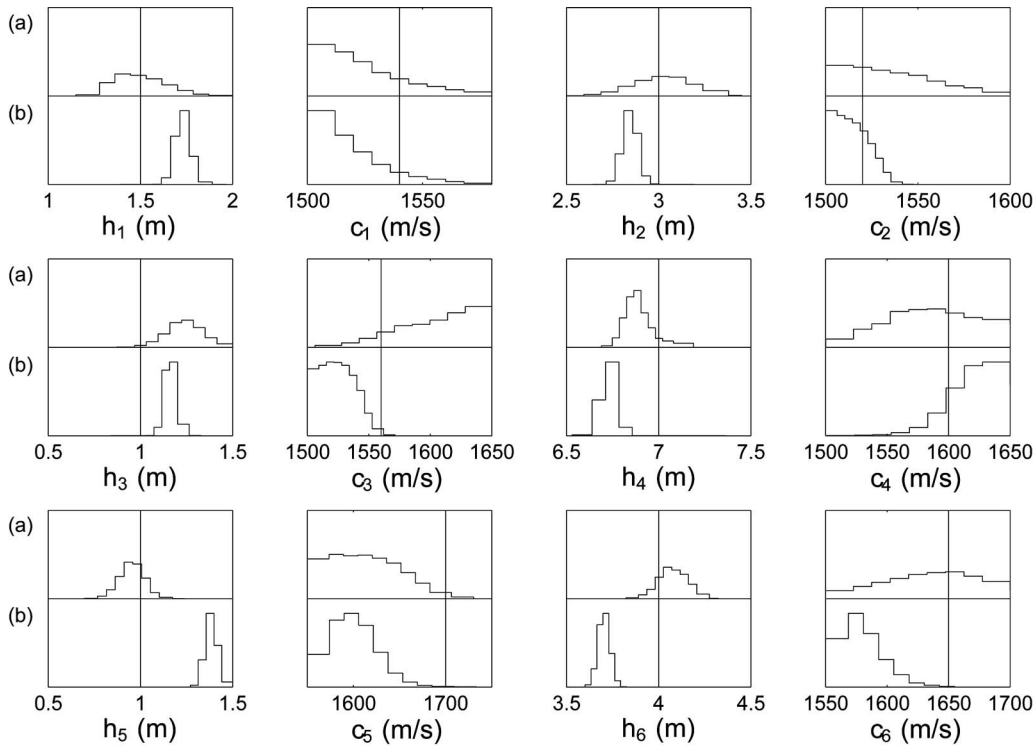


FIG. 11. Recovered marginal probability distributions for picked travel time data from simulated seismo-acoustic traces (a) including biases as unknown parameters and (b) ignoring biases. Ignoring the data biases results in biased and overly confident inversion results.

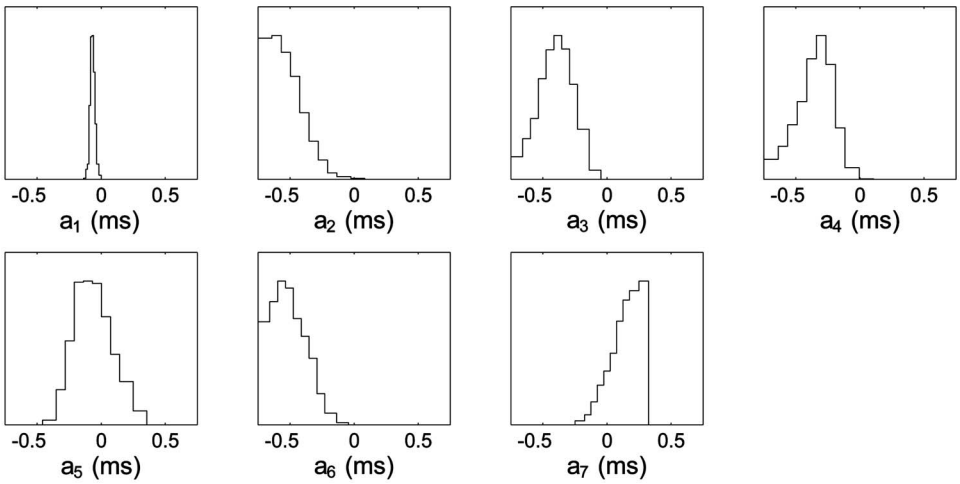


FIG. 12. Recovered marginal probability distributions of picking offsets of seven reflectors for the inversion of picked travel time data from simulated seismo-acoustic traces. The inversion is sensitive to the bias parameters and resolves all parameters within their prior bounds.

distributions for the physical parameters in Fig. 11 for two cases. First, travel time bias unknowns are included in the inversion, and second, the biases are ignored. It can be seen that ignoring picking biases causes unrealistic and biased inversion results. Figure 12 shows the marginal probabilities for the recovered bias parameters. Numerical values for the MAP parameter estimates and the credibility intervals for the 12 physical parameters are given in Table I. It can be seen that layer thicknesses are well resolved within their prior bounds.

It should be noted that data biases can be difficult to detect in measured data. The misfit of an inversion that includes unknown bias parameters can be very similar to the misfit of an inversion that neglects biases. However, neglecting biases can cause misleading inversion results. For reflection travel-time inversion, it is possible to generate synthetic data that are independent of the ray tracing forward model used in the inversion. By computing seismo-acoustic traces that include the actual source pulse, key aspects of the errors associated with picking travel times can be simulated. In particular, the model shown in Fig. 3 includes thin layers in comparison to the pulse length, leading to overlapping arrivals for adjacent reflectors. In such cases arrival first breaks are not clear, often leading to biases and correlated errors in picking. Hence, picking travel times from these synthetic traces produces realistic biased and nonstationary correlated errors. The actual data errors can then be recovered by ray tracing the true model. Hence, this simulation procedure is able to capture the main error processes that apply to measured data.

It is important to realize that the effect of including offset parameters may not be understood by only examining the

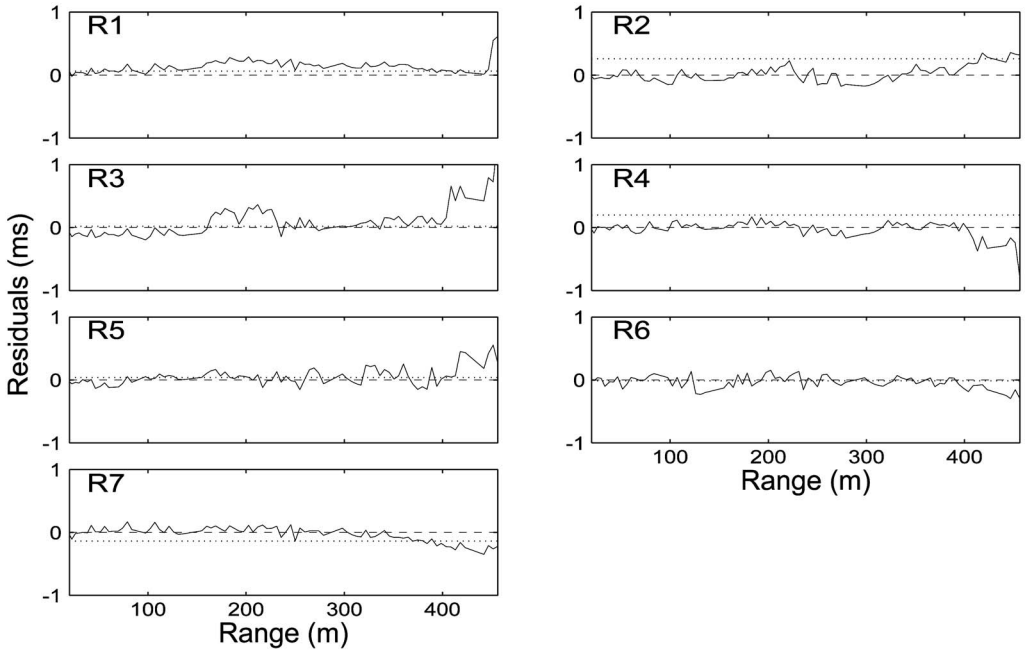


FIG. 13. Data residuals for picked and MAP replica data for seven reflectors (solid lines). Dotted lines indicate the recovered MAP time offset (bias) parameters. Reflectors R2 and R4 show some of the strongest recovered biases. Dashed lines indicate zero.

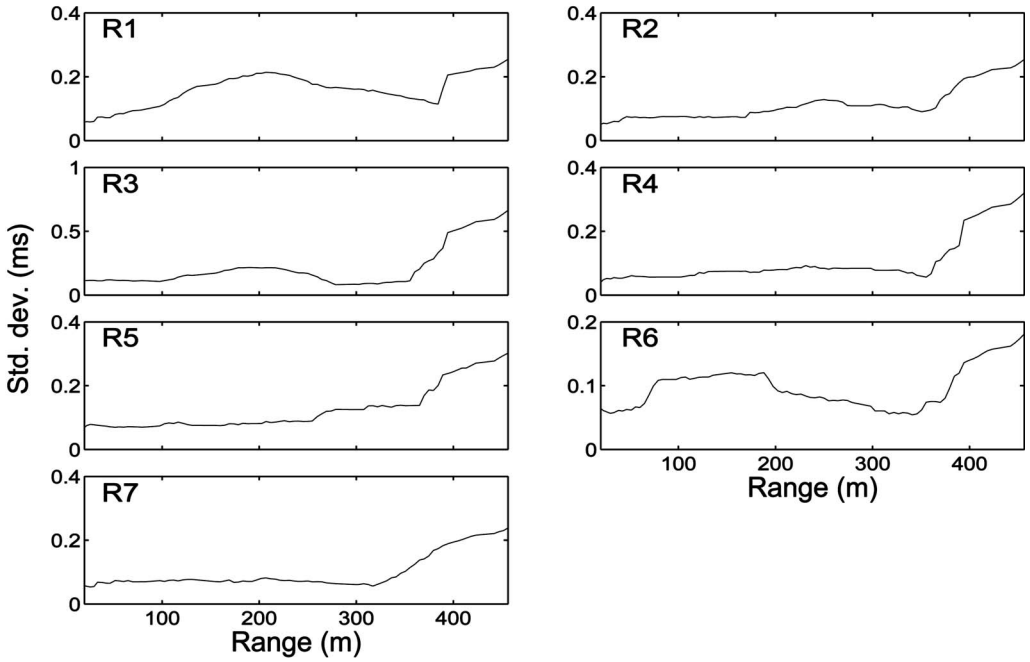


FIG. 14. Estimate of the standard deviation with 20 point running rms filter. Significant changes in standard deviation across range indicate nonstationarity of the residuals.

residuals and validating the statistical assumptions, as measured data may be fit well with or without including bias parameters in the inversion. Neglecting offsets can result in a substantial theory error that biases the resulting PPD as shown in Sec. III B. To quantify the effect of including offsets in the simulation, two different misfits can be calculated for the jth reflector

$$E_{1i} = \sum_j \left(\frac{t_{ij}^{\text{pick}} - t_{ij}^{(\text{true})}}{\sigma_{ij}} \right)^2$$

$$E_{2i} = \sum_j \left(\frac{t_{ij}^{(\text{pick})} - t_{ij}^{(\text{true})} - a_j}{\sigma_{ij}} \right)^2, \tag{20}$$

where σ_{ij} is estimated as described in Sec. II B, $t^{(\text{true})}$ refers to the true travel times obtained through ray tracing, and $t^{(\text{pick})}$ are the picked travel times used in the inversion and which contain biased, correlated errors. Here, E_{2i} is the misfit accounting for constant time offsets and E_{1i} is the misfit neglecting the offsets. Further, it is useful to look at the arith-

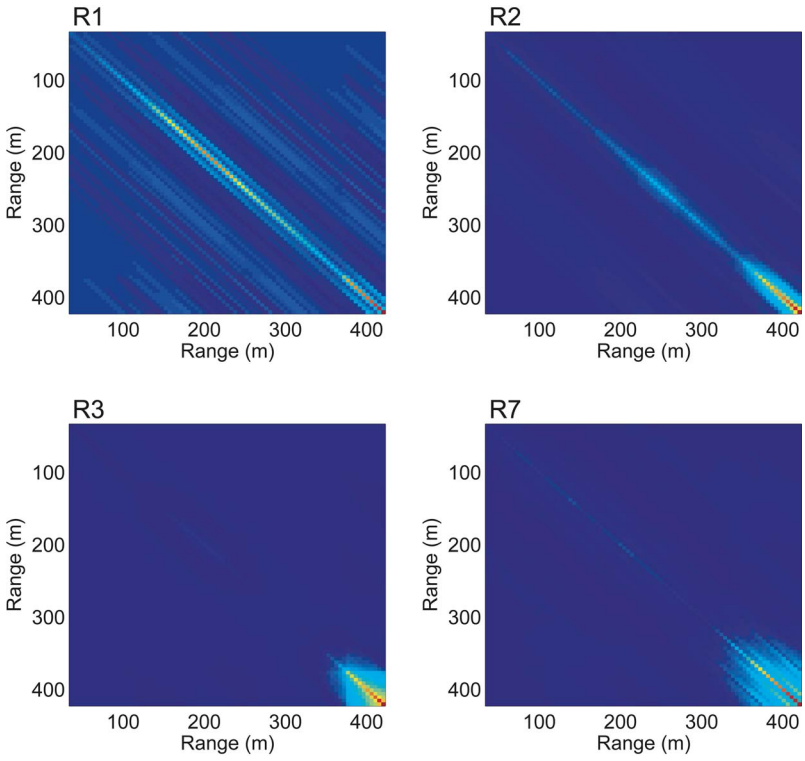


FIG. 15. (Color online) Selected travel time data covariance matrix estimates for reflectors one, two, three, and seven. The non-Toeplitz matrices indicate nonstationary data errors.

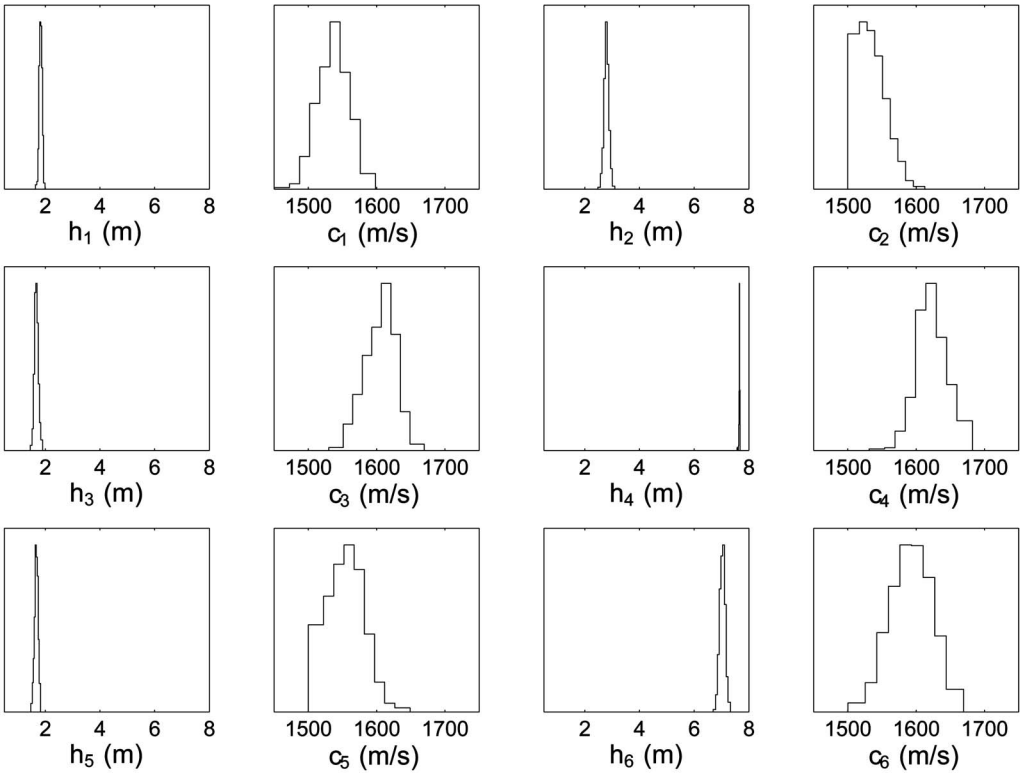


FIG. 16. Recovered marginal probability distributions of physical model parameters for the inversion of picked travel time data (Malta Plateau). The prior bounds for layer thicknesses h_i correspond to the width of the plots.

metic mean values which represent the mean offset

$$\bar{t}_{1i} = \mathrm{MEAN}(\mathbf{t}_i^{(\mathrm{pick})} - \mathbf{t}_i^{(\mathrm{true})}),$$

$$\bar{t}_{2i} = \mathrm{MEAN}(\mathbf{t}_i^{(\mathrm{pick})} - \mathbf{t}_i^{(\mathrm{true})} - a_i), \tag{21}$$

where the travel time vectors represent the travel times across range for the ith reflector. Similar to E_{1i} and E_{2i}, $\bar{t}_{1i}$ and $\bar{t}_{2i}$ give a measure of how well the true travel time data are fit. The closer $\bar{t}_{2i}$ is to zero, the better the fit. In the synthetic example of Fig. 3 the average values of the absolute values of all reflectors are $\bar{E}_1 = 68.336$, $\bar{E}_2 = 30.207$, $\bar{t}_1 = 0.1798$ ms, and $\bar{t}_2 = 0.0793$ ms. The fact that $\bar{E}_2$ and $\bar{t}_2$ are significantly smaller than $\bar{E}_1$ and $\bar{t}_1$, respectively, indicates that including bias parameters in the inversion fit the true travel time data better.

IV. MALTA PLATEAU DATA

This section applies the travel time inversion to seismo-acoustic data collected during the SCARAB98 experiment on the Malta Plateau, Mediterranean Sea. The goal is to extract multilayer sound-velocity structure for the upper parts of the seabed, including rigorous uncertainty estimates.

The acoustic data were generated with an electromechanical impulsive source (EG&G model 265 Uniboom, referred to as a "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 with 16 Benthos AQ-4 hydrophones. The hydrophone used in this data set was at 112 m depth and the water depth was 153 m. The sound-velocity profile was fairly constant with the sound velocity varying less than 5 m/s over the water column.

The seismo-acoustic traces collected at the experiment site have very low noise levels and are shown in Fig. 1. The

TABLE II. Numerical values of prior information and inversion results (MAP model and 95% HPD credibility intervals) for picked travel time data.

h (m)			c (m/s)		
Prior	MAP	95% HPD	Prior	MAP	95% HPD
0.10–8.00	1.79	1.63–1.97	1450–1600	1530	1457–1595
0.10–8.00	2.80	2.51–3.05	1500–1600	1508	1500–1596
0.10–8.00	1.66	1.45–1.87	1500–1700	1581	1537–1662
0.10–8.00	7.65	7.56–7.67	1500–1700	1637	1537–1679
0.10–8.00	1.61	1.47–1.82	1500–1650	1556	1500–1628
0.10–8.00	7.06	6.68–7.28	1500–1700	1569	1500–1664

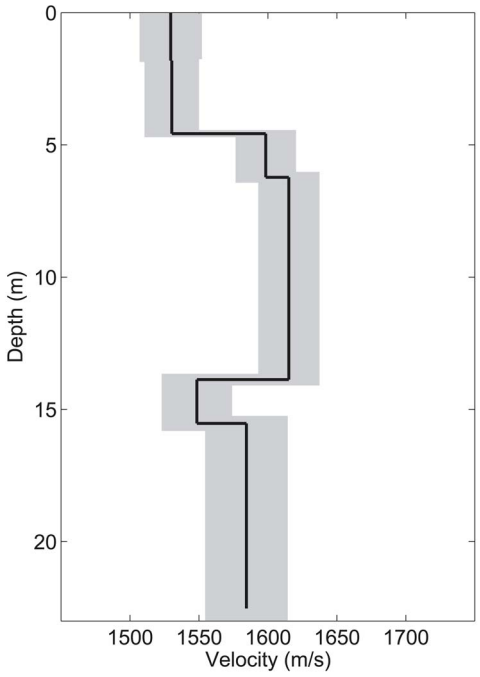


FIG. 17. (Color online) The recovered sound-velocity profile with one standard deviation credibility intervals. While layer thicknesses are well resolved, resolution for sound velocities is limited.

time series show a direct arrival as well as numerous seabed reflections, including the water-sediment interface and deeper layers. The direct arrival or water wave can be seen as the first event at ranges larger than 150 m. The sediment-water interface reflection occurs at approximately 0.130 s at the shortest range, followed by a number of subbottom reflections. The first water column multiple can be seen at approximately 0.275 s, which gives good separation from the main signal. Gaps due to missing traces in Fig. 1 are caused by data being written to disk, which interrupted the recording.

The travel time data were picked from the seismo-acoustic traces as discussed in Sec. III A. Particular attention was paid to pick the travel times objectively without altering the picks by interpreting the seismo-acoustic traces. The resulting picks are shown in Fig. 1. Seven reflectors including the sediment water interface were picked. Travel times were interpolated over the gaps to simplify the processing, particularly the calculation of covariance matrices and the statistical validation of assumptions, which are based on uniformly spaced residuals. Since the travel times can be assumed to be well behaved in these gaps, which were at most three traces wide, linear interpolation was found to be sufficient. This resulted in a travel time data set consisting of 83 picks per reflector for ranges of 30–420 m with a range spacing of ~4.7 m. The maximum range was limited because travel time differences (due to relatively thin layers) became too small to allow useful picking at larger ranges. The travel time data is inverted for a model consisting of six sediment layers and the geoacoustic parameters are layer thicknesses and sound velocities for these layers.

The data residuals between the arrival times predicted for an initial MAP model and the picked data are shown in Fig. 13 and indicate correlated errors. Figure 13 also shows recovered time offset parameters for all reflectors [as introduced in Eq. (18)]. The recovered offset parameters illustrate typical limitations in picking. For example, large biases are evident for reflectors R2 and R4, and first breaks for both these reflectors were difficult to pick since corresponding arrivals closely follow (and overlap with) earlier arrivals (Fig. 1). These complications can be due to single thin layers where the events of two different reflectors overlap, or can be due to multiple thin layers below the resolution limit of the source pulse. Figure 14 shows a 20-point rms average of the estimated standard deviation of the data residuals. The variable standard deviation level across range indicates nonstationarity of the data residuals. Due to the correlated and nonstationary errors, non-Toeplitz data covariance matrices were estimated as described in Sec. III A. Figure 15 shows the covariance matrix estimates for reflectors R1, R2, R3 and R7, with reflectors R3 and R7 showing strongly non-Toeplitz structure.

The Bayesian inversion (Sec. II) was then applied to the travel time data weighted according to the estimated data covariance matrices. The resulting marginal distributions for layer thicknesses and sound velocities are given in Fig. 16. It can be seen that all layer thicknesses are resolved well within the prior bounds (which correspond to the plot widths). The

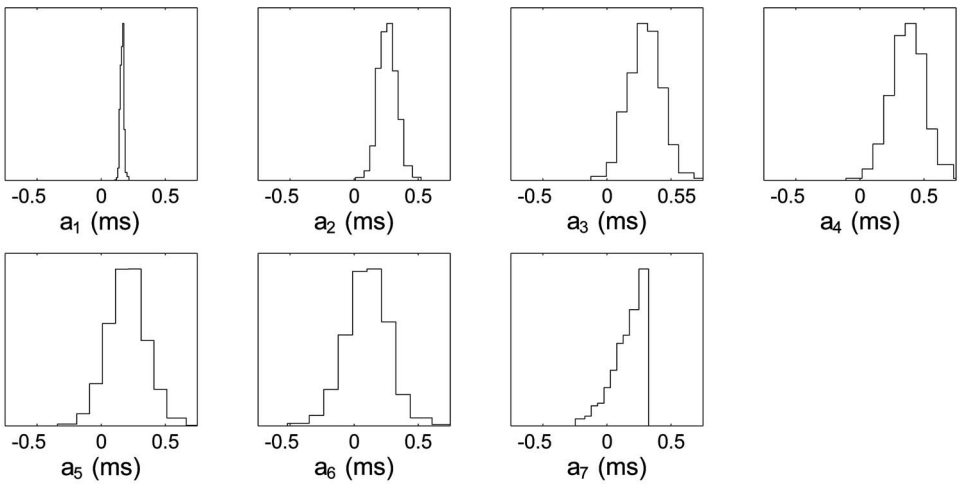


FIG. 18. Recovered marginal probability distributions of picking offset parameters of seven reflectors for the inversion of picked travel time data. The inversion shows good sensitivity for recovering the biases.

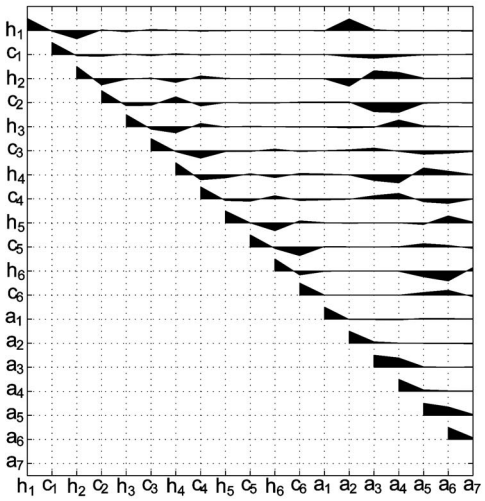


FIG. 19. Model parameter correlation matrix. For simplicity, only the upper triangular matrix of the symmetric matrix is plotted. Several strong correlations are visible with the most prominent ones being between physical parameters and bias parameters.

sound velocities are reasonably well resolved. The numerical values for prior bounds, 95% HPD credibility intervals, and the MAP model for the physical parameters are given in Table II. The mean profile with one standard deviation credibility bounds is given in Fig. 17. This result will be discussed further in Sec. V.

The recovered marginal probability distributions for the travel time offset parameters are shown in Fig. 18. The prior bounds for the pick offsets were set depending on the confidence level in the picks for a certain reflector. The first six reflectors were assigned intervals of $[-0.75, 0.75]$ ms. For the seventh reflector, a tighter interval of $[-0.3, 0.3]$ was chosen, since the arrival was well separated from other events and the picking was comparably straightforward. The marginal probability distributions indicate good resolution for all offset parameters, although the last reflector is constrained by the prior bounds (due to high confidence in the picking).

The model parameter correlation matrix for both physical and nuisance parameters is given in Fig. 19 and shows strong correlations between numerous parameters. It can be seen that the layer thicknesses of several adjacent layers show strong negative correlations (e.g., h_1 and h_2). This is due to the fact that the negative correlation conserves the total thickness of the two layers. A similar argument applies to negative correlations for sound velocities of adjacent layers. The strongest correlations, however, are between physical parameters and the picking offset nuisance parameters. These interparameter relationships are illustrated in Fig. 20 as joint marginal probability distributions. The two largest correlations are found between h_1 and a_2 (0.96) and between c_2 and a_3 (−0.75), and are due to the strong coupling between a time offset parameter and a layer thickness or sound velocity. The strong correlations are largely responsible for widening the marginal distributions for layer thicknesses and sound velocities after including the time offsets as nuisance parameters. Hence, including the time offsets is considered crucial to obtain meaningful uncertainty estimates and avoid overly optimistic (misleading) results.

It is of fundamental importance to check the quantitative inversion results by examining the assumptions made for the inversion (Sec. II C). In particular, the standardized residuals $\mathbf{L}(\mathbf{d}(\mathbf{m})-\mathbf{d})$ should resemble an uncorrelated Gaussian process and statistical tests should show an improvement over the results for the raw residuals (scaled only by their standard deviation). To examine the randomness of the data residuals, a runs test was performed and the resulting p values are shown in Fig. 21. The raw residuals fail the runs test for every reflector with p values $\sim 10^{-4}$. After applying the data covariance matrix estimate, five out of seven reflectors pass

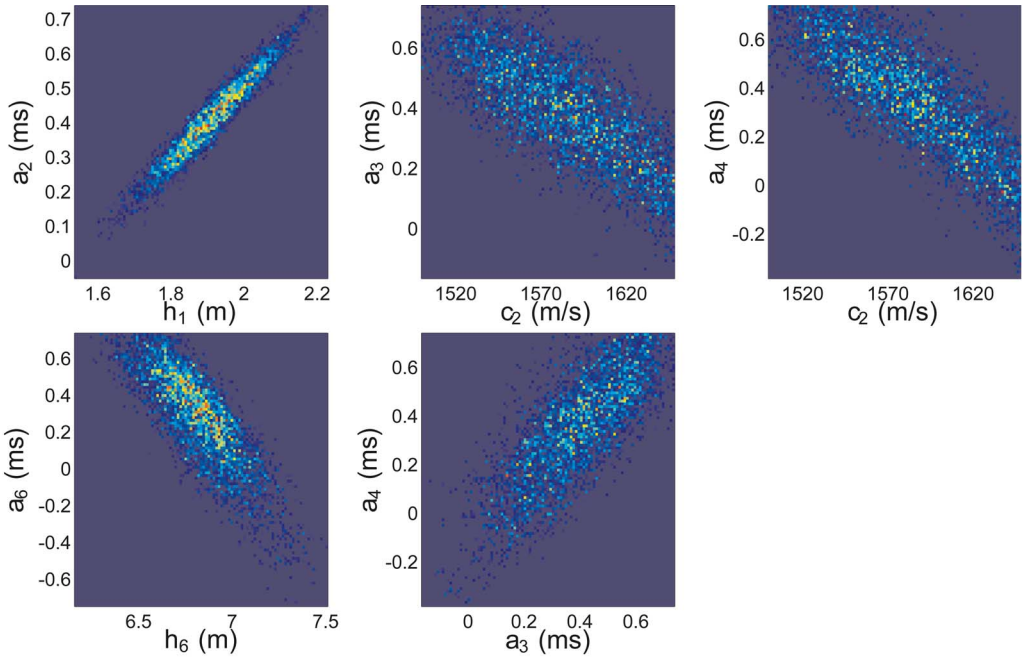


FIG. 20. (Color online) Selected joint marginal distributions between physical parameters (h_i and c_i) and nuisance parameters (constant time picking offsets, a_i). The narrow, diagonal shape indicates strong correlations.

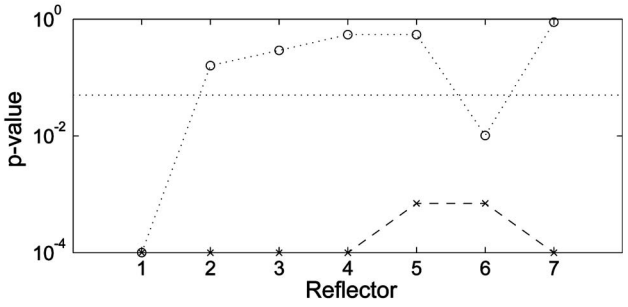


FIG. 21. Travel time inversion runs test p values. The crosses show the p values for the raw residuals, while the open circles show the p values for the standardized residuals. The dotted line indicates the 0.05 confidence level.

the runs test at a 0.05 level. This indicates that the data covariance matrix estimates account for much (but not all) of the data error covariances.

To examine the Gaussianity of the residuals, Fig. 22 shows histograms of raw and standardized residuals. In addition, the theoretical Gaussian distribution is included on each plot. It can be seen that the data residuals are reasonably Gaussian for most reflectors (with the notable exception of R1) and do not contain large outliers. To quantify the Gaussianity of the residuals, the KS test was performed. The p values were evaluated from a table of critical values that was computed for the appropriate number of data (Sec. II C). The p values for the test are shown in Fig. 23. It can be seen that applying the covariance matrix does not significantly change the Gaussianity of the residuals. For the standardized residuals, four out of the seven reflectors pass at an 0.05 level and reflector five almost passes with a p value of 0.035. Reflector R1 clearly failed the validation process. This is somewhat surprising, since R1 (the water-sediment interface) was particularly clear to pick. However, it should be noted that the inversion evaluates all reflectors at once and that good fit of several reflectors can sometimes be obtained at the cost of another reflector.

While the above-presented validation of assumptions is not completely satisfactory in a statistical sense, the tests show that the assumptions appear reasonable and that the inversion likely gives useful estimates and uncertainties for the geoacoustic parameters given the model parametrization.

V. SUMMARY AND CONCLUSIONS

This paper developed a nonlinear Bayesian inversion approach for seabed reflection data that included rigorous uncertainty estimation and examination of statistical assumptions. Travel time data were picked on seismo-acoustic traces and inverted for sediment layer thicknesses and sound velocities. Particular attention was paid to picking errors. As picking errors are often biased, correlated, and nonstationary, non-Toeplitz data covariance matrices were estimated from data residual analysis and included along with unknown travel time offset (bias) parameters in the inversion. Simulated experiments with realistic data errors showed that neglecting error covariances and biases can cause misleading inversion results with unrealistically high confidence.

Statistical assumptions were examined *a posteriori* to ensure meaningful inversion results. The form of the residuals distribution was quantitatively tested with the Kolmogorov-Smirnov test, and randomness was quantified with the runs test. Histograms of the residuals were also plotted as qualitative checks.

While validating statistical assumptions about data errors is important to support the inversion results, it is difficult to test for systematic errors and biases in measured data. This paper developed realistic computer simulations emulating many aspects of the actual experiment which permitted an

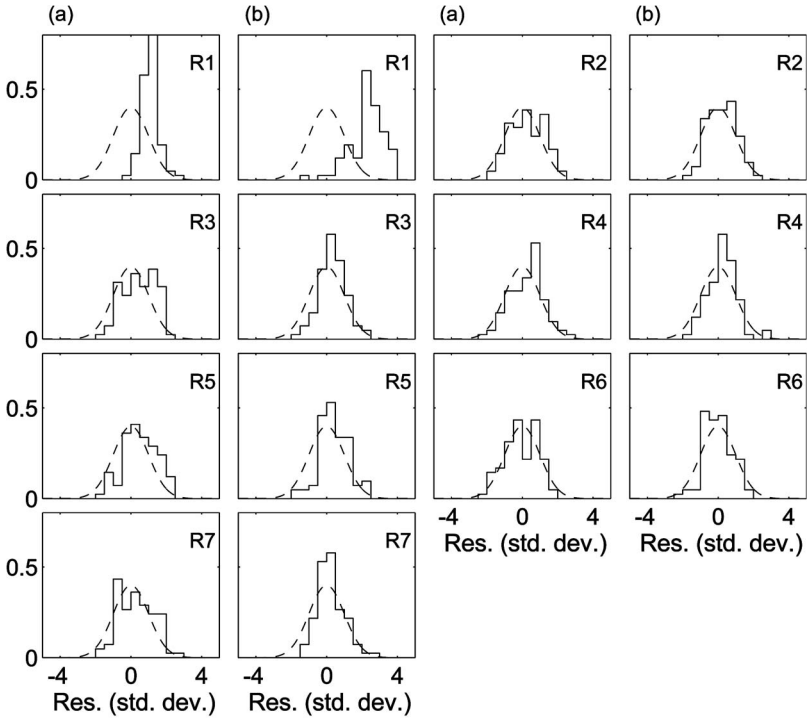


FIG. 22. Histograms of raw (a) and standardized (b) residuals as a qualitative measure of Gaussianity of data errors. The dashed lines indicate theoretical zero mean Gaussian distributions of unit standard deviation. The agreement is reasonable for most reflectors for raw and standardized residuals.

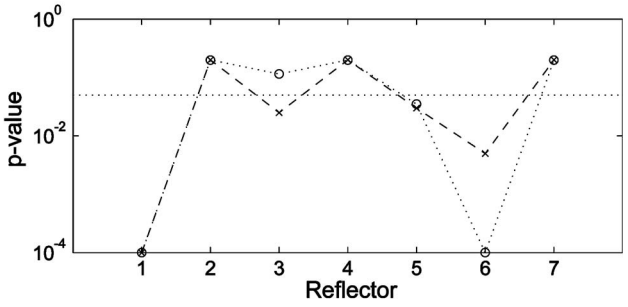


FIG. 23. Travel time inversion KS test p values. The crosses show the p values for the raw residuals, while the open circles show the p values for the standardized residuals. The dotted line is the 0.05 level of confidence.

evaluation of realistic errors (including biases and correlations). It was found to be important to include travel time offset parameters in the inversion to account for biased travel time picks and avoid strongly biased inversion results.

The travel time inversion was also applied to shallow-water reflection data from the Malta Plateau. The model parameters consisted of layer thicknesses and sound velocities for six sediment layers. Biased, correlated, and nonstationary errors were accounted for as described earlier. The recovered travel time offsets were consistent with the picking confidence. Within the prior search bounds, layer thicknesses were well resolved while sound velocities were less well resolved. However, the ratio of the prior bound width relative to the parameter value is much larger for thicknesses than for velocities (i.e., in a relative sense, velocity resolution is comparable to thickness resolution). Finally, the statistical assumptions for the inversion were examined *a posteriori*.

It should be noted that parameter uncertainties will generally depend strongly on the model itself. For example, thin layers lead to error-prone arrival picking as well as a general lower sensitivity to velocity. Hence, larger geoacoustic uncertainties were obtained for the simulated data which involved thinner layers than what appear to be present at the Malta Plateau site. Further, even smaller parameter uncertainties could be expected in environments involving thicker layers than those considered here.

A key point in this paper is that reliable estimation of posterior probabilities require a physically meaningful specification of the likelihood function. This involves at least three components. First, an appropriate form for the error distribution (e.g., Gaussian) must be chosen and examined *a posteriori*. Second, data error correlations must be quantified and included in the inversion by means of data covariance matrices. Third, possible error biases must be accounted for. All three aspects are straightforward to implement in the numerical Bayesian approach used here and were applied in the inversion of data measured on the Malta Plateau.

Finally, since the reflection travel-time inversion presented here constrains layer thicknesses well, this inversion could be combined with another approach, such as reflection coefficient inversion, which has higher sensitivity to velocities.[14]

ACKNOWLEDGMENTS

The authors thank the NATO Undersea Research Center for supporting the at-sea data collection. We also gratefully acknowledge the support of the Office of the Naval Research. The first author was supported by the Natural Sciences and Engineering Research Council of Canada and the University of Victoria during Ph.D studies.

[1]C. M. Ferla and F. B. Jensen, "Are current environmental databases adequate for sonar predictions in shallow water?," *Impact of Littoral Environmental Variability on Acoustic Predictions and Sonar Performance* (Kluwer Academic, Dordrecht, 1987).
[2]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).
[3]G. M. Bryan, "The hydrophone-pinger experiment," J. Acoust. Soc. Am. **68**, 1403–1408 (1980).
[4]G. M. Bryan, *Physics of Sound in Marine Sediments* (Plenum, New York, 1974), pp. 119–130.
[5]A. Tarantola, *Inverse Problem Theory: Methods for Data Fitting and Model Parameter Estimation* (Elsevier, Amsterdam, 1987).
[6]M. Sambridge and K. Mosegaard, "Monte Carlo methods in geophysical inverse problems," Rev. Geophys. **40**, 3-1–3-29 (2002).
[7]B. A. Brooks and L. N. Frazer, "Importance reweighting reduces dependence on temperature in Gibbs samplers: An application to the coseismic geodetic inverse problem," Geophys. J. Int. **161**, 12–20 (2005).
[8]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).
[9]S. E. Dosso, "Quantifying uncertainty in geoacoustic inversion. I. A fast Gibbs sampler approach," J. Acoust. Soc. Am. **111**, 129–142 (2002).
[10]S. E. Dosso and C. W. Holland, "Geoacoustic uncertainties from viscoelastic inversion of seabed reflection data," IEEE J. Ocean. Eng. (in press).
[11]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).
[12]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).
[13]J. Dettmer, S. E. Dosso, and C. W. Holland, "Geoacoustic inversion with strongly correlated errors," Can. Acoust. **32**, 194–195 (2004).
[14]J. Dettmer, S. E. Dosso, and C. W. Holland, "Bayesian inversion of reflection data for seabed properties of multi-layered systems," in Proceedings of the Eighth European Conference on Underwater Acoustics, 2006, pp. 473–478.
[15]K. Mosegaard and M. Sambridge, "Monte Carlo analysis of inverse problems," Inverse Probl. **18**, R29–R54 (2002).
[16]D. H. Rothman, "Automatic estimation of large residual statistics corrections," Geophysics **51**, 337–346 (1986).
[17]S. Geman and D. Geman, "Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images," IEEE Trans. Pattern Anal. Mach. Intell. **6**, 721–741 (1984).
[18]C. F. Mecklenbräuker and P. Gerstoft, "Objective functions for ocean acoustic inversion derived by likelihood methods," J. Comput. Acoust. **8**, 259–270 (2000).
[19]F. J. Massey, "The Kolmogorov-Smirnov test for goodness of fit," J. Am. Stat. Assoc. **46**, 68–78 (1951).
[20]J. E. Freund, *Modern Elementary Statistics* (Prentice-Hall, Englewood Cliffs, NJ, 1967).
[21]H. W. Lilliefors, "On the Kolmogorov-Smirnov test for normality with mean and variance unknown," J. Am. Stat. Assoc. 399–402 (1967).
[22]S. E. Dosso and G. R. Ebbeson, "Array element localisation and survey design," Can. Acoust. **33**, 16–26 (2005).
[23]H. Schmidt, "OASES version 2.2: User guide and reference manual," Technical Report, MIT, Cambridge, 1999.

Observing individual fish behavior in fish aggregations: Tracking in dense fish aggregations using a split-beam echosounder

Nils Olav Handegard[a)]
Institute of Marine Research, Bergen, Norway

(Received 2 May 2006; revised 5 April 2007; accepted 21 April 2007)

Acoustic instruments are important tools for observing the behavior of aquatic organisms. This paper presents a simple but efficient method for improving the tracking of closely spaced targets using a split-beam echosounder. The traditional method has been a stepwise approach from the detection of echoes, rejection of apparently multiple targets and then tracking the remainder. This is inefficient because the split-beam angles are not included in the initial detection; rather they are only used in the rejection criteria before the subsequent tracking. A simple track-before-detection method is presented, where the phase angles, echo intensities, ranges, and times are used simultaneously, resulting in better detection and tracking of the individual fish. Two test data sets were analyzed to determine the effectiveness of this method at discriminating individual tracks from within dense fish aggregations. The first data set was collected by lowering a split-beam transducer into a herring layer. The second data set, also collected with a split-beam transducer, was from a caged aggregation of feeding herring larvae. Results indicate the potential of target tracking, using a split-beam echosounder, as a tool for understanding interindividual behavior. © *2007 Acoustical Society of America.* [DOI: 10.1121/1.2739421]

PACS number(s): 43.30.Sf [KGF] Pages: 177–187

I. INTRODUCTION

Ecosystem studies depend on knowledge of the individual components. Several studies have shown the feasibility of using various acoustic methods for measuring the behavior of individual targets *in situ*, both for fish (Arrhenius *et al.*, 2000; Torgersen and Kaartvedt, 2001) and plankton (Jaffe *et al.*, 1999; Klevjer and Kaartvedt, 2003). A particularly nice example is Genin *et al.* (2005), where the observed swimming behavior of zooplankton relative to water currents has been used to explain the observed aggregation patterns. Among other acoustical studies concerning the behavior of individuals are the behavior of over-wintering herring in the Ofotfjord (Huse and Ona, 1996), vertical search patterns in fish (Cech and Kubecka, 2002), diel differences in swimming patterns in fish (Gjelland *et al.*, 2004) and zoo-plankton (De Robertis *et al.*, 2003), behavioral changes induced by a trawling vessel (Handegard *et al.*, 2003; Handegard and Tjøstheim, 2005), and the feasibility for devices to prevent fish entering hydroelectric turbine intake (McKinstry *et al.*, 2005). McQuinn and Winger (2003) used manual tracking to investigate the impact of diel-dependent fish behavior on target strength. Riverine and shallow-water research is another large field where acoustic methods have been used to observe fish behavior, with special emphasis on migratory behavior and counting (Enzenhofer *et al.*, 1998; Mulligan and Chen, 1998; Mulligan and Kieser, 1996). This interest is motivated by the fact that echo integration is difficult to apply in a riverine environment.

Schooling behavior is a spectacular pattern in nature, and there have been many attempts to uncover the dynamics of this phenomenon. Parr (1927) introduced the idea of simple repulsive and attractive "forces" between individuals, and these ideas were further developed by Breder (1954) and Sakai (1973). The first individual-based data simulation was reported by Aoki (1982). Similar model approaches have been described by Reynolds (1987) and Huth and Wissel (1992). All these models demonstrate that simple rules on the individual can result in complex school dynamics. However, data to support these models are scarce, and methods capable of quantifying interindividual behavior are needed, in particular for closely spaced individuals. The latter problem is the main motivation for the present study, but the method is general and is also useful for other tasks involving the detection of single individuals, e.g., target strength measurements. The goal of this work, therefore, is to develop an improved method for tracking closely spaced individual targets using split-beam echosounders.

A. The state of the art

Two different acoustic instruments for observing behavior are the multibeam sonar, see, e.g., Jaffe *et al.* (1995), and the split-beam echosounder (Brede *et al.*, 1990; Ehrenberg and Torkelson, 1996). The multibeam sonar can handle several targets at a given range, but the resolution is limited by the number of beams and their opening angles. There are, however, methods to compensate for this problem (Jaffe, 1999; Schell and Jaffe, 2004), but at the cost of being able to observe fewer animals at the same range. The somewhat simpler split-beam echosounder is ineffective when multiple targets are located at the same range (Foote, 1996). The split-

[a)]Electronic mail: nils.olav.handegard@imr.no

beam echosounder does not depend on the grid-cell volume in the same way as the multibeam sonar, but it needs a fair signal-to-noise ratio to work properly (Kieser *et al.*, 2000). Both methods have limitations when observing dense aggregations of targets. This study focuses on the use of split-beam echosounders because of their relative simplicity and availability to researchers.

The split-beam echosounder transmits an echo pulse into the water column, and the backscattered signal is received on four quadrants of the transducer face. The phase differences between the four quadrants are used to estimate the direction to the target, so each sample (or pixel in the echogram) is associated with an intensity and two angles, in addition to the receive time and corresponding range as given by the location in the echogram. If there is a single target at a given range, the angles are representative of the position of that target. However, if there is no target or there are multiple targets at the same range, the angles do not represent the position of a single target. It is thus not possible for two fish at the same range to provide valid observations.

Traditionally, obtaining target tracks from acoustic data has been a two-step process (Ehrenberg and Torkelson, 1996, p. 329). First, the targets are detected with a single-echo-detection (SED) algorithm, and then these detections are combined into tracks making use of their positions in successive pings. The potential of this method for observing fish behavior has been acknowledged for a long time (Foote *et al.*, 1986, 1984; Ona, 1994). In general, target tracking is a well-established field, see, e.g., Blackman and Popoli (1999), and has been further developed for split-beam data (Handegard *et al.*, 2005; Xie, 2000).

The SED targets were originally used to estimate the target strength (TS) of the individual fish within the echo beam (Brede *et al.*, 1990; Foote *et al.*, 1986, 1984; Ona, 1999). For this purpose, it is crucial to avoid two targets being considered as one since this would positively bias the results (Foote, 1996). Different methods to reject SED targets contaminated with multiple targets include the use of phase, amplitude, and echo-duration information from the returned echo, see Soule *et al.* (1996) for an evaluation of these methods. When successful, the results of SED algorithms are high-quality targets with corresponding estimates of TS and location in the beam. However, the SED algorithms are not designed for tracking purposes, and there are often missing pings within a track due to the strict SED rejection criteria. The SED algorithm works on ping-by-ping data, and little effort has been applied to use the temporal dimension of the data to improve single-echo detection. One exception is Balk and Lindem (2000), who use the information in adjacent samples (range and time) to decide whether a sample in the echogram belongs to a target or not, a technique known as cross-filter detection which is used to aid the SED algorithm. This, along with other tracking tools, is implemented in the SONAR 5 software (SONAR 5 user manual, Helge Balk, University of Oslo, Norway). The idea of using the temporal dimension is intellectually appealing, since it utilizes information that "conventional" detectors discard. The SONAR 5 software can also interpolate sample data between already-detected SED targets, leading to better tracking performance than conventional methods.

B. Posttracking detection

In this paper, the target angles, echo intensity, time, range, and the actual tracking results are considered in one single step. It is not based on the traditional stepwise process of detection, rejection, and tracking, where the angles are used as rejection criteria only, and single targets must be passed by the SED algorithm to initiate tracks. The idea is inspired by the track-before-detect approach (Blackman and Popoli, 1999, p. 18). In order to achieve this, all samples above a threshold are initially treated as single targets. Note that a target is typically composed of several samples. The threshold is set lower than the expected intensity of the target echoes, ensuring that no targets are lost. Consequently, low intensity samples where no fish are present are treated as valid targets. Each sample has its own apparent position and intensity. The range and time are determined by the sample (pixel) position in the echogram, while the angles and intensities are given by the pixel values in the echogram and "anglegrams,"[1] respectively. The low threshold results initially in many false targets, but the advantage is that no information is lost in this initial step, as opposed to the traditional approach, where the SED algorithm rejects many targets. This calls for a different approach when associating samples to tracks, by postponing the quality screening until the tracks have been established, which may be denoted "track rejection" as opposed to the single-target rejection applied in the SED algorithm. The main objective of this paper is to develop these techniques.

II. MATERIALS AND METHOD

This section is divided into three main parts. First the test data sets are presented. The second part is the actual method of data association, i.e., associating samples to tracks. Finally the track rejection and track quality algorithms are described. The other aspects of the tracking, like track estimation, incorporating platform movement, etc., are the same as described in Handegard *et al.* (2005). That paper is somewhat technical, and it is not necessary to fully understand the details there to appreciate the ideas presented here. However, track estimation is an important part of any tracking system, and therefore a paragraph in Sec. IV briefly addresses these questions.

A. Test data

Test data set I [see Figs. 1(a)–1(d)] was obtained by a Simrad EK60 split-beam echosounder with a Simrad 38DD, a 38-kHz, 7°-beamwidth circular transducer. This transducer is depth stabilized and certified to be used down to 1500 m depth. The test set is taken looking horizontally into a herring (*Clupea harengus*) layer. The reason for aligning the sounder horizontally was originally to investigate the horizontal-aspect TS for herring, but it is also the preferred orientation when observing interindividual fish behavior, since there are indications that the fish will orientate relative to the neighboring fish horizontally more so than vertically

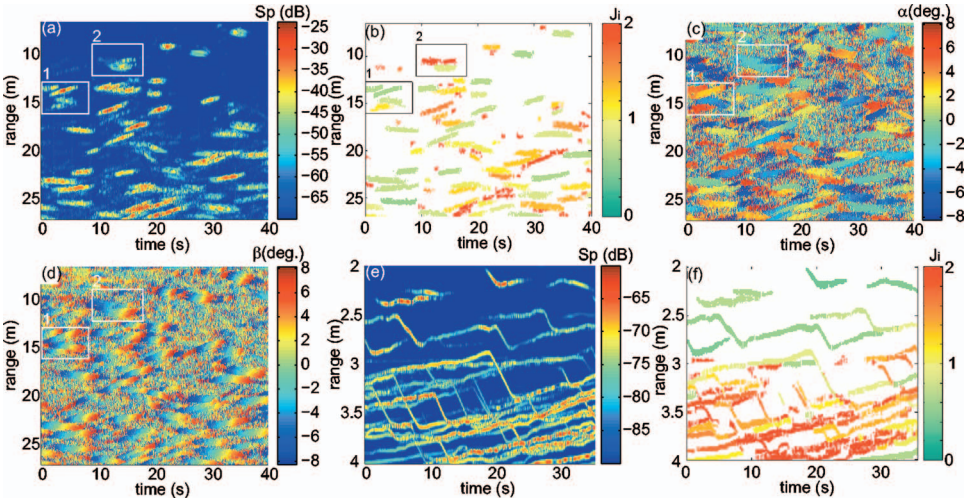


FIG. 1. (a) The example echogram for data set I. The echogram is given in S_p (dB re m^2) units. The boxes labeled 1 and 2 are the subsets 1 and 2, which are used as examples of successful and unsuccessful tracking, see Figs. 4 and 5, respectively. (b) The track quality for each track, J_i, where red indicates rejected tracks ($J_i > 1$) and green indicates accepted tracks ($J_i \leq 1$). Tracks with $J_i > 2$ are shown as $J_i = 2$. (c), (d) The corresponding alongship and athwart ship angles, respectively, in degrees. (e) The echogram for data set II and (f) the resulting quality of the tracks in test data set II, J_i.

(Grünbaum)[2]. The range of the probing echosounder was set to 27 m allowing for a very high ping rate of 17 Hz. In addition the sample interval was set to 9.4 cm with a pulse duration time τ of 512 μs. The output power P_t was set to 200 W. All data recorded closer than 6 m were discarded due to near field effects. The received power was converted to $S_p = 10\log(s_p)$ with unit dB re m^2, where

$$s_p = \frac{P_{r,i} 16\pi^2}{P_t g_0^2 \lambda^2} r_i^4 10^{2\alpha r_i}.$$

Here P_t is the output power, g_0 is the on axis gain, λ is the wavelength, α is the absorption coefficient, $P_{r,i}$ is the received power in sample i, and r_i is the corresponding range. In fisheries acoustic terms, this corresponds to time varied gain (TVG) of 40 log r and no beam pattern compensation.

Test data set II [see Figs. 1(e) and 1(f)] was obtained by a Simrad EK60 Split-beam echosounder with a Simrad ES 200-7C, a 200-kHz, 7°-beamwidth circular transducer, looking vertically into a cylindrical container of black polyethylene sheeting. The aim of this experiment was to investigate feeding behavior of herring larvae in a wide and physically controlled volume of sea water (a mesocosm). The bag was supported by ropes attached to a stainless steel circular ring connected to an open raft in a sheltered bay. The bag and the raft constituted a single entity floating on the sea surface. The pulse duration time τ was 64 μs, with a corresponding sample interval of 1.2 cm, and a ping rate of 16.8 Hz. The output power P_t was set to 1000 W.

In order to compare the present method to the original SED algorithm, single targets are detected in test data set I by the SED algorithm incorporated in the EK60 and by the SONAR 5 software. The echo-length acceptance criteria are set wide and the phase deviations high in order to reject less targets (Table I).

B. The method

Since the sample data are being used directly, no prior detection is necessary. The idea is that samples are treated as targets and that several samples from the same ping may be associated with the same single target, somewhat similar to the concept of joint probabilistic data association (Blackman and Popoli, 1999, pp. 353–355), but probabilities of detection are not considered and multiple samples may not be shared by two tracks.

A tracking system consists of several steps, including track estimation, track prediction, gating, data association, and track support. Gating, data association, and track support are presented in detail in the following, but track estimation and track prediction are only summarized. The details are presented in Handegard *et al.* (2005).

Prior to tracking, a threshold is set to initially remove samples not being targets. This threshold is set low to ensure that weak targets are not missed in the subsequent algorithm. The thresholds for data sets I and II are −70 and −90 dB, respectively, and are set based on visual inspection of the echograms [Figs. 1(a) and 1(e)] with different thresholds.

Each sample above the threshold is treated as an observation, and consists of $\mathbf{y} = [\alpha\ \beta\ r\ I]$, where α and β are the alongship and athwartship angles, respectively, r is the range, and I is the energy of the sample. The intensity measure used here is S_p.

TABLE I. The EK60 and SONAR 5 single-echo-detection (SED) settings for the case I test set. The parameters are chosen to allow more detections than the typical settings. The "echo lengths" are given as a factor multiplied with the pulse duration time τ. Maximum phase deviation is the maximum allowed average electrical phase jitter between samples inside an echo from a single target. For the echosounder and transducer used here (case I), one phase step corresponds to 0.064°. The recommended setting for weak targets is four to ten phase steps. Maximum gain compensation is the correction value from the one-way model of the transducer beam pattern. For the Simrad 38DD transducer, this corresponds to a maximum acceptable off-axis angle of ~5°. The threshold is applied to the echogram in S_p (dB re m^2) units.

	EK60	SONAR 5
Software version	1.4.4.66	v5.9.6
Minumum echo length (s)	0.2τ	0.2τ
Maximum echo length (s)	2.7τ	2.7τ
Maximum phase deviation (phase steps)	10.0	23.0
Maximum gain compensation (dB)	6.0	6.0
Threshold (dB)	−50	−50
Multiple peak suppression	N/A	Off
Min distance between detections (cm)	N/A	1

TABLE II. The tracking parameters for data set I and data set II: The threshold (TH), the gate parameters (α_g, β_g, r_g, and I_g), the track initiation parameters (α_0, β_0, r_0, I_0, and N_0), the track termination parameters, including the maximum within gate conflicts (N_c), the maximum number of missing samples between the first and last sample in range (N_e), the maximum number of successive missing pings (N_m), and the track rejection parameters, including the missing pings to track-length ratio (NM), the track length (TL), and the samples to track length ratio (NL).

	TH (dB)	α_G β_G (deg)	r_G (m)	I_G (dB)	α_0 β_0 (deg)	r_0 (m)	I_0 (dB)	N_0	N_c	N_e	N_m	MN	TL	NL
Data set I	−70	2.80	0.44	20	2.80	0.12	20	5	3	2	1	0.80	8	2.00
Data set II	−90	1.80	0.10	20	2.00	0.03	20	5	2	2	1	0.80	8	2.00

At each time step, the track state $\mathbf{x}$ (position, velocity, and echo intensity) for each live track is estimated using a Kalman filter, and assuming constant velocity, the state is predicted at the next time step [Handegard *et al.*, 2005, their Eq. (5)]. The predicted state is denoted $\tilde{\mathbf{x}}$.

To compare predictions with observations, the predicted state $\tilde{\mathbf{x}}$ is mapped to observation space, i.e., $\tilde{\mathbf{x}} \rightarrow \tilde{\mathbf{y}}$. For the position, this involves mapping the Cartesian position of the track to alongship angle, athwartship angle, and range [Handegard *et al.*, 2005, their Eq. (4)]. The velocity is not part of the mapping, and the echo intensity is the same in both spaces. This allows us to define a distance metric between a sample $\mathbf{y}$ and a live track $\tilde{\mathbf{y}}$ (predicted position in observation space). The metric is the so-called gate distance defined in the following.

1. Gating

The first part of the data association is the gating. This decides which samples should be considered to be parts of the track. In the following, i, j, and k denote track number, sample number, and ping number, respectively. The difference between the prediction from track i and sample j is calculated for all predictions and samples at ping k, i.e.,

$$\hat{\boldsymbol{\epsilon}}_{ijk} = \mathbf{y}_{jk} - \tilde{\mathbf{y}}_{ik}. \tag{1}$$

The gate distance

$$d_{ijk} = \hat{\boldsymbol{\epsilon}}_{ijk} G \hat{\boldsymbol{\epsilon}}_{ijk}^{T} \tag{2}$$

is a measure of closeness between samples and predictions. If $d_{ijk} \leqslant 1$, sample j is inside the gate of track i at ping k. Here T is matrix transpose and

$$G = \begin{bmatrix} \alpha_G^2 & 0 & 0 & 0 \\ 0 & \beta_G^2 & 0 & 0 \\ 0 & 0 & r_G^2 & 0 \\ 0 & 0 & 0 & I_G^2 \end{bmatrix}^{-1}. \tag{3}$$

Note that the intensity is also included in the gate. The elements in G are set as parameters in this case, α_G, β_G, r_G, and I_G being the maximum allowed deviances between the observation and the prediction along each dimension (see Table II). This means that if the maximum deviation occurs in α, no deviance is allowed in any of the other dimensions. Consequently, the maximum deviance rarely occurs. The interpretation is that the observation must be within a hyperellipsoid defined by $d \leqslant 1$ (see Fig. 2). There are ways to set the gate parameters based on detection probabilities and innovation covariances, i.e., the covariance of $\hat{\boldsymbol{\epsilon}}$, but this is not done here, cf. the discussion to follow.

2. Data association

The next step is to associate the samples inside the gates to tracks. At each ping a gate is calculated for each live track, but a given sample may be inside more than one gate, and a given gate may contain more than one sample. If there are no conflicts, i.e., no samples lie within more than one gate, each sample is associated to the corresponding track. If there are conflicting observations, they are associated to the "closest" track in terms of d. However, if the number of conflicting observations inside a gate is higher than the *ad hoc* parameter N_c, the shortest track is automatically terminated. In general the N_c parameter is a crude method to avoid wrong associations, and tuning the gates are preferred over decreasing N_c. The result of the data association step is that each sample j at ping k that are deemed part of a track i is associated with track number i. The set of samples j associated with track i at ping k is denoted A_{ik}. Similarly, the set A_i is defined as the set of all samples associated with track i over the full duration of the track.

The observations j that are associated to track i at ping k are then combined into a composite observation by

$$\mathbf{y}'_{ik} = \frac{\sum_{j \in A_{ik}} w_{ijk} \mathbf{y}_{jk}}{\sum_{j \in A_{ik}} w_{ijk}}, \tag{4}$$

where $w_{ijk} = \exp(-d_{ijk}^2)$. This weighted observation is used in the Kalman update equation. This approach is analogous to joint probabilistic data association, where the weights are

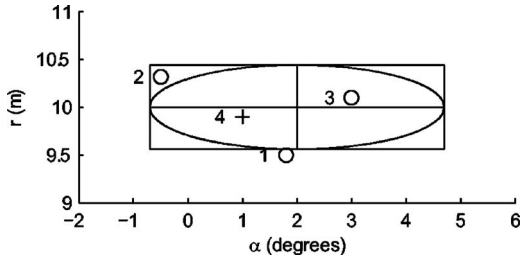


FIG. 2. Two-dimensional projection of the four-dimensional gating process; the center of the ellipsoid is the prediction by the Kalman filter algorithm. Four samples are given for illustrative purposes. Sample 1 is outside the range gate r_G, sample 2 is inside the range of r_G and α_G separately (illustrated by the rectangle), but outside the ellipsoidal gate since the deviations are summed. Sample 3 is inside the gate of r_G and α_G (illustrated by the ellipse), but the gate is four dimensional, and this sample lies outside the ellipsoidal gate, because of a large offset in β. Finally, sample 4 lies within the ellipsoidal gate and is associated with the prediction.

determined by closeness to the prediction, but detection probabilities are not considered here. The updated state variables are used to obtain predictions for the next time step, and then the process is repeated.

3. Initiating and terminating tracks

Before proceeding to the next time step, samples not associated with tracks are used as candidates for new tracks. Initiating new tracks is difficult since there are usually several sample values for each new target. The method chosen for initiating new tracks is similar to the tracking of already established tracks. The nonassociated samples within a ping are compared in pairs (each sample is paired with the next adjacent sample). By taking the first sample as a prediction and the second as an observation, Eqs. (1) and (2) can be used to calculate the distance between them, using the track-initiation gate

$$G_0 = \begin{bmatrix} \alpha_0^2 & 0 & 0 & 0 \\ 0 & \beta_0^2 & 0 & 0 \\ 0 & 0 & r_0^2 & 0 \\ 0 & 0 & 0 & I_0^2 \end{bmatrix}^{-1} \tag{5}$$

instead of the regular gate. If the difference between these samples is less than unity, i.e., the second is within the gate of the first, the second sample is compared with the next adjacent sample is range, etc., forming a chain of samples that are within each other's gates. When a sample does not fall within the gate of the previous, the chain is determined. If the length of the chain is larger than N_0, a new track is formed based on the mean of these samples. If many adjacent samples (in range) meet the criteria, there is a possibility for combining several new targets into one, and this will occasionally happen. Several tools have been developed to remove or to flag these tracks as low-quality results, cf. Sec. II C.

Track termination is the method applied to close a track during tracking. Three methods for track termination are implemented. The gate-conflict parameter N_c has already been defined as the maximum number of conflicting targets within a gate. Furthermore, when tracking very weak targets, the background reverberation may be tracked. These "tracks," however, typically have several missing samples within them. To terminate these tracks, the number of missing samples between the first and last sample (in range) are counted, and if they exceed N_e samples, the track is terminated. Finally, the number of successive missing pings, N_m, is used. However, after a track is terminated, other techniques can be used to test its quality or to reject it, cf. Sec. II C.

C. Quality control

When tracking fish in dense registrations, misassociations and track-split errors will inevitably occur, and a method to reject or to flag each track with an association-quality measure is important. The track quality algorithm is based on measures of association errors, and the track rejection algorithms are based on more *ad hoc* criteria. The criteria adopted for rejecting tracks are simply the number of missing pings to the track length ratio (MN) the track length (TL), and the total number of samples to the track length ratio (NL). These are set as parameters.

Previously, association errors have been investigated for single targets, i.e., false or missing associations between single detections (Handegard *et al.*, 2005, see, e.g., their Sec. III D). However, that measure cannot be used here because target detection and tracking are not separated. Since the algorithm presented here works on samples, misassociation in range may also occur, both as connection errors (one track may consist of several fish) and track-split errors (several adjacent tracks may be formed from one single fish). Four measures of track quality have been defined. The first two measures take a global approach where no particular misassociation type is addressed, whereas the two last measures deal with track connection and track-split errors in range, respectively.

One way to investigate association errors is to compare the results from a forward and backward run through the data set. If questionable associations have occurred, it is likely that a different result would be obtained by running the data association backwards, i.e., starting from the end of the data set and progressing to the beginning. For a given track in the primary data set, e.g., the forward run, the samples associated with track i are given by A_i. Let B_l be the set of samples associated with track l from the backward run. The intersection is given by $A_i \cap B_l$, and let N_{A_i}, N_{B_l}, $N_{A_i \cap B_l}$ be the number of samples in the respective sets. Two measures comparing the backward and forward runs are defined. Let

$$J_{a,i} = 1 - \max_l (N_{A_i \cap B_l})/N_{A_i}. \tag{6}$$

This results in a measure of how well the "best" backward track overlaps the forward track, and the identifier l for that track. The range of J_a is [0 1], where a high value indicates that the forward track contains false associations or the backward track is split. Let

$$J_{b,i} = 1 - (N_{A_i \cap B_l})/N_{B_l}, \tag{7}$$

where l is the backward track that maximized $N_{A_i \cap B_l}$ in Eq. (6). Again the range is [0 1], and a high J_b value indicates the occurrence of track-split errors in the forward track or false association in the backward track. Since the index l is taken from Eq. (6), J_a do not necessarily become J_b when using the backward run as the forward run. This procedure is illustrated in Sec. III on a subsample of the test data set I, cf. Fig. 5.

The track initialization algorithm may be stricter than that used to continue an already initiated track; one result of this could be a late initiation of tracks. By combining high-quality-forward and backward tracks in terms of J_a and J_b, the dependence on track initiation parameters may be reduced. The new track is defined by the union $A_i \cup B_l$ of the matched tracks. If weak targets along the edges of the tracks are a concern, the intersection between the forward and backward track, $A_i \cap B_l$, can be used instead.

A track-split error may occur if the tracks are initiated incorrectly, i.e., when two tracks are initiated for one fish, or

if the gates from two tracks cover the same samples. This can be observed in the echogram as two adjacent tracks, where the tracks are not separated by any nonassociated samples. A measure to detect this effect is simply to check the sample above and below the track. If that sample belongs to another track, a track-split error may have occurred, i.e.,

$$J_{\text{split},ik} = \frac{N_a}{2}, \tag{8}$$

where $N_a \in \{0,1,2\}$ if there is, or not, an adjacent track above, below, or both at ping k for track i. The factor $1/2$ is for scaling the measure to the range [0 1]. For the whole track

$$J_{\text{split},i} = \frac{1}{L_i}\sum_{k}^{L_i} J_{\text{split},ik}, \tag{9}$$

where L_i is the length of the track (excluding missing pings).

A track merging error in range occurs when one track is formed from two fish, e.g., if the initiation and tracking gates are set too wide. It is desirable to set the gates wide to capture rapid changes in behavior, but at the possible cost of misassociation, both in range and time. However, using standard SED rejection criteria, misassociations in range can be monitored. Following the recommendations of Soule *et al.* (1996) for successful multiple-target rejection in SED algorithms, the phase angle deviation over one target in one ping is calculated as a measure of track-combination error in range, i.e.,

$$J_{\text{merge},ik} = \frac{1}{2}\left[\frac{1}{N_{A_{ik}}-1}\sum_{j\in A_{ik}}(\bar{\alpha}_{ik}-\alpha_{jk})^2\right]^{1/2} + \frac{1}{2}\left[\frac{1}{N_{A_{ik}}-1}\sum_{j\in A_{ik}}(\bar{\beta}_{ik}-\beta_{jk})^2\right]^{1/2}, \tag{10}$$

where $N_{A_{ik}}$ is the number of samples for track i at ping k and $\bar{\alpha}_{ik}$ and $\bar{\beta}_{ik}$ are the mean athwartship and alongship angles for track i and ping k, respectively. To use this measure for each track i,

$$J_{\text{merge},i} = \max_k (J_{\text{merge},ik}) \tag{11}$$

is defined. The maximum value is chosen since one erroneous association can severely bias the velocity estimate for the whole track. For applications where false associations are of less concern, the mean value could be used instead.

Four measures of quality control have been described. Each of these can be monitored individually, or they can be scaled relative to each other like

$$J_i = \left\{\left(\frac{J_{\text{a},i}}{J_{\text{a},0}}\right)^2 + \left(\frac{J_{\text{b},i}}{J_{\text{b},0}}\right)^2 + \left(\frac{J_{\text{merge},i}}{J_{\text{merge},0}}\right)^2 + \left(\frac{J_{\text{split},i}}{J_{\text{split},0}}\right)^2\right\}^{1/2} \leqslant 1, \tag{12}$$

where $J_{\text{merge},0}=1°$, $J_{\text{split},0}=0.3$, $J_{a,0}=0.3$, and $J_{b,0}=0.3$ are acceptable values for the different types of errors. This is, however, dependent on the application and is here implemented as changeable parameters.

Finally, the impact of various parameter settings on $J=(1/N)\Sigma_i^N J_i$, where N is the number of tracks, is investigated for both test data sets. The sensitivity measure is defined as

$$S_{pa} = 0.5\left(\frac{|\Delta J_{+10\%}|}{J} + \frac{|\Delta J_{-10\%}|}{J}\right)\left(\frac{|\Delta x|}{x}\right)^{-1}, \tag{13}$$

where $\Delta J_{+10\%}$ and $\Delta J_{-10\%}$ are the changes in J when perturbing parameter $pa \pm 10\%$, and $|\Delta x|/x=0.1$, except for the parameters that are integers. To test the sensitivity to integer parameters, the parameter value is increased or decreased by one and $|\Delta x|/x=1/N_0$, where N_0 is the unperturbed value. A similar method was used to test the sensitivity to data association in Handegard *et al.* [2005, their Eq. (15)].

III. RESULTS

The performance of the algorithm is demonstrated by its ability to associate samples in the test-data sets to tracks. First the results from the full data sets are presented (Fig. 1). Then subset 1 from data set I is used to demonstrate the ability to track closely spaced targets as compared to a traditional SED algorithm. Finally, subset 2 from data set I is presented as an example where the tracking fails. Here the importance of the track quality algorithm is shown.

The ability of the method to associate samples into individual tracks is presented in Figs. 1(b) and 1(f). The parameters for the association method are specified in Table II, and the general parameters used for the tracking, i.e., to obtain the prediction, etc., are similar to those in Handegard *et al.* (2005, their Table V, case I).

By inspecting the intensity echogram of test-data set I [Fig. 1(a)], distinctive tracks can be distinguished by eye. When simultaneously looking at the anglegrams [Figs. 1(c) and 1(d)], the tracks are more clearly separable. Actually, the angles are remarkably stable over a track, even if the echo intensity for the track is low. When visually comparing the echograms and anglegrams with the classification image [Fig. 1(b)], it is seen that the tracks are well detected. The distinctive tracks are classified as good tracks, whereas dubious registrations are marked as low-quality tracks. The same can be seen in the results for test-data set II [Fig. 1(e)].

The tracking algorithm can be used to separate background reverberation samples and track samples. Here, background reverberation is defined as the signal from a nontargeted scatterer. The empirical distribution for intensity and angles for both background reverberation and signal samples are presented [Figs. 3(a) and 3(b), respectively]. There is an overlap between the signal and background reverberation distributions in intensity. This may be caused by the failure to detect the beginning of some of the low-quality tracks, which are thus being classified as background reverberation. Similarly, some background reverberation samples will be falsely associated with tracks. This is defined as association errors, see the following. When looking at the angle distributions, a typical "circular" distribution is seen for the targets [Fig. 3(b)]. If the targets were uniformly distributed within the beam with r degrees opening angle, the probability density along one angle, e.g., α, times the number of

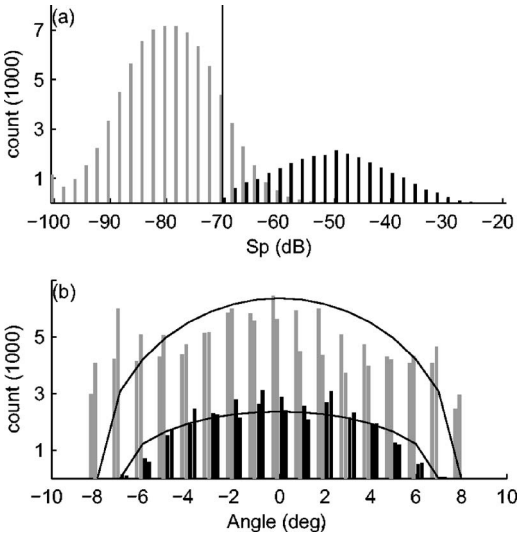


FIG. 3. The signal and background reverberation distributions. Signals are defined to be samples that belong to a track [see Fig. 1(b)], and background reverberation is defined to be samples more than two pixels away from a defined track. (a) The intensity distribution in S_p units (dB re m^2), where light gray bars are from background reverberation samples and black bars are from signal samples. The vertical line is the −70 dB threshold. (b) The angle distribution, where the light gray bars are background reverberation and black bars are signal. The first and second adjacent similarly colored bars are the alongship and athwart ship angles, respectively. The lines are the resulting angle distributions given uniformly distributed targets within $\alpha^2+\beta^2 \leq r^2$.

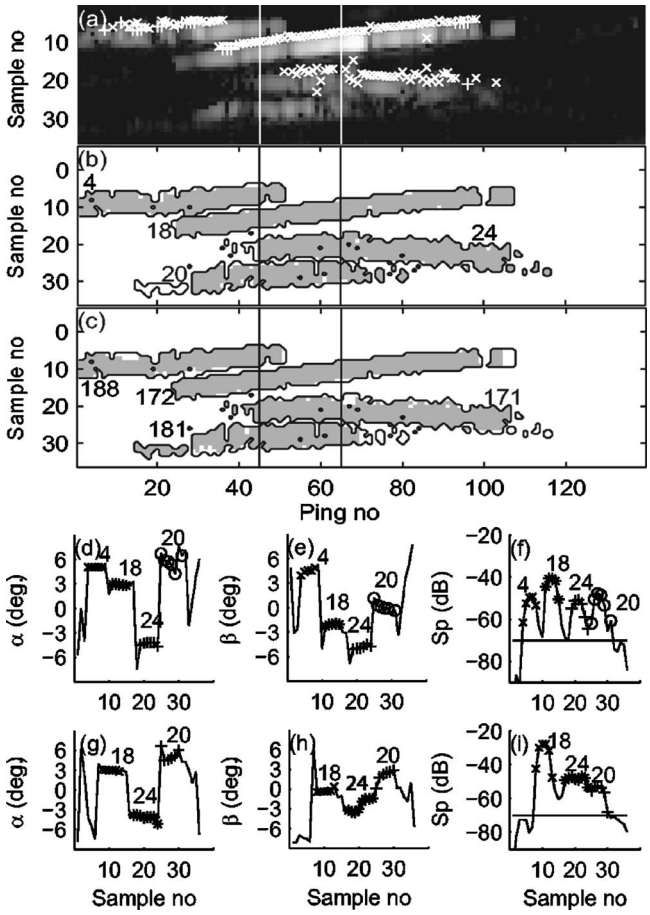


FIG. 4. Successful tracking of closely spaced targets. (a) The intensity echogram, where the white "+" and "×" denote the SED targets from EK60 and SONAR 5, respectively. The white vertical lines are the examples shown in (d)–(f) and (g)–(i), respectively. (b), (c) The forward and backward data association, respectively. Numbers indicate track numbers. The outlines denote the combination of the forward and backward tracking. Panels (d)–(f) and (g)–(i) show the intensity and angles for the samples indicated by vertical lines in (a)–(c). (×), (*), (+), and (○); Samples that have been associated with different tracks. The track numbers are printed above.

samples would be given by $y \sim N\sqrt{1-\alpha^2/r^2}/2\pi$, where N is the total number of samples. This seems to fit the results well, indicating that the targets ($r=7°$) were indeed randomly positioned across the beam. In the case of the background reverberation samples, the circular distribution is a less good fit (using $r=8°$). This distribution seems to be a combination of a uniform and a circular distribution. The uniform part is analogous to the model assumption in Kieser *et al.* [2000, their Eq. (4)]. The circular distribution component may originate from nondetected tracks or other nonfish scatters in the water column.

A subset of test-data set I, where closely spaced data are successfully tracked, is used to compare the ability of the tracker against the traditional SED, and to present the combination of forward and backward tracks (Fig. 4). The algorithm detects four tracks, but the EK60 SED does not accept many targets in this case. This may be beneficial when estimating TS, but is not necessarily an advantage for tracking purposes. SONAR 5 accepts more targets, but may also include more false targets. This is not so crucial, since the tracking algorithm would reject the false targets. However, both methods present several missing detections along a track. To illustrate the information used by the SED algorithm, the data contained in two separate pings are presented [See Figs. 4(d)–4(f) and Figs. 4(g)–4(i)]. To separate the targets, the conventional SED initially attempts to identify the peaks based only on the echo intensities. This may be possible in the first example [Fig. 4(f)] but is clearly not possible in the second example [Fig. 4(i)]. Success would also depend on the SED settings (Table I). If the threshold is set too low, several targets will be within the window and are thus rejected. However, when taking the angles into account, the tracks are separable. Tracks 20 and 24 are clearly separated by the alongship angles [Fig. 4(g)]. Without taking the angles into consideration, it would not be possible to properly track these signals. This is also demonstrated by the complete failure of both SED algorithms to detect track 20 [Fig. 4(a)]. The combination of forward and backward tracks are presented in Figs. 4(b) and 4(c). If the tracks meet the criteria of overlap i.e., $J_a < 0.25$ and $J_b < 0.25$ in this case, the forward and backward runs are combined into a single merged track. The tracks in this subsample meet the stated criteria, and are successfully merged [Fig. 4(b)]. Note that the track numbers are different in the forward and backward case.

When the density is increasing, the limitations inherent in the split-beam principle will cause any tracking algorithm to fail. The track quality algorithm is thus an important part of the tracker, as it flags the cases where the tracker fails. Subset 2 of test data set I, where the algorithm fails, is used as an example (Fig. 5). Note that it is also difficult to determine tracks by visually inspecting the echogram and anglegrams [Figs. 1(a), 1(c), and 1(d), inside second box]. Since

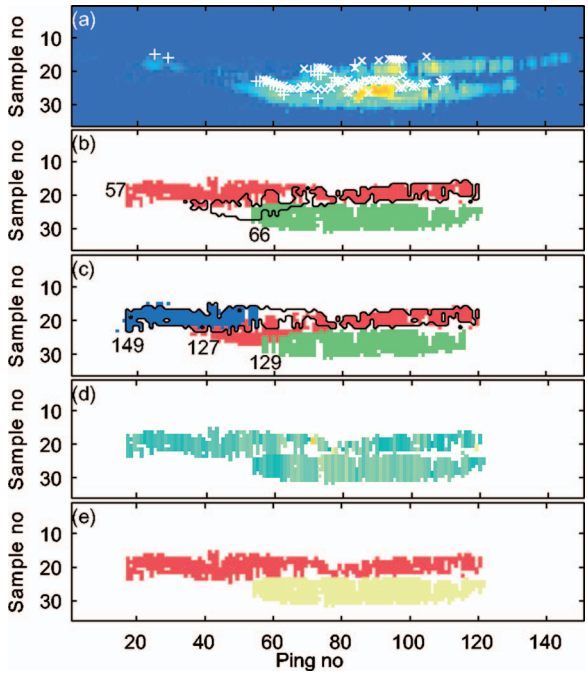


FIG. 5. Example of unsuccessful tracking. (a) The intensity echogram. The color scale is similar to Fig. 1(a). (b), (c) The results from the forward and backward tracking, respectively, on which J_a and J_b are based. Different colors indicate different tracks. Track 57 in the forward run (b) is taken as an example. The boundary from this track is transported to (c), i.e., the result from the backward run. It is seen that backward track 127 covers the largest proportion of this boundary, i.e., 46%. The J_a value is then $J_a=1-0.46=0.54$. Then the boundary of the backward track 127 is transported to the forward case panel (b). Track 57 covers 66% of this area, and thus $J_b=1-0.66=0.34$. (d) $J_{\text{merge},ik}/J_{\text{merge},0}$ with the color-scale range from [0 2], similar to Figs. 1(b) and 1(f). (e) The resulting quality of the tracks, with the color scale range from [0 2], similar to (d).

the forward and backward tracking do not agree [Figs. 5(b) and 5(c)], and the merge and track-split errors are high (Table III), the algorithm marks the tracks as low quality. Further, the SED targets of the two algorithms are inconsistent in this case, and tracking these would yield doubtful results. The classification is done for both test-data sets [Figs. 1(b) and 1(f), Table III].

Sensitivity. The sensitivity test investigates the impact of

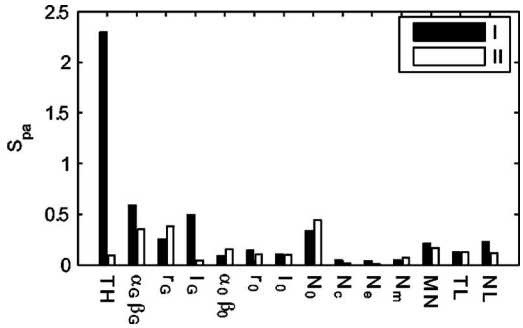


FIG. 6. The sensitivity S_{pa} for the different parameters, for data sets I and II, respectively.

parameter values on the track quality, giving an indication of which parameters are most important in the algorithm. The sensitivity for both test-data sets is given in Fig. 6. In general, the threshold, TH, is most important, followed by the gate and track initiation parameters. The lowest sensitivity is found in the track-termination parameters (N_c, N_e, and N_m). The sensitivity to the track-rejection parameters (MN, TL, and NL) is low, but this result depends on the initial parameter setting. If they are set to reject a large portion of the tracks, the rejection parameters become more important. Consequently, this test should be interpreted with care.

IV. DISCUSSION

The failure of the SED algorithm when tracking inside dense fish registrations was the motivation for this work, but there are other reports of problems with the SED procedure. Cronkite *et al.* (2004, their Sec. 4.3) discuss the problem of detecting closely spaced targets. They argue that the poor detection of closely spaced targets is caused by the echosounder SED algorithm, rather than acoustic shadowing within the aggregation. The inadequacy of the SED algorithm is thus not unique to the test cases reported here. In addition, the focus in fisheries acoustics has shifted from addressing technical problems like calibration, etc., to a more biological perspective, addressing behavioral considerations. This shift is evident in the Proceedings of the International

TABLE III. The first column indicates the test set I or II, and, where applicable, subset 1 or 2. The subsets refer to the boxes in Fig. 1(a), and Figs. 5 and 4, respectively. Where no track number is given, the mean of the respective data is presented, i.e., I,1 indicates the mean values from the tracks in subset 1 in test set I. I and II indicate the mean values of the full test sets. TrNoF and TrNoB indicate track numbers in the forward and backward cases, respectively, L_i is the track length (or mean length for the mean values) for the forward case, J_i (or J for the mean values), J_a, J_b, J_{merge}, and J_{split} are the quality measures defined in Sec. II C.

	TrNoF	TrNoB	L_i	J_i	J_a	J_b	J_{merge}	J_{split}
I,1	4	188	51	0.672	0.120	0.079	0.014	0.029
I,1	18	172	83	0.638	0.080	0.026	0.015	0.018
I,1	20	181	56	1.183	0.191	0.194	0.023	0.052
I,1	24	171	72	0.784	0.123	0.078	0.017	0.045
I,1			66	0.819	0.129	0.094	0.017	0.036
I,2	57	127	104	2.247	0.535	0.340	0.034	0.063
I,2	66	129	68	0.901	0.182	0.060	0.017	0.096
I,2			86	1.574	0.359	0.200	0.026	0.079
I			51	1.419	0.237	0.244	0.020	0.065
II			60	1.858	0.312	0.401	0.016	0.189

Council for Exploration of the Sea (www.ices.dk), which has sponsored symposia on fisheries acoustics (Craig, 1984; Karp, 1990; MacLennan *et al.*, 2003; Margetts, 1977; Massé *et al.*, 2003; Simmonds, 1996). Consequently, an improved methodology for addressing these questions is valuable.

Different applications make different demands on the tracking and detection algorithms. The SED algorithm was originally designed to estimate the TS of species of interest, and achieves this by rejecting low-quality targets (Soule *et al.*, 1996). By accepting only the high-quality-single targets, many pings may not register along the track in the subsequent tracking. This adversely results in split tracks and position errors, which in turn make the estimation of target velocity and position difficult. Apart from that, there are other problems with rejecting weak and low-quality targets before tracking. For example, estimates of target strength are improved when there are samples over several pings from the same fish (Ehrenberg and Torkelson, 1996; Foote *et al.*, 1984), and tracking can be used to obtain behavioral observations relevant to the target strength estimation (Chu *et al.*, 2003; McQuinn and Winger, 2003). Consequently, improving our ability to track single individuals may lead to better TS estimates, and if bias in intensity is a concern, the quality could be controlled by track quality parameters like the J_{merge}, cf. the following discussion of track quality. Another effect of the traditional approach is that weak and low-quality targets are undetected, which is a problem when addressing the interindividual behavior issues mentioned in Sec. I. Here it is crucial that low-quality tracks are detected, to avoid false nearest-neighbor pairing. This is also important when investigating the spatial distribution from single-target observations, as in, e.g., Pedersen (1996); Trenkel *et al.* (2004). Consequently, the method presented here improves the feasibility of experiments that rely on our ability to resolve single targets, including TS estimation.

Target tracking is a large field with an extensive literature, and several methods have been described, see, e.g., Blackman and Popoli (1999) as a general reference. Other data-association methods include multiple-hypothesis tracking. This approach delays the decision on data association for several pings, and keeps track of the most likely target associations. This results in several combinations (hypotheses) of the targets. This is the recommended data-association method according to Blackman and Popoli (1999, Chap. 6). However, keeping track of several track-combination hypotheses when working with echo samples, as opposed to single targets, would yield a vast number of combinations. It is possible that the method could be refined to fit this approach, but the complexity would increase. Other approaches includes particle filtering, e.g., the probability-hypothesis-density (PHD) method. The PHD filter has been implemented for forward-scan sonar images (Clark and Bell, 2005), and one advantage of this approach is its ability to filter signals in high clutter, reducing the number of spurious measurements. However, data association, the main consideration in this paper, is not presently addressed in the PHD method.

The novelty in the present work is the track-before-detect approach. This implies the use of angle samples in the detection, not only for rejection in the SED algorithms and in subsequent tracking. This is achieved by treating each sample above a threshold as a single target, and allowing for several associations of samples to each prediction (track). The approach is simple, but some further consideration of why it works may be appropriate. Humans have a remarkable capability to extract information from images, and that is one of the reasons that tracking targets in intensity echograms by eye often performs better than automatic tracking. However, when we include both anglegrams and existing track predictions in addition to the intensity echogram, our ability is less superior because we have to interpret the information in several images simultaneously. The computer has no problem in handling this multidimensionality.

A. Parameter sensitivity and track quality

The method is very flexible, and calls for caution when setting the parameters. In general, one should start with the most sensitive parameters and progress to the less sensitive parameters. The threshold should initially be set low to not miss any targets, and should subsequently be increased if necessary. The N_n parameter indicating the number of samples needed to initiate a track should be approximately the number of samples per ping from each target (5 in test data I). This is dependent on the sampling interval, the pulse length, and the target size range. The track-initiation gate range parameter, r_0, should be set slightly higher than the resolution in range (9.4 and 1.2 cm, for test data I and II, respectively). The track gate, G, depends on the fish behavior, the background reverberation, and the density. The track-termination and track-rejection parameters should initially be set high and then decreased, if necessary. After the initial settings are decided, a subsample of the data should be used for tuning the parameters. The parameters, G and G_0 should be tuned first. A larger initiation gate starts the tracks earlier, but potentially gives more false targets. If the track-initiation gate is narrowed too much, track-split errors in range may occur. If the track gate is increased, more rapid changes in behavior may be detected but several targets may be erroneously associated within one track, both in time and range. If the track gate is decreased, tracks may split and rapid changes in behavior will be undetected, but there will be less false associations. The track-termination and track-rejection parameters should be set to terminate and reject more tracks if the gate tuning does not improve the performance.

The performance of the tracker depends on the parameter values, and the above-outlined setting procedure involves subjectivity. An important aspect of the track-quality algorithm is to remove some of this subjectivity. As mentioned earlier, the definition of high or low quality may depend on the application, and these trade-offs can be achieved by different weights on the various quality measures. Further, a low-quality track may convert to a high-quality track by tuning the tracking parameters correctly, while the definition of track quality remains the same. Consequently, the track quality indicates the comparative success of the different parameter settings, but is also an absolute (as opposed to rela-

tive) measure of the track quality. This distinction is important because it gives us the possibility to assess whether tracking is feasible or not in a given situation, beyond the information on how well the parameters are set.

Although track quality measures have been established, the ultimate test is to visually compare the automatic classification into tracks with the appearance of the echogram and the anglegrams. The value of carefully inspecting these should not be underestimated, because humans are highly skilled at extracting information from images. Some researchers argue that manual tracking is better and more accurate than automatic tracking. Although this may be true in some cases, the manual approach is less consistent. Manual tracking may yield different results depending on who scrutinizes the data. Algorithms with given parameter settings do not have this problem.

B. Track estimation

An important part of any tracking system is the track-estimation process. Here, track estimation is defined as the information one can obtain from the samples comprising a single track, e.g., the estimated positions, the corresponding velocities, TS, etc. For example, if swimming velocities are required, simply dividing the distance between target fixes by the time difference is not good enough, because the observation errors are ignored, cf. Mulligan and Chen (2000) for a convincing illustration. Although this paper focuses on data association, a short discussion of different track estimation techniques is included.

The Kalman filter is a common method for estimating speed and positions in tracking applications. This is a powerful tool when real-time tracking is required, e.g., in the data association process when predictions are needed or when tracking airplanes or missiles. When estimating fish swimming trajectories, however, we can estimate the tracks after association. This enables us to use a whole suite of estimation techniques. Several methods have been tested on split-beam data in Handegard *et al.* (2005). Both smoothing splines and Kalman smoothing (Kalman smoothing traverse the Kalman filter estimates backwards to reduce dependence of the initial conditions) were found to work well for position estimates, but the Kalman smoother failed to estimate the velocities. In the test cases presented here, however, the Kalman smoother produced sensible results. If the process and measurement errors are well known, the Kalman smoother is an appropriate tool since it separates the process and observation errors, i.e., it reduces the number of *ad hoc* parameters. Another approach is to use smoothing splines where the smoothing is set by cross validation. For the test data in this paper, the cross-validation method detected periodicity in the position estimates. This has the potential to give grossly overestimated swimming-speed estimates if not corrected for. The importance of plotting the tracks to assess this effect should not be underestimated. However, the periodicity is believed to be caused by the tail-beat frequency of individual fish, and this could potentially be associated with the swimming speed (Bainbridge, 1960).

Another aspect, not addressed in this paper, is the quality of the angle measurements in each sample. This should be considered when estimating fish position and velocity. For example, low signal-to-noise ratio gives higher variability in the angles, but may also bias them toward the center of the beam (Kieser *et al.*, 2000). This is especially important when using sonars close to boundaries, as would apply in a riverine environment or with horizontal transmissions near the bottom or the water surface.

V. CONCLUSIONS

The main advantages of tracking samples directly without prior target detection are simplicity and efficiency. The process is simple, because no prior detection is necessary as the process of tracking and detection is done in one step. The efficiency is improved by the target angles being utilized along with the intensity and range data, as opposed to the conventional SED algorithm that works on intensity and range only. The benefit of this approach has been clearly demonstrated.

The combination of the improved data association and the quality control substantially improves our ability to investigate the behavior of closely spaced single targets, providing the means to learn more about the behavior of the individual animals within aquatic ecosystems.

ACKNOWLEDGMENTS

This work was carried out during a stay at the University of Washington and NOAA Alaska Fisheries Science Center, Seattle, funded by the Norwegian Research Counsel under the Leiv Erikson program. I would like to thank Egil Ona and Lucio Calise for the test data sets I and II, respectively, Karl Øystein Gjelland for processing the SONAR 5 data, and Vidar Hjellvik, Alex De Robertis, and Patrick H. Ressler for comments on the manuscript.

[1]I use the term "anglegram" to describe a ping-by-ping display of targets which are color-coded to indicate one of the bearing angles.
[2]Personal communication, based on camera observations of fish in a tank.

Aoki, I. (**1982**). "A simulation study on the schooling mechanism in fish," Bull. Japanese Soc. Sci. Fisheries **48**, 1081–1088.
Arrhenius, F., Benneheij, B. J., Rudstam, L. G., and Boisclair, D. (**2000**). "Can stationary bottom split-beam hydroacoustics be used to measure fish swimming speed in situ?," Fisheries Research **45**, 31–41.
Bainbridge, R. (**1960**). "Speed and stamina in three fish," J. Exp. Biol. **37**, 129–153.
Balk, H., and Lindem, T. (**2000**). "Improved fish detection in data from split-beam transducers," Aquatic Living Resources **13**, 297–303.
Blackman, S. S., and Popoli, R. (**1999**). *Design and Analysis of Modern Tracking Systems* (Artech House, Boston).
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.
Breder, C. M. J. (**1954**). "Equations descriptive of fish schools and other animal aggregations," Ecology **35**, 361–370.
Cech, M., and Kubecka, J. (**2002**). "Sinusoidal cycling swimming pattern of reservoir fishes," J. Fish Biol. **61**, 456–471.
Chu, D., Jech, J. M., and Lavery, A. (**2003**). "Inference of geometrical and behavioural parameters of individual fish from echo-trace-analysis," Deep-Sea Res., Part I **50**, 515–527.
Clark, D., and Bell, J. (**2005**). "Bayesian multiple target tracking in forward scan sonar images using the phd filter," IEE Proc., Radar Sonar Navig. **152**, 327–334.

Craig, R. ed. (**1984**). ICES Symposium on Fisheries Acoustics, Bergen, Norway, 1982, Rapp. P.-v. Reun.-Cons. Int. Explor. Mer **184**.
Cronkite, G. M., Enzenhofer, H. J., and Gray, A. P. (**2004**). "Split-beam sonar observations of targets as an aid in the interpretation of anomalies encountered while monitoring migrating adult salmon in rivers," Aquatic Living Resources **17**, 1–12.
De Robertis, A., Schell, C., and Jaffe, J. (**2003**). "Acoustic observations of the swimming behaviour of the euphasiid *Euphasia pacifica* Hansen," ICES J. Mar. Sci. **60**, 885–898.
Ehrenberg, J. E., and Torkelson, T. C. (**1996**). "Application of dual-beam and split-beam target tracking in fisheries acoustics," ICES J. Mar. Sci. **53**, 329–334.
Enzenhofer, H. J., Olsen, N., and Mulligan, T. J. (**1998**). "Fixed-location riverine hydroacoustics as a method of enumerating migrating adult pacific salmon: Comparison of split-beam acoustics vs. visual counting," Aquatic Living Resources **11**, 61–74.
Foote, K. G. (**1996**). "Coincidence echo statistics," J. Acoust. Soc. Am. **99**, 266–271.
Foote, K. G., Aglen, A., and Nakken, O. (**1986**). "Measurement of fish target strength with a split-beam echo sounder," J. Acoust. Soc. Am. **80**, 612–621.
Foote, K. G., Kristensen, F. H., and Solli, H. (**1984**). "Trial of a new, split-beam echo sounder," Coun. Meet. Int. Coun. Explor. Sea B:21, Copenhagen, Denmark.
Genin, A., Jeffe, J. S., Reef, R., Richter, C., and Franks, P. J. S. (**2005**). "Swimming against the flow: A mechanism of zooplankton aggregation," Science **308**, 860–862.
Gjelland, K. Y., Bøhn, T., Knudsen, F. R., and Amundsen, P.-A. (**2004**). "Influence of light on the swimming speed og coregonids in subarctic lakes," Ann. Zool. Fennici **41**, 137–146.
Handegard, N. O., Michalsen, K., and Tjøstheim, D. (**2003**). "Avoidance behaviour in cod (*gadus morhua*) to a bottom-trawling vessel," Aquatic Living Resources **16**, 265–270.
Handegard, N. O., Patel, R., and Hjellvik, V. (**2005**). "Tracking individual fish from a moving platform using a split-beam transducer," J. Acoust. Soc. Am. **118**, 2210–2223.
Handegard, N. O., and Tjøstheim, D. (**2005**). "When fish meets a trawling vessel: Examining the behaviour of gadoids using a free floating buoy and acoustic split-beam tracking," Can. J. Fish. Aquat. Sci. **62**, 2409–2422.
Huse, I., and Ona, E. (**1996**). "Tilt angle distribution and swimming speed of overwintering norwegian spring spawning herring," ICES J. Mar. Sci. **53**, 863–873.
Huth, A., and Wissel, C. (**1992**). "The simulation of the movement of fish schools," J. Theor. Biol. **156**, 365–385.
Jaffe, J. S. (**1999**). "Target localization for a three-dimensional multibeam sonar imaging system," J. Acoust. Soc. Am. **105**, 3168–3175.
Jaffe, J. S., Ohman, M. D., and De Robertis, A. (**1999**). "Sonar estimates of daytime activity levels of *Euphausia pacifica* in Saanich Inlet," Can. J. Fish. Aquat. Sci. **56**, 2000–2010.
Jaffe, J. S., Reuss, E., McGehee, D., and Chandran, G. (**1995**). "FTV, a sonar for tracking macrozooplankton in 3-dimensions," Deep-Sea Res., Part I **42**, 1495–1512.
Karp, W., ed. (**1990**). ICES Symposium on Developments in Fisheries Acoustics, Seattle, WA, 1987, Rapp. P.-v. Reun. Cons. Int. Explor. Mer **189**.
Kieser, R., Mulligan, T., and Ehrenberg, J. (**2000**). "Observation and explanation of systematic split-beam angle measurement errors," Aquatic Living Resources **13**, 275–281.
Klevjer, T. A., and Kaartvedt, S. (**2003**). "Split-beam target tracking can be used to study the swimming behaviour of deep-living plankton in situ," Aquatic Living Resources **16**, 293–298.
MacLennan, D. N., Massé, J., and Gerlotto, F., eds. (**2003**). Sixth ICES Symposium on Acoustics in Fisheries and Aquatic ecology, Part 1, Montpellier, France, 2002, ICES J. Mar. Sci. **60**.
Margetts, A., ed. (**1977**). ICES Symposium on Hydro-Acoustics in Fisheries Research, Bergen, Norway, 1973, Rapp. P.-v. Reun. Cons. Int. Explor. Mer **170**.
Massé, J., Gerlotto, F., and MacLennan, D. N., eds. (**2003**). Sixth ICES Symposium on Acoustics in Fisheries and Aquatic ecology, Part 2, Montpellier, France, Aquatic Living Resources **16**.
McKinstry, C. A., Simmons, M. A., Simmons, C. S., and Johnson, R. L. (**2005**). "Statistical assessment of fish behavior from split-beam hydroacoustic sampling," Fisheries Research **72**, 29–44.
McQuinn, I., and Winger, P. (**2003**). "Tilt angle and target strength: Target tracking of atlantic cod (gadus morhua) during trawling," ICES J. Mar. Sci. **60**, 575–583.
Mulligan, T., and Chen, D. (**2000**). "Comment on 'can stationary bottom split-beam hydroacoustics be used to measure fish swimming speed in situ?' by Arrhenius *et al.*," Fisheries Research **49**, 93–96.
Mulligan, T. J., and Chen, D. G. (**1998**). "A split-beam echo counting model: Development of statistical procedures," ICES J. Mar. Sci. **55**, 905–917.
Mulligan, T. J., and Kieser, R. (**1996**). "A split-beam echo-counting model for riverine use," ICES J. Mar. Sci. **53**, 403–406.
Ona, E. (**1994**). "Recent developments of acoustic instrumentation in connection with fish capture and abundance estimation," in *Marine Fish Behaviour in Capture and Abundance Estimation*, edited by A. Fernö and S. Olsen (Fishing News Books, Oxford, UK), pp. 200–214.
Ona, E., ed. (**1999**). "Methodology for Target Strength Measurements," ICES Cooperative Research Report (International Council for the Exploration of the Sea), Vol. **235**.
Parr, A. E. (**1927**). "A contribution to the theoretical analysis of the schooling behaviour of fishes," Occasional Papers of the Bingham Oceanographic Collection **1**, 1–32.
Pedersen, J. (**1996**). "Discrimination of fish layers using the three-dimensional information obtained by a split-beam echo-sounder," ICES J. Mar. Sci. **53**, 371–376.
Reynolds, C. W. (**1987**). "Flocks, herds, and schools: A distributed behavioural model," Comput. Graphics **21**, 25-34.
Sakai, S. (**1973**). "A model for group structure and its behaviour," Biophysics (Engl. Transl.) **13**, 82–90.
Schell, C., and Jaffe, J. S. (**2004**). "Experimental verification of an interpolation algorithm for improved estimates of animal position," J. Acoust. Soc. Am. **116**, 254–261.
Simmonds, E. J., ed. (**1996**). ICES Symposium on Fisheries and Plankton Acoustics, Aberdeen, Scotland, 1995, ICES J. Mar. Sci. **53**.
Soule, M., Hampton, I., and Barange, M. (**1996**). "Potential improvements to current methods of recognizing single targets with a split-beam echo-sounder," ICES J. Mar. Sci. **53**, 237–243.
Torgersen, T., and Kaartvedt, S. (**2001**). "In situ swimming behaviour of individual mesopelagic fish studied by split-beam echo target tracking," ICES J. Mar. Sci. **58**, 346–354.
Trenkel, V. M., Godø, O. R., Handegard, N. O., and Patel, R. (**2004**). "Studying the relationship between spatial fish distributions and trawl catches," Coun. Meet. Int. Coun. Explor. Sea R:26, Copenhagen, Denmark.
Xie, Y. (**2000**). "A range-dependent echo-association algorithm and its application in split-beam sonar tracking of migratory salmon in the fraser river watershed," IEEE J. Ocean. Eng. **25**, 387–398.

Autocorrelation based denoising of manatee vocalizations using the undecimated discrete wavelet transform

Berke M. Gur[a] and Christopher Niezrecki
Department of Mechanical Engineering, University of Massachusetts-Lowell, Lowell, Massachusetts 01854

(Received 13 July 2006; revised 3 April 2007; accepted 4 April 2007)

Recent interest in the West Indian manatee (*Trichechus manatus latirostris*) vocalizations has been primarily induced by an effort to reduce manatee mortality rates due to watercraft collisions. A warning system based on passive acoustic detection of manatee vocalizations is desired. The success and feasibility of such a system depends on effective denoising of the vocalizations in the presence of high levels of background noise. In the last decade, simple and effective wavelet domain nonlinear denoising methods have emerged as an alternative to linear estimation methods. However, the denoising performances of these methods degrades considerably with decreasing signal-to-noise ratio (SNR) and therefore are not suited for denoising manatee vocalizations in which the typical SNR is below 0 dB. Manatee vocalizations possess a strong harmonic content and a slow decaying autocorrelation function. In this paper, an efficient denoising scheme that exploits both the autocorrelation function of manatee vocalizations and effectiveness of the nonlinear wavelet transform based denoising algorithms is introduced. The suggested wavelet-based denoising algorithm is shown to outperform linear filtering methods, extending the detection range of vocalizations. [DOI: 10.1121/1.2735111]

PACS number(s): 43.30.Sf, 43.60.Hj [EJS] Pages: 188–199

I. INTRODUCTION

The West Indian manatee (*Trichechus manatus latirostris*) was added to the endangered species list in 1967 and detailed records of manatee mortalities have been kept since 1974. In 1980, the United States Fish and Wildlife Service established a manatee protection plan. Within this plan, watercraft collisions were identified as the most significant cause of manatee deaths. Accordingly, idle-speed or no-wake zones were designated throughout the Florida waterways where manatee-watercraft collisions are most likely to occur. However, it has been reported that the rate of manatee-watercraft collision related mortalities continue to remain high, despite measures taken (U.S. Fish and Wildlife Service, 2001). The problem also has an economical aspect. The boating industry in Florida has been continuously expanding (24.8% annual growth rate over the last 25 years, 4% annual growth rate since 2001) and reached $18.4 billion in 2005 (Marine Industries Association of Florida Inc., 2007). The passive speed zones are increasing the travel time within the channels and have a negative impact on boating, the boating industry, real estate, and development.

Recent research has focused on developing more effective solutions that can reduce mortalities while forcing minimal restrictions on boaters. Boater warning systems based on active sonar (Jaffe *et al.*, 2007), above sea level infrared (Keith, 2002), and passive acoustic detection (Herbert *et al.*, 2002; Mann *et al.*, 2002; Niezrecki and Beusse, 2002) methods have been previously suggested for this purpose while Gerstein and Blue (2004) investigated an active acoustic manatee warning system. Passive acoustic detection methods exploit the harmonic nature of manatee vocalizations to discriminate them from background noise. However, the harmonic content of manatee vocalizations becomes less discernible as the signal-to-noise ratio (SNR) of the received acoustic signal decreases [see Figs. 1(a) and 1(b)]. Detection becomes impossible when the SNR drops below the critical detection threshold of the detector. Several factors affect the SNR of the received signal such as background noise levels and the distance of the manatee to the receiver hydrophone. Phillips *et al.* (2006) reported that in the absence of any boat noise, the detection range decreased from 250 to 2.5 m when the background noise sound pressure level (SPL) was increased from 70 to 100 dB, for a detection threshold of 6 dB. Taking into account the fact that typical overall SPL generated by boats are approximately 140 dB, detection of manatee calls at long distances in the presence of boat noise becomes a challenging task. Therefore, denoising of underwater acoustic signals, and particularly eliminating the contribution of boat noise is crucial for successful acoustic detection of manatee calls at reasonable ranges, as well as other marine mammal vocalizations within similar noisy habitats. This paper addresses the problem of denoising manatee vocalizations contaminated with boat noise.

A typical manatee vocalization lasts between 0.2 and 0.5 s and has a fundamental frequency followed by several harmonics (see Fig. 1). Both the fundamental and the harmonics generally are frequency modulated and the vocalizations are considered nonstationary. The fundamental frequency ranges between 2 and 5 kHz but can be as low as 600 Hz. In general, the first harmonic dominates the vocalization and contains the majority of the acoustic energy. Harmonics can be found at up to 18 kHz (Schevill and Watkins, 1965; Hartman, 1979; Steel, 1982; Bengston and Fitzgerald, 1985; Gerstein *et al.*, 1999; Nowacek *et al.*, 2003; Phillips *et*

[a]Electronic mail: berke_gur@student.uml.edu

 0001-4966/2007/122(1)/188/12/$23.00

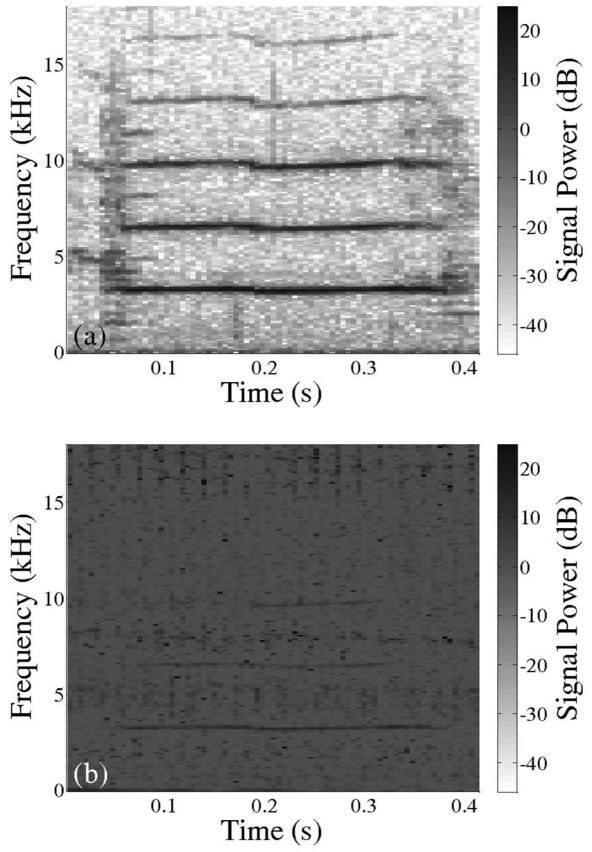


FIG. 1. (a) The spectrogram of a typical manatee call and (b) the spectrogram of the manatee call corrupted by boat noise (SNR=−5 dB).

al., 2004). The underwater acoustic environments of the shallow Florida channels are primarily corrupted with broadband boat noise. Snapping shrimp noise also contributes to the background noise and is more significant in brackish or saltwater environments. Typical snapping shrimp noise is broadband (up to or above 100 kHz), contains high energy spikes, with short durations (~1 ms).

The problem of denoising, detecting, and classifying underwater bioacoustic signals has been studied in the past with a majority of the effort directed toward marine mammals, in particular cetacean vocalizations. A variety of methods, including simple spectral energy ratios (Mellinger and Clark, 1994), matched filtering (Stafford *et al.*, 1998), time warping (Buck and Tyack, 1993), spectral correlation (Mellinger and Clark, 2000), and artificial neural networks (Potter *et al.*, 1994; Deecke and Janik, 2006) have been implemented for this purpose.

In an attempt to detect manatee vocalizations in noisy environments, bandpass and high-pass filters have been incorporated into a harmonic content detector (Niezrecki *et al.*, 2003). Yan *et al.*, (2005, 2006), benchmarked linear adaptive filtering algorithms against the denoising capability of a tenth-order Butterworth bandpass filter. Using test signals with a SNR between 0 and −15 dB, the denoising performance of the feedback adaptive line enhancer (FALE) was the finite-impulse-response filter structured adaptive line enhancer (FIR-ALE).

In the last two decades, the wavelet transform has emerged as an alternative to the short-time Fourier transform (STFT) for time-frequency analysis of nonstationary signals and has been used in analyzing transient signals in sonar applications (Chen *et al.*, 1998; Carevic, 2005). In addition, several researchers including Learned and Willsky (1995), Bailey *et al.* (1998), Huynh *et al.* (1998), and recently, Adam (2006) have implemented the wavelet transform in favor of the STFT in feature extraction, detection, and classification of marine mammal vocalizations.

However, since its introduction, the wavelet transform has found many other applications besides time-frequency analysis. In their innovative work, Donoho and Johnstone (1994) demonstrate that thresholding the wavelet coefficients obtained through an orthogonal wavelet transform results in a simple, nonlinear estimator that surpasses all previous linear estimators in estimating functions of a wide class of smoothness from noisy observations. The authors propose two nonlinear thresholding rules (hard thresholding and soft thresholding), along with an estimation scheme (VISUSHRINK) that performs thresholding on the wavelet coefficients of the noisy signal. The VISUSHRINK estimator is shown to be nearly asymptotically optimal in the minimax sense for the mean-square-error (MSE) risk. Following VISUSHRINK, Donoho and Johnstone (1995) derived the SURESHRINK estimator which utilizes level-dependent thresholding. Coifman and Donoho (1995) suggested that the wavelet decomposition be performed by the undecimated discrete wavelet transform (UDWT) and reported improvements in denoising performance compared to the discrete wavelet transform (DWT) algorithms. The UDWT was further employed in the work of Xu *et al.* (1994), Lang *et al.* (1996), and Pan *et al.* (1999).

The VISUSHRINK and SURESHRINK methods are simple and effective. Despite this, researchers implementing these algorithms emphasized that optimizing the estimator for minimax MSE performance was not suitable for some applications. This led to the development of a variety of wavelet domain denoising schemes. In medical image processing and image restoration applications, several researchers developed alternate thresholds and noise level estimation methods (Chang *et al.*, 2000; Bao and Zhang, 2003; Pizurica *et al.*, 2003). Weiss and Dixon (1997) applied the continuous wavelet transform (CWT) for denoising underwater high-frequency pulse returns.

A wavelet-based denoising strategy is fully described by the wavelet family, level of decomposition, thresholding rule, and the threshold. The selection of these parameters is dependent on the time, frequency, smoothness, and statistical characteristics of the target and noise signals, as well as the risk criteria (Taswell, 2000). The scope of this paper is limited to a single receiver hydrophone setup and possible improvements through more sophisticated multichannel techniques such as beamforming are not included in the discussion. Accordingly, the manatee vocalizations are referred to as the target signal and any signal other than the vocalizations are assumed to be noise. Denoising is defined as estimating the wave form of manatee vocalizations from received signals contaminated with typical underwater noise.

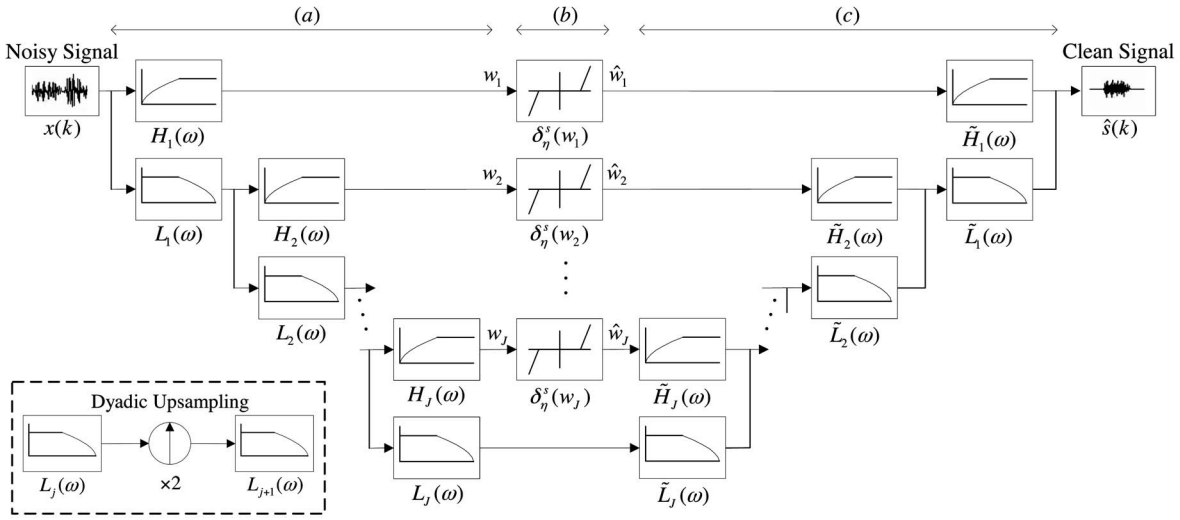


FIG. 2. The wavelet denoising scheme consists of (a) the forward UDWT; (b) thresholding; and (c) the inverse UDWT, where $L_j(\omega)$, $H_j(\omega)$, $\tilde{L}_j(\omega)$, $\tilde{H}_j(\omega)$, and $\delta_\eta^s(w_j)$ are the decomposition/reconstruction QMF pairs and the thresholding rule associated with level j. The low-/high-pass decomposition filter pairs at two consecutive levels are related through upsampling, as shown in the dashed box.

To the authors' knowledge, wavelet-based denoising of underwater acoustic signals for the detection of manatee vocalizations has not yet been investigated and, thus, represents the focus of this paper. In this study, a simple wavelet-domain thresholding scheme is proposed and tested. This paper is organized as follows. Section II provides an overview of the theoretical development of the wavelet transform, classical wavelet shrinkage denoising, and the newly developed autocorrelation-based denoising scheme. The performances of the denoising algorithms introduced in Sec. II are evaluated through a simulation setup defined in Sec. III, while the results of these simulations are presented in Sec. IV. Lastly, Sec. V presents conclusions and suggestions for future research.

II. THEORETICAL DEVELOPMENT

A. The undecimated discrete wavelet transform

The one-dimensional DWT of an analog signal $x(t) \in L^2(\mathbb{R})$ is the integral transform given in

$$\{\mathcal{W}_\psi x\}(j,k) = \int_{-\infty}^{\infty} x(t)\psi_{j,k}^*(t)dt, \tag{1}$$

where $\{\mathcal{W}_\psi x\}(j,k)$ are the wavelet coefficients of the signal at a scale $j \in \mathbb{Z}^+$ and translation $k \in \mathbb{Z}$, and $\psi_{j,k}(t) = 2^{-j/2}\psi(2^{-j}t-k)$ is a set of basis functions. The wavelet analysis of discrete signals is performed within the multirate signal processing and subband decomposition framework, in which the integral transform and the continuous basis functions $\psi_{j,k}(t)$ of Eq. (1) are replaced by the convolution operation and digital filters, respectively. In general, the decomposition is achieved through a pair of low/high pass FIR quadrature mirror filters (QMF), i.e., $L(\omega)$ and $H(\omega)$. Since the frequency content is reduced by half at the output of each filter, only half of the coefficients carry distinct frequency information. Therefore, following the decomposition, the output of each filter is downsampled (i.e., even samples of the signal are dropped) to reduce computational effort in the next step. This filtering/downsampling sequence is repeated on the output of the low-pass filter up to a desired level of decomposition, resulting in a tree-like structure. Each level of decomposition splits the frequency spectrum of the low frequency coefficients in half. Convolution of finite length filters and signals introduce boundary distortions which can be eliminated by zero padding or extending the signal sufficiently and taking only the central portion of the output of the convolution operation. The inverse transform or reconstruction of the decomposed signal is accomplished by upsampling the wavelet coefficients followed by convolving with the reconstruction QMF pairs. The pair of low/high pass filters used for decomposition cannot be used for reconstruction because this leads to aliasing and signal distortion. Therefore, reconstruction is accomplished by another set of low/high pass filters, i.e., $\tilde{L}(\omega)$ and $\tilde{H}(\omega)$, which are derived from of the decomposition QMF pair.

The choice of the coefficients to be downsampled is not trivial and conflicting decomposition results are obtained by choosing a different set of coefficients to eliminate in the downsampling process. This property is often referred to as the translation variance of the DWT, in the sense that a circular shift of the signal will result in a different set of coefficients compared to the unshifted signal. Initially, several methods were suggested to overcome this translation variance of the DWT. One such method, referred to as the UDWT, eliminates the downsampling (and therefore the translation variance) and accommodates for this by upsampling the filter coefficients following each decomposition level, as shown in Fig. 2. The lack of the downsampling step introduces redundancy to the transform.

B. Wavelet domain denoising

The core concept of wavelet-based denoising algorithms is that a target signal of a certain smoothness class can be represented in the wavelet domain with a few, relatively large amplitude coefficients, while noise is assumed to be

mapped to a large number of coefficients with small magnitudes. For example, it is well known that the mapping of white Gaussian noise (WGN) to the wavelet domain results in WGN with the same noise level. Therefore, it is anticipated that eliminating the small amplitude wavelet coefficients will result in eliminating the coefficients due to noise and the reconstruction of the remaining wavelet coefficients will result in a denoised signal. The outcome of the elimination scheme is determined by a thresholding rule and a threshold. The hard thresholding rule eliminates all coefficients below the threshold while the ones above the threshold remain unchanged. Although this method has a better MSE performance, it results in discontinuities and visually unappealing results. The soft thresholding rule, given in Eqs. (2)–(4), eliminates the coefficients below the threshold while the remaining coefficients are shrunk toward zero by an amount equal to the threshold. This shrinkage of the wavelet coefficients eliminates the discontinuities associated with hard thresholding,

$$\delta_\eta^s(\{\mathcal{W}_\psi x\}(j,k)) = \mathrm{sgn}(\{\mathcal{W}_\psi x\}(j,k))(|\{\mathcal{W}_\psi x\}(j,k)| - \eta)_+ \tag{2}$$

where $\delta_\eta^s(\cdot)$ denotes the soft thresholding operator using the threshold η and

$$\mathrm{sgn}(\{\mathcal{W}_\psi x\}(j,k)) = \begin{cases} +1 & \text{if } \{\mathcal{W}_\psi x\}(j,k) > 0 \\ 0 & \text{if } \{\mathcal{W}_\psi x\}(j,k) = 0 \\ -1 & \text{otherwise,} \end{cases} \tag{3}$$

$$(|\{\mathcal{W}_\psi x\}(j,k)| - \eta)_+ = \begin{cases} (|\{\mathcal{W}_\psi x\}(j,k)| - \eta) \\ \text{if } (|\{\mathcal{W}_\psi x\}(j,k)| - \eta) \geq 0 \\ 0 \quad \text{otherwise.} \end{cases} \tag{4}$$

The VISUSHRINK algorithm introduced by Donoho and Johnstone (1994) uses the soft thresholding rule with the universal threshold of $\eta_u = \sigma_v[2\log(N)]^{1/2}$ where the noise level σ_v is estimated through the median absolute deviation (MAD) estimate of the high frequency coefficients of the first level, i.e., $\sigma_v \approx \hat{\sigma}_{w_1} = \mathrm{median}(|w_1|)/0.6745$. The VISUSHRINK is near asymptotically optimal in the minimax sense. Following VISUSHRINK, Donoho and Johnstone (1995) introduced the SURESHRINK algorithm that minimizes Stein's unbiased estimate of risk, given as $\eta_{j,\mathrm{SURE}} = \mathrm{argmin}_\eta[N - 2\cdot\#\{i:|w_j(i)| \leq \eta\} + \Sigma_i \min(|w_j(i)|, \eta)^2]$, where $\#\{\cdot\}$ denotes the number of arguments satisfying the relation given inside the curly braces, for each level of decomposition. In both VISUSHRINK and SURESHRINK, the low frequency coefficients are not thresholded.

C. Optimum threshold denoising

Donoho and Johnstone (1994) argue that an ideal thresholding scheme can be achieved if the denoising algorithm is told which coefficients to eliminate. They show that the estimation error of such an ideal estimator is the lower limit of any estimator and that VISUSHRINK comes close to the performance of this ideal estimator for functions of certain smoothness. All the algorithms presented in this paper are tested using artificially generated test signals; hence, noise-free target signals are available. Therefore, a similar approach to that of the ideal thresholding scheme is adopted to determine the upper limit of denoising performance. The optimum threshold for a soft thresholding scheme is determined by minimizing the MSE between the target and noisy signals' wavelet coefficients, as shown in the following:

$$\eta_{j,\mathrm{opt}} = \arg\min_\eta \left[\sum_k (\delta_\eta^s(\{\mathcal{W}x\}(j,k)) - \{\mathcal{W}s\}(j,k))^2 \right] \tag{5}$$

for each level $j = 1, 2, \ldots, J$. The optimal threshold is not a realizable denoising method, as it calculates the denoised signal with prior knowledge of the noise free signal and represents the upper bound on the performance of any wavelet-based thresholding denoising scheme for a similar setup.

D. Undecimated discrete wavelet transform denoising using an autocorrelation-based threshold

The underwater acoustic noise signal is best modeled as additive, shown in the following:

$$x(n) = s(n) + v(n), \tag{6}$$

where $x(n)$ is the received noisy signal, $s(n)$ is the target, and $v(n)$ is the noise signal, respectively, with $n = 0, 1, 2, \ldots, N-1$. The target signals considered in this paper are manatee vocalizations that are known to have slow decaying and essentially harmonic autocorrelation functions. The noise signal, on the other hand, can be modeled as uncorrelated, normally distributed, nonstationary random sequence with a rapidly decreasing autocorrelation function (Yan *et al.*, 2005). However, the underwater acoustic signals that fall within the scope of this text can be assumed to be wide sense stationary (WSS) within a short time window of 3 ms or 128 samples at a sampling rate of 48 kHz. The distinct behaviors of the autocorrelation functions for both the manatee vocalizations and noise do not change when the signals are mapped to the wavelet domain via the UDWT, as shown in Fig. 3. The autocorrelation function of WSS signals can be estimated using

$$r_{xx}(\tau) = \frac{1}{N}\sum_{i=1}^{N-\tau} (x(i) - \bar{x})(x(i-\tau) - \bar{x}) \tag{7}$$

where τ is the lag and $\bar{x}$ is the mean of the signal. The biased autocorrelation estimator is preferred because it results in better noise reduction performance.

In this paper, the authors suggest an alternative thresholding scheme based on the autocorrelation function of the wavelet coefficients. The proposed autocorrelation-based UDWT algorithm (AUDWT) accomplishes denoising as follows. The received noisy signal is measured with a hydrophone and is processed as 128 sample buffers. The signal is then filtered through a Butterworth tenth-order bandpass filter with a passband ranging from 0.6 to 18 kHz to eliminate any signal contamination that does not overlap with the vocalization bandwidth. The performance of the proposed AUDWT method greatly depends on the effective mapping of the harmonic content of a manatee vocalization to the wavelet domain. This can be accomplished by selecting a wavelet

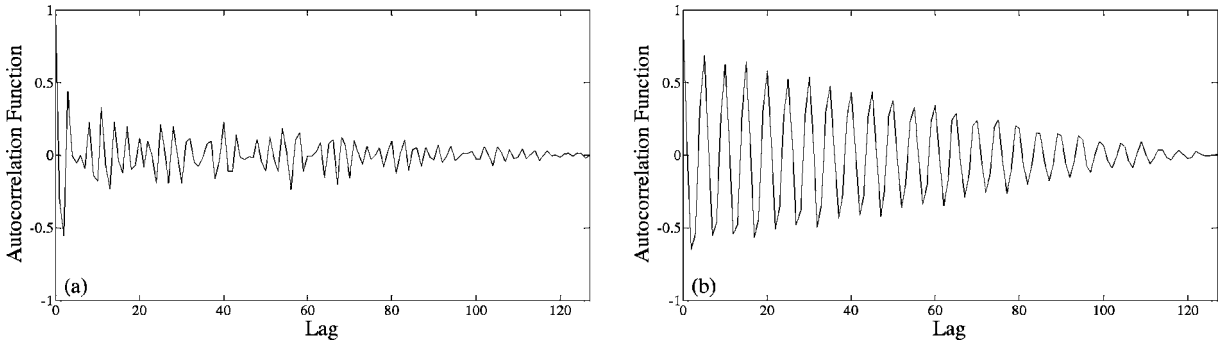


FIG. 3. The normalized autocorrelation function estimates of the detail wavelet coefficients of (a) background noise only and (b) of a manatee call plus background noise for a scale that contains vocalization energy.

with a sufficient number of vanishing moments (i.e., oscillations). Practical implementation of a subband filtering algorithm requires finite filter impulse responses. Furthermore, achieving high temporal resolution is also important since it is desired to suppress short duration signals such as snapping shrimp. The Daubechies family of wavelets achieves a maximum number of vanishing moments for a given temporal support (Daubechies, 1992). Therefore, the filtered signal, sampled at 48 kHz, is decomposed to five levels ($J=5$) using the UDWT with the Daubechies-8 (db8) wavelet. Considering the range of the fundamental frequency and the harmonic structure of manatee vocalizations, a five-level decomposition leads to the most appropriate frequency bands that result in approximate and detail coefficients $\{w_1, w_2, w_3, w_4, w_5, v_5\}$ representing the frequency bands [12–24], [6–12], [3–6], [1.5–3], [0.75–1.5], and [0–0.75] kHz, respectively. Following the UDWT decomposition, the autocorrelation function is estimated using Eq. (7). To distinguish a decaying autocorrelation function due to noise from that of a slow varying autocorrelation of a manatee call within a subband, the rms of the autocorrelation function beyond delay $\tau=20$ is calculated. The calculated rms value of the autocorrelation function is fed into a 16-point moving average filter to eliminate any contamination due to noise transients. The output of the filter is compared to a predetermined threshold constant for a given level of decomposition to determine whether thresholding will be applied to the signal or not. If the output of the filter exceeds the predetermined threshold constants, then the wavelet coefficients are shrunk using the soft thresholding rule. Several thresholds, including thresholds suggested for the UDWT (Berkner and Wells, 2001), were evaluated in terms of the MSE performance (see Sec. III). The universal threshold was determined to be the best performing threshold and is implemented in the AUDWT. If the output of the filter falls short of the threshold constants, then the wavelet coefficients are set to zero. This assures that the silent intervals within calls have no or very few unwanted transients that can increase the false alarm rate of a detector. The approximate coefficients are not included in the threshold rule and are set to zero because no or very little vocalization energy is expected within that subband. Following thresholding, the denoised signal is recovered through an inverse UDWT and the output can be applied to a signal detector or used as an input to a localization system. The block diagram implementation of the AUDWT is given in Fig. 4.

III. SIMULATION

In this study, the target signals used for simulation purposes are obtained from the manatee vocalization library developed by Yan *et al.* (2006). This library is composed of ten different categories of vocalizations based on the structure of the harmonics of the calls (the categories are labeled 0000, 1000, 1010, 1011, 1100, 1110, 1111, 1200, 1210, and 1211). Each category consists of ten calls bringing the total number of calls in the library to 100. All ten calls of each category are collected in a single file in which the vocalizations are spread with 2 s intervals, starting with the first call at 1 s.

Since the main concern of this paper is to evaluate the performance of denoising algorithms in the presence of boat noise, two noise recordings representing different boat speed configurations are used to contaminate the manatee vocaliza-

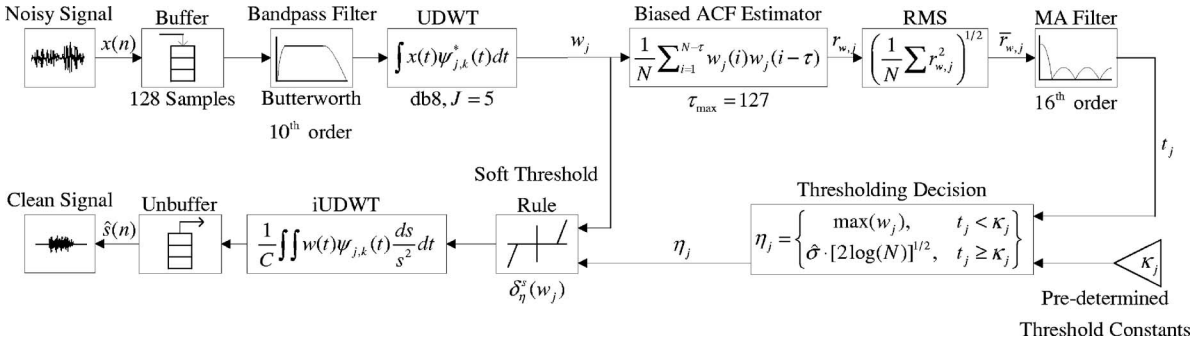


FIG. 4. The block diagram of the AUDWT method.

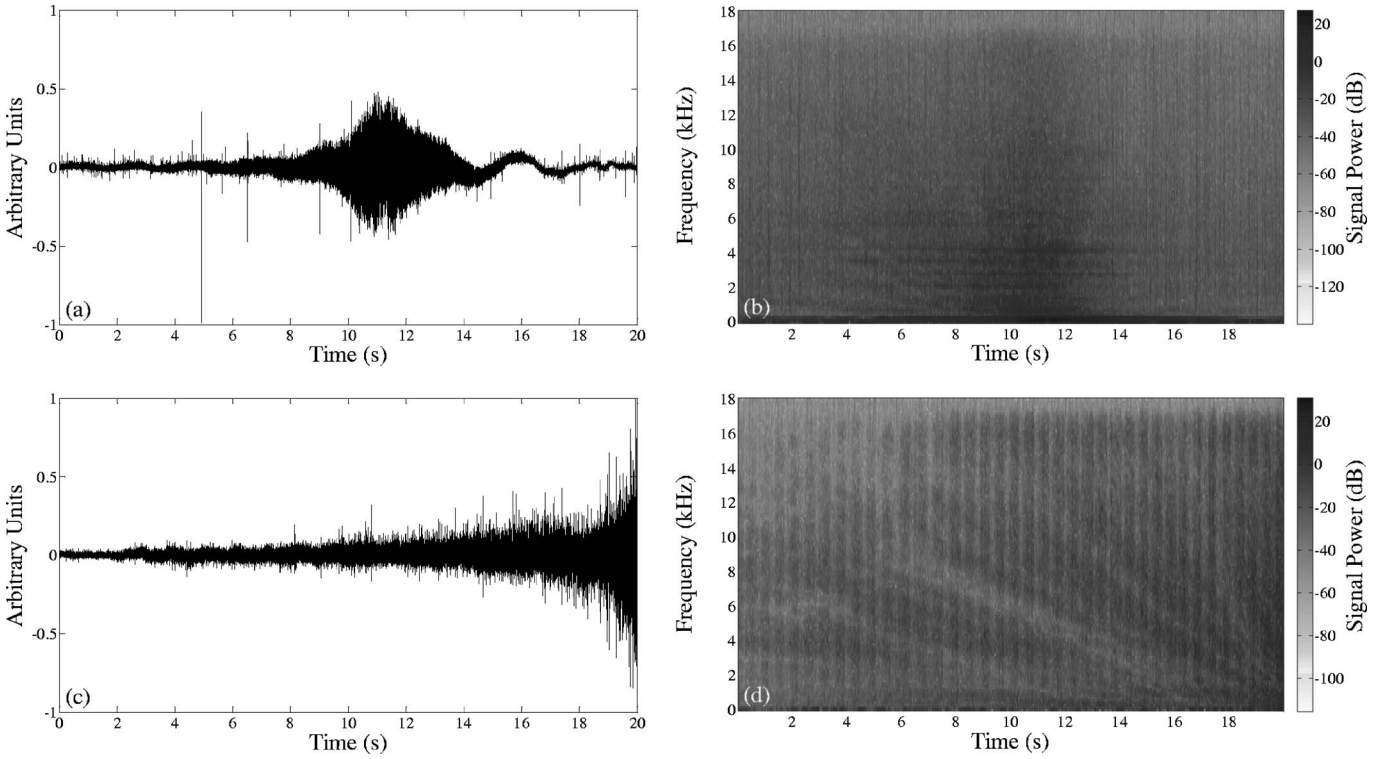


FIG. 5. Boat noise recordings used for simulation and the corresponding frequency spectra. (a), (b) Fast moving boat noise and (c), (d) slow moving boat noise.

tions. These noise recordings are selected from a collection of recordings that were made within Florida. The first recording consists of a boat making a high speed pass in a channel with snapping shrimp noise. The recording was made at Crystal River, FL in 2003. The second recording consists of sounds from a slow moving boat. This recording was made at Cedar Keys, FL in 2004. The time domain plots and spectra for both noise files are shown in Fig. 5. Both simulations consisted of files for each of the ten categories of calls contaminated with their respective noise files (slow or fast moving boat).

In order to evaluate the denoising performances of the methods, the signal power of manatee vocalizations were adjusted such that a constant SNR of −5 dB was achieved throughout the recordings. Yan *et al.* (2006), discuss the difficulty associated with accurately calculating the SNR of nonstationary signals after denoising and has suggested the following formula:

$$\overline{\mathrm{SNR}} = 10 \log_{10}\left[\sum_{i=1}^{10} P_i^2 \Big/ \sum_{i=1}^{10} Q_i^2\right], \tag{8}$$

where P_i is the rms value of the ith vocalization within a category and Q_i is the corresponding rms value of the preceding noisy section and is calculated over a length equal to that of P_i. The improvements in SNR of the vocalizations after denoising were calculated by taking the difference in SNR before and after denoising, both calculated using Eq. (8). Details of Eq. (8) can be found in the work of Yan *et al.* (2006) and the same formulation is adopted in this paper for all SNR measures. However, the SNR can be misleading when used for evaluating denoising performance. As was pointed out by Yan *et al.* (2006), the denoising algorithms in general fail to eliminate noise completely and some residue is left. This residue results in overestimation of the SNR performance. Nonetheless, SNR is an important measure in evaluating detection ranges and therefore is used as a means of measuring how well noise was suppressed. However, the SNR measure is complimented with another performance measure, the normalized mean square error (NMSE). Some applications require that the wave form of the acoustic signal be preserved. Such cases arise when the acoustic signal is used to localize the source and the performances of localizers degrade as the wave form is altered, or when the signals are used for classification and the changes in the wave form result in increased classification errors. Therefore, the NMSE, given in Eq. (9) is also used to evaluate the performance of denoising,

$$\mathrm{NMSE} = \sum_{i=1}^{10} |\hat{s}_i - s_i|^2 \Big/ \sum_{i=1}^{10} \tilde{N}_i, \tag{9}$$

where $\tilde{N}_i$ is the duration of each vocalization within the recording. The noisy vocalization has an initial NMSE, and it is assumed that the denoising reduces this value, while a NMSE of 0 corresponds to perfect preservation of the vocalization. The NMSE values reported in this paper are scaled such that the initial NSME of the noisy vocalization corresponds to 1. Therefore, the resulting NMSE performances of the denoising algorithms fall within the range of [0-1], with a lower value indicating better denoising performance. The NMSE of the optimal threshold method is the true upper performance of denoising and SNR values higher than those

TABLE I. The improvement in SNR for bandpass filtering, VISUSHRINK, FIR-ALE, and FALE, the AUDWT, and the optimal threshold methods for signals corrupted with a fast moving boat noise (original SNR=−5 dB).

Category	Bandpass filter	VISUSHRINK	FIR-ALE	FALE	AUDWT	Optimum
0000	5.95	6.30	14.89	15.65	113.62	68.69
1000	6.17	7.55	19.72	20.08	115.40	67.94
1010	5.85	7.06	17.98	18.42	114.10	69.11
1011	5.85	7.15	19.00	19.56	115.56	69.00
1100	4.43	3.20	16.71	17.16	114.10	67.18
1110	5.16	4.10	16.52	17.04	114.76	68.33
1111	6.09	5.85	18.09	19.04	16.22	69.49
1200	5.96	.48	18.80	19.43	116.32	68.54
1210	6.00	3.58	16.73	17.67	115.79	68.87
1211	5.80	4.14	18.28	18.96	116.60	68.62
Average	5.73	5.14	17.67	18.30	115.25	68.58

predicted by the optimum threshold indicate the presence of noise residue within the vocalization. All artificial recordings are generated such that they contain $N=960\,000$ samples, corresponding to a duration of 20 s for a sampling rate of $f_s=48$ kHz.

The performance of the AUDWT is compared to a bandpass filter, the VISUSHRINK of Donoho and Johnstone (1994), the FIR-ALE and FALE of a Yan *et al.* (2005, 2006), and the optimal threshold method. The bandpass filter used for comparison is the same tenth-order Butterworth bandpass filter with a passband of 0.6–18 kHz used within the AUDWT algorithm. It was observed that the SURESHRINK algorithm offered no distinct advantages compared to the VISUSHRINK method and therefore is not discussed any further. The wavelet domain algorithms (VISUSHRINK, AUDWT, and optimal threshold) process the data in buffers of 128 samples, corresponding to a window of 3 ms. The parameters of the linear adaptive filtering algorithms are set based on the suggestions given in the corresponding papers of Yan *et al.* (2005, 2006).

IV. RESULTS AND DISCUSSION

Previous papers test the performance of wavelet shrinkage algorithms with relatively clean signals having a SNR in the range of 7–20 dB (Donoho and Johnstone, 1994; Coifman and Donoho, 1995; Donoho and Johnstone, 1995; Lang *et al.*, 1996; Pan *et al.*, 1999). The results obtained using these signals are promising. However, the universal threshold of the VISUSHRINK algorithm is based on the asymptotic properties of normally distributed random numbers. Specifically, Donoho (1995) made use of the observation that for a normally distributed sequence of N random numbers $z=[z(0),z(1),\ldots,z(N-1)]$, the maximum absolute value within the sequence becomes bounded with a high probability, as shown in the following:

$$\Pr\{\max|z| \leqslant [2\log(N)]^{1/2}\} \rightarrow 1 \tag{10}$$

as $N\rightarrow\infty$. As the SNR of the signal decreases, the energy contribution of the target signal becomes weaker and the universal threshold estimation becomes too crude.

The resulting improvements in SNR (from an initial SNR of −5 dB) and NMSE of the denoised recordings for a fast traveling boat are given in Table I and Table II, respectively, for the bandpass filtering, VISUSHRINK, FIR-ALE, FALE, AUDWT, and optimal threshold methods. Similarly, the average SNR and NMSE results for denoising of vocalizations contaminated with a slow boat are given in Table III and Table IV, respectively. The bandpass filter performs much better for a fast moving boat than for a slower boat and can increase the SNR up to 6 dB. The fast moving boat noise file has more contribution from sources besides boat noise

TABLE II. The scaled NMSE for bandpass filtering, VISUSHRINK, FIR-ALE, and FALE, the AUDWT, and the optimal threshold methods for signals corrupted with a fast moving boat noise (original SNR=−5 dB).

Category	Bandpass filter	VISUSHRINK	FIR-ALE	FALE	AUDWT	Optimum
0000	0.14	0.13	0.33	0.34	0.14	0.09
1000	0.13	0.11	0.35	0.35	0.09	0.08
1010	0.09	0.07	0.26	0.27	0.06	0.05
1011	0.08	0.06	0.25	0.25	0.05	0.04
1100	0.20	0.20	0.21	0.21	0.17	0.16
1110	0.16	0.15	0.15	0.16	0.13	0.12
1111	0.12	0.12	0.22	0.23	0.09	0.07
1200	0.16	0.18	0.30	0.31	0.14	0.10
1210	0.14	0.15	0.20	0.19	0.13	0.08
1211	0.15	0.15	0.20	0.22	0.12	0.09
Average	0.14	0.13	0.25	0.25	0.11	0.09

TABLE III. The improvement in SNR for bandpass filtering, VISUSHRINK, FIR-ALE, and FALE, the AUDWT, and the optimal threshold methods for signals corrupted with a slow moving boat noise (original SNR= −5 dB).

Category	Bandpass filter	VISUSHRINK	FIR-ALE	FALE	AUDWT	Optimum
0000	0.18	1.39	4.40	3.30	23.39	50.68
1000	0.22	1.41	12.99	20.77	34.71	51.13
1010	0.23	0.53	11.61	18.45	103.63	53.38
1011	0.22	1.40	12.76	20.51	27.11	54.39
1100	0.26	1.22	11.30	19.07	31.66	53.54
1110	0.21	0.91	9.29	13.15	56.94	54.02
1111	0.25	1.11	10.11	17.27	32.69	54.21
1200	0.19	0.61	10.27	16.22	22.20	52.87
1210	0.20	0.90	7.57	14.231	24.09	52.92
1211	0.26	1.25	11.15	18.72	34.80	54.42
Average	0.22	1.07	10.14	16.17	39.12	53.45

such as snapping shrimp and wind. The better performance of the bandpass filter for this noise case can be attributed to the suppression of such noise sources that fall outside the bandwidth of manatee vocalizations. In both scenarios, in terms of the SNR, VISUSHRINK makes no significant improvement to the results achieved by bandpass filtering. The FALE works better than the FIR-ALE, especially in the case of a slow traveling boat, and results in a SNR improvement of 15–20 dB. It is evident from these results that the AUDWT performs best among the tested algorithms for both scenarios, both in the SNR and NMSE measures. However, the effects of noise residue on the SNR performance of the AUDWT are evident for the case of a fast moving boat. The improvements in SNR for the AUDWT are in the range of 115 dB while the predicted upper limit on denoising performance is about 70 dB. This difference is attributed to noise residue; hence the AUDWT has a poorer NMSE performance than the optimum threshold method. For the slow moving boat, the AUDWT achieves an average SNR improvement of 39.1 dB. It should be noted that the thresholds for the AUDWT method were determined empirically from tests conducted with a training data set such that the average NMSE performance was maximized. These thresholds were not modified during the tests with the two boat noise recordings. The noise recordings used to contaminate the manatee vocalizations in these two tests and in the training process come from different boats, set to different configurations and recorded at different locations. Thus, it can be concluded that the performance of the AUDWT algorithm appears to be robust to different sources of contamination. Plots demonstrating the typical resulting denoised signals for the tested algorithms are given in Figs. 6 and 7 for the fast and slow moving boats, respectively. A close inspection of Figs. 7(b) and 7(c) reveals that bandpass filtering alone is not effective in reducing the background noise, particularly for the slow moving boat.

The boat noise and manatee vocalization recordings were made independently, at different locations, using different equipment. To be able to convert the denoising performances to detection ranges, precise knowledge on the source levels of the boat noise and manatee vocalizations is required. Unfortunately, the environment in which these recordings are made makes it difficult to obtain such precise information. Even so, one can infer the detection ranges by assuming typical values for the unknown source levels. Several researchers have investigated background and boat noise levels of manatee habitats. Specifically, background noise levels have been reported to be in the range of 70–105 dB (ref 1 μPa) while the results for boat noise levels vary. However, an average source level of 140 dB (ref 1 μPa, at 1 m)

TABLE IV. The scaled NMSE for bandpass filtering, VISHUSHRINK, FIR-ALE, and FALE, the AUDWT, and the optimal threshold methods for signals corrupted with a slow moving boat noise (original SNR=−5 dB).

Category	Bandpass filter	VISHUSHRINK	FIR-ALE	FALE	AUDWT	Optimum
0000	1.20	0.53	0.28	0.27	0.26	0.21
1000	0.89	0.29	0.20	0.21	0.17	0.12
1010	0.98	0.31	0.19	0.18	0.27	0.17
1011	0.91	0.31	0.31	0.30	0.17	0.12
1100	0.90	0.32	0.25	0.24	0.22	0.15
1110	0.91	0.32	0.27	0.28	0.27	0.17
1111	0.89	0.31	0.26	0.26	0.21	0.14
1200	0.93	0.33	0.26	0.25	0.24	0.18
1210	0.92	0.33	0.26	0.25	0.26	0.18
1211	0.90	0.31	0.28	0.27	0.19	0.14
Average	0.94	0.34	0.26	0.25	0.23	0.16

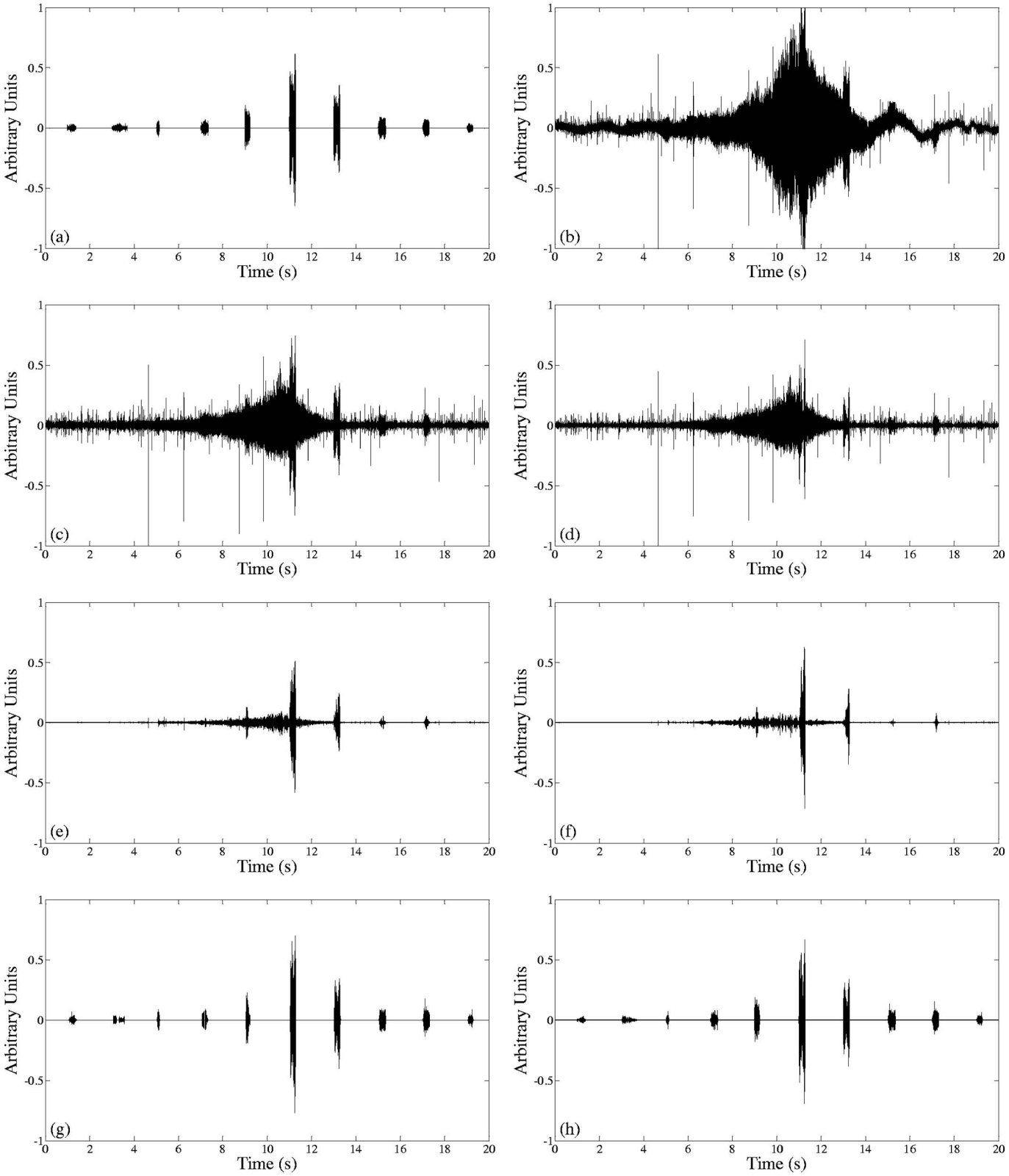


FIG. 6. (a) Noise free manatee vocalizations of type 1210, scaled for −5 dB; (b) test file obtained by the superposition of the scaled manatee vocalizations and fast moving boat noise; (c) denoising results of bandpass filtering; (d) denoising results of VISUSHRINK; (e) denoising results of FIR-ALE; (f) denoising results of FALE; (g) denoising results of AUDWT; and (h) denoising results of the optimal threshold method.

with a noise floor of 120 dB is typical for boat noise. Average manatee source levels (ref 1 μPa, at 1 m) have been reported to vary between 110 and 118 dB (Nowacek *et al.*, 2003; Phillips *et al.*, 2006). Once the boat and manatee source levels are known, the resulting SNR of a denoised recording can be converted to the distance of the manatee to the receiver hydrophone via the sonar equation given the following:

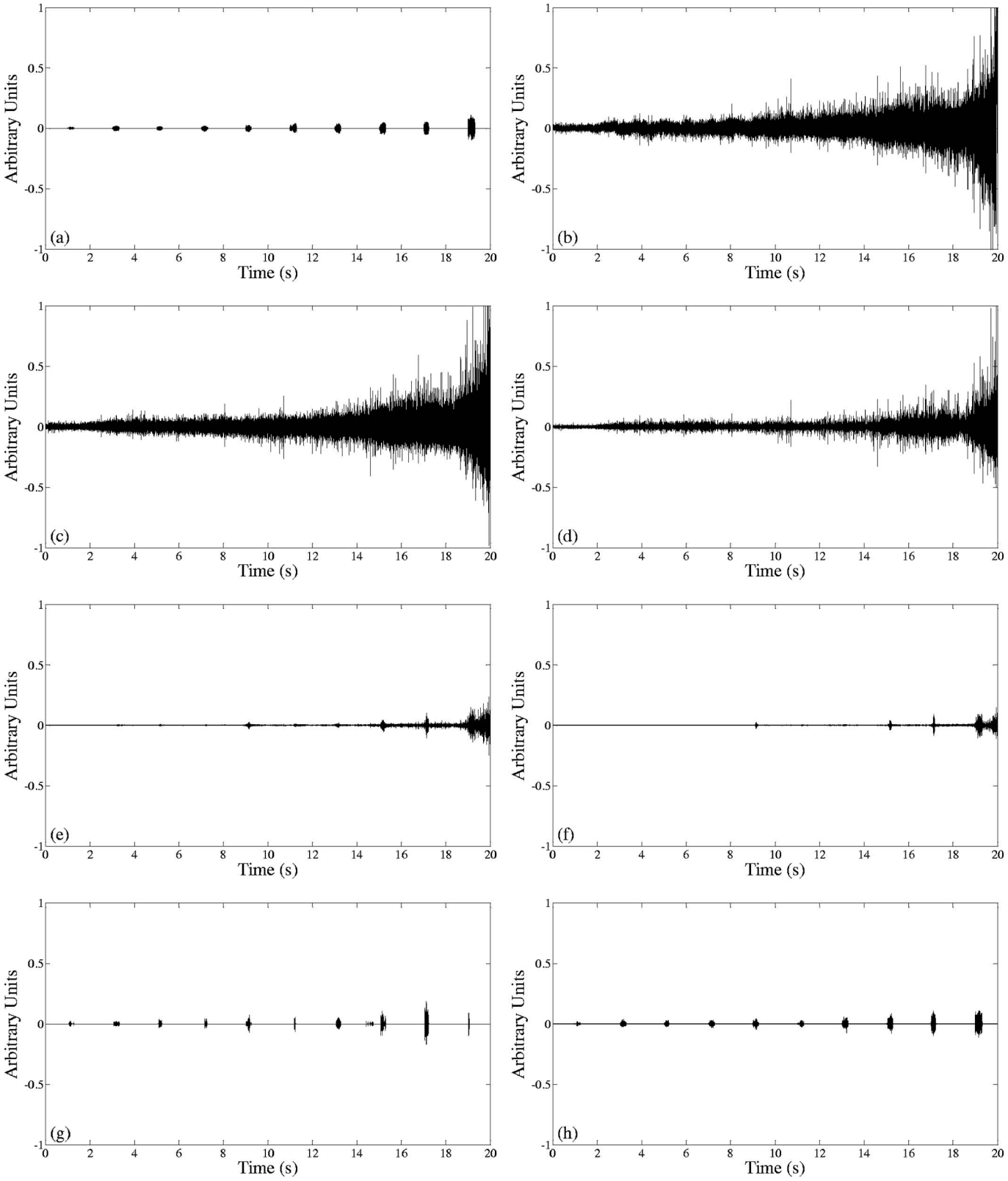


FIG. 7. (a) Noise free manatee vocalizations of type 1110, scaled for −5 dB; (b) test file obtained by the superposition of the scaled manatee vocalizations and slow moving boat noise; (c) denoising results of bandpass filtering; (d) denoising results of VISUSHRINK; (e) denoising results of FIR-ALE; (f) denoising results of FALE; (g) denoising results of AUDWT; and (h) denoising results of the optimal threshold method.

$$\mathrm{SNR}_f = \mathrm{SL}_m - 15\log_{10}(r_m) - \max[\mathrm{BL}, \mathrm{SL}_b - 15\log_{10}(r_b)] + \mathrm{DG}, \quad (11)$$

where SNR_f is the SNR after denoising, SL_m and r_m are the source level and distance of the manatee to the receiver hydrophone, respectively, BL is the background noise level, SL_b and r_b are the boat source level and range to hydrophone, respectively, and DG is the denoising gain. For detec-

tion at a range r_m, the resulting SNR_f for that range must be at least as high as the detection threshold (minimum required SNR) of a harmonic content detector, within the frequency bandwidth of interest. The noise floor of boat noise can be assumed to be 120 dB (ref 1 μPa, at 1 m) within the bandwidth of manatee vocalizations. Further, the manatee vocalizations are assumed to have a source level of 118 dB (ref 1 μPa, at 1 m). Using a mixed spreading model, the SPL of the boat noise received at the hydrophone can be estimated to peak at 105 dB when the boat is passing only 10 m away from the hydrophone. Assuming a detection threshold of 3 dB, the detection range can be calculated to be 4.6 m. Since the AUDWT can improve the SNR above the 3 dB detection threshold from −5 dB, this would correspond to an increase in the detection range up to 16 m. In an alternate scenario, keeping the source levels the same, a boat passing 50 m away from the receiver hydrophone results in a peak SPL of 94.5 dB at the hydrophone. Again, for a detection threshold of 3 dB, the detection range could be increased from 23.2 to 79.4 m.

V. CONCLUSIONS AND FUTURE RESEARCH

The problem of eliminating underwater background noise for detection of manatee vocalizations is challenging, mainly due to typical low SNR of the received signals. Bandpass filtering and the previously suggested wavelet domain VISUSHRINK method fail to achieve acceptable results at such low SNR. This paper demonstrates that the unique behavior of the autocorrelation function of manatee vocalizations can be used to perform nonlinear wavelet-based denoising when the vocalizations are corrupted with nonstationary broadband boat noise. In essence, the proposed AUDWT method outperforms linear filtering methods and has a NMSE performance approaching that of the optimal threshold method. This study demonstrates that nonlinear methods have a greater capability of improving the SNR of noisy manatee vocalizations over linear methods, thereby increasing the probability of detection and detection ranges of manatee vocalizations. Further improvement in noise reduction performance can be achieved by using narrower frequency bins to better isolate the harmonics. This can be accomplished in the wavelet domain using the wavelet packet transform. The increased computational workload associated with the wavelet packet transform can be eliminated to some degree by using an entropy-based cost function to select the most efficient decomposition tree. Noise reduction using the wavelet packet transform to denoise manatee vocalizations will be discussed in a future paper.

The AUDWT has been tested with signals of SNR lower than −5 dB and promising results were obtained. However, the denoising performance of the AUDWT degrades as the SNR approaches −10 dB for the current threshold settings. Therefore, future work can be directed toward identifying the maximum detection range (and thus the anticipated initial SNR of the noisy signals) that satisfies desired detection criteria in terms of correct and false detection rates. The proposed AUDWT makes a thresholding decision based on the presence of a coherent signal within a given frequency band. However, not all coherent signals within the manatee habitat are due to manatee vocalizations (e.g., dolphin vocalizations) and a dedicated manatee vocalization detector is necessary for achieving a satisfactory detection performance. The AUDWT noise reduction scheme can be implemented as a preprocessor to a dedicated manatee vocalization detector, such as the harmonic content detector developed by Niezrecki *et al.* (2003), resulting in a practical passive acoustic detection system to warn boaters of the presence of manatees.

ACKNOWLEDGMENTS

The authors would like to express their sincere appreciation to the Florida Sea Grant, Florida Fish and Wildlife Conservation Commission, and University of Florida Marine Mammal Program in supporting this research. The authors would also like to thank the two anonymous reviewers for their constructive comments.

Adam, O. (**2006**). "Advantages of the Hilbert Huang transform for marine mammals signals analysis," J. Acoust. Soc. Am. **120**, 2965–2973.

Bailey, T. C., Sapatinas, T., Powell, K. J., and Krazanowski, W. J. (**1998**). "Signal detection in underwater sound using wavelets," J. Am. Stat. Assoc. **93**, 73–83.

Bao, P., and Zhang, L. (**2003**). "Noise reduction for magnetic resonance images via adaptive multiscale products thresholding," IEEE Trans. Med. Imaging **22**, 1089–1099.

Bengston, J. L., and Fitzgerald, S. M. (**1985**). "Potential role of vocalizations in West Indian manatees," J. Mammal. **66**, 816–819.

Berkner, K., and Wells, R. O., Jr. (**2001**). "Denoising via nonorthogonal wavelet transforms," in *Wavelet Transforms and Time-Frequency Analysis*, edited by L. Debnath (Birkhäuser, Boston), pp. 68–80.

Buck, J. R., and Tyack, P. L. (**1993**). "A quantitative measure of similarity for *tursiops truncatus* signature whistles," J. Acoust. Soc. Am. **94**, 2497–2506.

Carevic, D. (**2005**). "Adaptive window-length detection of underwater transients using wavelets," J. Acoust. Soc. Am. **117**, 2904–2913.

Chang, S. G., Yu, B., and Vetterli, M. (**2000**). "Adaptive wavelet thresholding for image denoising and compression," IEEE Trans. Image Process. **9**, 1532–1546.

Chen, C. H., Lee, J. D., and Lin, M. C. (**1998**). "Classification of underwater signals using wavelet transforms and neural networks," Math. Comput. Modell. **27**, 47–60.

Coifman, R. R., and Donoho, D. L. (**1995**). "Translation-invariant denoising," in *Wavelets and Statistics*, edited by A. Antoniadis and G. Oppenheim (Springer, New York), pp. 125–150.

Daubechies, I. (**1992**). *Ten Lectures on Wavelets* (SIAM, Philadelphia), Chap. 6, pp. 167–213.

Deecke, V. B., and Janik, V. M. (**2006**). "Automated categorization of bioacoustic signals: Avoiding perceptual pitfalls," J. Acoust. Soc. Am. **119**, 645–653.

Donoho, D. L. (**1995**). "Denoising by soft-thresholding," IEEE Trans. Inf. Theory **41**, 613–627.

Donoho, D. L., and Johnstone, I. M. (**1994**). "Ideal spatial adaptation by wavelet shrinkage," Biometrika **81**, 425–455.

Donoho, D. L., and Johnstone, I. M. (**1995**). "Adapting to unknown smoothness via wavelet shrinkage," J. Am. Stat. Assoc. **90**, 1200–1224.

Gerstein, E. R., and Blue, J. E. (**2004**). "Investigation of the potential utility of manatee alerting devices, phase one: Acoustic warning system to alert manatees of approaching vessels," Report for the Fish and Wildlife Research Institute, http://www.floridamarine.org/features/view_article.asp?id=14362 (Last access date: 3 April 2007).

Gerstein, E. R., Gerstein, L., Forsythe, S. E., and Blue, J. E. (**1999**). "The underwater audiogram of the West Indian manatee (*Trichechus manatus*)," J. Acoust. Soc. Am. **105**, 3575–3583.

Hartman, D. S. (**1979**). "Ecology and behavior of the manatee (*Trichechus manatus*) in Florida," *Special Publication No. 5* (The American Society of Mammalogists, Lawrence, KS), pp. 98–100.

Herbert, T., Hitz, G., Mayo, C. C. S., Dobeck, G., Manning, B., Sandlin, M., Hansel, J., Bowden, T., and Artman, D. (**2002**). "Proof-of-concept for off

the shelf technology to identify acoustic signature to detect presence of manatee(s)," Report for the Fish and Wildlife Research Institute, http://www.floridamarine.org/features/view_article.asp?id=14362 (Last access date: 3 April 2007).
Huynh, Q. Q., Cooper, L. N., Intrator, N., and Shouval, H. (**1998**). "Classification of underwater mammals using feature extraction based on time-frequency analysis and BCM theory," IEEE Trans. Signal Process. **46**, 1202–1207.
Jaffe, J. S., Simonet, F., Roberts, P. L. D., and Bowles, A. E. (**2007**). "Measurement of the acoustic reflectivity of sirenia (Florida manatees) at 171 kHz," J. Acoust. Soc. Am. **121**, 158–165.
Keith, E. O. (**2002**). "Boater manatee awareness system," Report for the Fish and Wildlife Research Institute, http://www.floridamarine.org/features/view_article.asp?id=14362 (Last access date: 3 April 2007).
Lang, M., Guo, H., Odegard, J. E., Burrus, C. S., and Wells, R. O., Jr. (**1996**). "Noise reduction using an undecimated discrete wavelet transform," IEEE Signal Process. Lett. **3**, 10–12.
Learned, R. E., and Willsky, A. S. (**1995**). "A wavelet packet approach to transient signal classification," Appl. Comput. Harmon. Anal. **2**, 265–278.
Mann, D. A., Nowacek, D. P., and Reynolds, J. I. (**2002**). "Passive acoustic detection of manatee sounds to alert boaters," Report for the Fish and Wildlife Research Institute, http://www.floridamarine.org/features/view_article.asp?id=14362 (Last access date: 3 April 2007).
Marine Industries Association of Florida Inc. (**2007**). 7800 SW 57 Avenue, Suite 302, Miami, FL 33143, http://www.boatflorida.org/custom_pages/site_page_2708/index.html (Last access date: 3 April 2007).
Mellinger, D. K., and Clark, C. W. (**1994**). "A publicly accessible database for marine mammal call classification," J. Acoust. Soc. Am. **96**, 3298.
Mellinger, D. K., and Clark, C. W. (**2000**). "Recognizing transient low-frequency whale sounds by spectrogram correlation," J. Acoust. Soc. Am. **107**, 3518–3529.
Niezrecki, C., and Beusse, D. O. (**2002**). "A system for warning boaters in the presence of manatees," Report for the Fish and Wildlife Research Institute, http://www.floridamarine.org/features/view_article.asp?id=14362 (Last access date: 3 April 2007).
Niezrecki, C., Phillips, R., Meyer, M., and Beusse, D. O. (**2003**). "Acoustic detection of manatee vocalizations," J. Acoust. Soc. Am. **114**, 1640–1647.
Nowacek, D. P., Casper, B. M., Wells, R. S., Nowacek, S. M., and Mann, D. A. (**2003**). "Intraspecific and geographic variation of West Indian manatee (*Trichechus manatus spp*.) vocalizations (L)," J. Acoust. Soc. Am. **114**, 66–69.
Pan, Q., Zhang, L., Dai, G., and Zhang, H. (**1999**). "Two denoising methods by wavelet transform," IEEE Trans. Signal Process. **47**, 3401–3406.
Phillips, R., Niezrecki, C., and Beusse, D. O. (**2004**). "Determination of West Indian manatee vocalization levels and rate," J. Acoust. Soc. Am. **115**, 422–428.
Phillips, R., Niezrecki, C., and Beusse, D. O. (**2006**). "Theoretical detection ranges for acoustic based manatee avoidance technology," J. Acoust. Soc. Am. **120**, 153–163.
Pizurica, A., Phillips, W., Lemahieu, I., and Acheroy, M. (**2003**). "A versatile wavelet domain noise filteration technique for medical imaging," IEEE Trans. Med. Imaging **22**, 323–331.
Potter, J. R., Mellinger, D. K., and Clark, C. W. (**1994**). "Marine mammal call discrimination using artificial neural networks," J. Acoust. Soc. Am. **96**, 1255–1262.
Schevill, W. E., and Watkins, W. A. (**1965**). "Underwater calls of *Trichechus* (manatee)," Nature (London) **205**, 373–374.
Stafford, K., Fox, C. G., and Clark, D. S. (**1998**). "Long-range acoustic detection and localization of blue whale calls in the northeast Pacific Ocean," J. Acoust. Soc. Am. **104**, 3616–3625.
Steel, C. (**1982**). "Vocalization patterns and corresponding behavior of the West Indian manatee (Trichechus manatus)," Ph.D. dissertation, Biological Sciences, Florida Institute of Technology, Melbourne, FL.
Taswell, C. (**2000**). "The what, how, and why of wavelet shrinkage denoising," Comput. Sci. Eng. **2**, 12–19.
U.S. Fish and Wildlife Service (**2001**). "Florida manatee recovery plan," 5430 Gorsvenor Lane, Suite 110, Bethesda, MD 20814.
Weiss, L. G., and Dixon, T. L. (**1997**). "Wavelet-based denoising of underwater acoustic signals," J. Acoust. Soc. Am. **101**, 377–383.
Xu, Y., Weaver, J. B., Healy, D. M., Jr., and Lu, J. (**1994**). "Wavelet transform domain filters: A spatially selective noise filteration technique," IEEE Trans. Image Process. **3**, 747–758.
Yan, Z., Niezrecki, C., and Beusse, O. D. (**2005**). "Background noise cancellation for improved acoustic detection of manatee vocalizations," J. Acoust. Soc. Am. **117**, 3566–3573.
Yan, Z., Niezrecki, C., Cattafesta, L. N., III, and Beusse, O. D. (**2006**). "Background noise cancellation of manatee vocalizations using an adaptive line enhancer," J. Acoust. Soc. Am. **120**, 145–152.

Detection performances of experienced human operators compared to a likelihood ratio based detector

Ildar R. Urazghildiiev[a] and Christopher W. Clark
Bioacoustics Research Program, Cornell Laboratory of Ornithology, Ithaca, New York, 14850-1999

(Received 3 November 2006; revised 20 March 2007; accepted 4 April 2007)

This paper compares the detection performances of experienced human operators and an automatic detector that is based on the generalized likelihood ratio test (GLRT). Test data consist of polynomial phase signals and additive white Gaussian noise with signal-to-noise-ratios of 18, 20, 24, and 24 dB. Test results demonstrate that, for a given human operator false alarm probability, the GLRT-based detector provided a higher probability of detecting signals than human operators. [DOI: 10.1121/1.2735114]

PACS number(s): 43.30.Wi, 43.60.Bf [WWA] Pages: 200–204

I. INTRODUCTION

A general problem in bioacoustic applications is that of detecting the calls produced by acoustically active species of interest (see, e.g., Refs. 1–10). Fundamental difficulties in bioacoustic signal detection occur due to signal variability and noise variability. The signal is often transient, occurs unpredictably, and has variable features for which the ranges of variability are not well constrained.[1–10] Ambient noise is typically a non-stationary random process with unknown statistical and spectral properties. In many cases a large portion of the ambient noise consists of a continuous, wideband, random process whose power spectral density does not change essentially over tens of seconds or minutes. This component of the noise is referred to as background noise, and in certain circumstances much of the background noise can be modeled as a Gaussian locally stationary random process.[11] Another confounding component of noise includes impulsive sounds from such things as human activities, other acoustically active animals, and mechanical interaction of the sensor with the medium. When impulsive noise features are similar to signal features (e.g., frequency band, instantaneous bandwidth, duration, frequency modulation) signal detection can be an extremely difficult problem.

Currently, the detection of signals of interest is typically accomplished through a combination of visual inspection of spectrograms and listening to the data.[1–10] A human is considered an experienced human operator (EHO) after lengthy periods of spectrogram inspection and listening combined with visual verification of the species responsible for the sound. An important property of the spectrogram is that it displays the frequency modulation of the signals. This property is used by an EHO for distinguishing between signals and impulsive noises. However, the spectrum of background noise covers the whole frequency band occupied by the signals so that when the signal-to-background noise ratio (SNR) is low, the signals cannot be detected in the spectrogram. As a result, an EHO misses signals with low SNR, and detection probability becomes dependant on the subjective ability of an EHO to visually and acoustically distinguish signals from background noise.

Other detection methods that are widely used in bioacoustics are based on statistically optimal or suboptimal tests (see, e.g., Refs. 11–13). These methods are implemented by computing a statistic that is a function of input data and comparing the statistic with a critical threshold. In the applications where a human makes the final decision, using an automatic detector can reduce the amount of labor required by rejecting a potentially huge portion of the data that contain no signal. To be effective, this procedure assumes that the automatic detector detects a high percentage of true signals and a low percentage non-signals (false alarm detections.) In practice, some percentage of the automatic detections are rejected by an EHO because the signal is not clearly visible in the spectrogram. In such cases, there is no way to check whether the rejections are signals or not. On the other hand, an EHO will select many weak signals that are not detected by the automatic detector. The confidence of such selections is low if the SNR is low. Since the combination of an automatic and a human-based detector can potentially provide many benefits, comparing the detection performances of an EHO and automatic detectors is of practical interest.

Because of the huge variability of signals of interest in bioacoustics, many different strategies can be applied to the automatic detector design. In this paper, we restrict our study by the class of optimal detectors considered in statistical decision theory. An important advantage of this class of detectors is that they minimize the average number of detection errors and provide nearly the best possible detection performance. There are many optimal methods that could be applied, such as Bayesian, Neyman–Pearson, maximum likelihood and others, but all of them are similar in the sense that they use the likelihood ratio test.[14–16] Therefore, the goal of this paper is to estimate the detection performances of an EHO and compare these with the performance of a statistically optimal detector based on the generalized likelihood ratio test (GLRT). To solve this problem, a number of empirical tests using a synthetic data set are conducted. Since many acoustically active species produce transient, fre-

[a]Electronic mail: iru2@cornell.edu

quency modulated (FM) signals, we use synthetic data composed from additive white Gaussian noise and polynomial phase signals (PPS). This data set consists of a total of 400 synthetic calls with different SNRs and examined by five experienced human operators.

Note that in the presence of Gaussian noise, the detection performances of GLRT-based detectors are completely specified by the SNR and do not depend on the type of frequency modulated signals.[15,16] Therefore, the basic results of this paper are applicable to the problem of detecting species that produce transient FM calls such as birds,[1–3] elephants,[4,5] and whales.[6–11] Note also that for certain types of bioacoustical signals, the statistically optimal detector may be unknown or impractical. Therefore, a similar approach as we present in the paper can be used to evaluate the performances of other automatic detectors.

II. DATA MODEL AND PROBLEM FORMULATION

To model the observed process, we computed time series data that combined stationary white Gaussian noise and signals to generate a set of signals with different SNR. The following discrete time series formula was used to generate these data:

$$x(t) = \sum_i A_i s_i(t - \tau_i) + w(t), \quad t = 1, 2, \ldots, \tag{1}$$

where $s_i(t)$ is the ith signal of interest, A_i is the unknown positive nonrandom scalar representing the signal's amplitude, τ_i is the time of arrival TOA of the ith signal, and $w(t)$ is noise. Noise is modeled as a white Gaussian random process with zero-mean and variance σ_0^2. Signals are represented as transient, FM processes with instantaneous frequency given by

$$f(t) = \sum_{m=0}^{M-1} f_m t^m, \quad t = 1, \ldots, N, \tag{2}$$

where f_m is the mth polynomial coefficient and the number of samples N is chosen to ensure a given signal duration. Let $\lambda = (f_0, \ldots, f_{M-1})^T \in U_\lambda$ be the M-dimensional vector of signal parameters and U_λ be the admissible set of λ. Then, the signal can be represented as

$$s(t, \lambda) = \cos[\theta_\lambda(t) + \varphi_0], \tag{3}$$

where $\theta_\lambda(t) = 2\pi \Sigma_{m=1}^{M} (m)^{-1} f_{m-1} t^m$ is the phase and φ_0 is the initial phase. The initial phase is a random value uniformly distributed over the interval $0 - 2\pi$.

A similar model was used in Ref. 11 to simulate contact calls produced by North Atlantic right whales (NARW), *Eubalaena glacialis*. Therefore, without loss of generality, we use the set of model parameters, $U_\lambda = \{\lambda \,|\, f_m^{\min} \leqslant f_m \leqslant f_m^{\max}, m = 0, 1, 2\}$, obtained from an empirical distribution of the parameters of NARW contact calls. Here $f_m^{\min}$, $f_m^{\max}$ are the lower and the upper bounds of the polynomial coefficient f_m, respectively. This type of calls was used in this study because of the availability of human operators who were very experienced with detecting NARW calls. We stress that the basic results obtained using this particular type of signal are applicable to the problem of detecting other species producing similar FM calls.

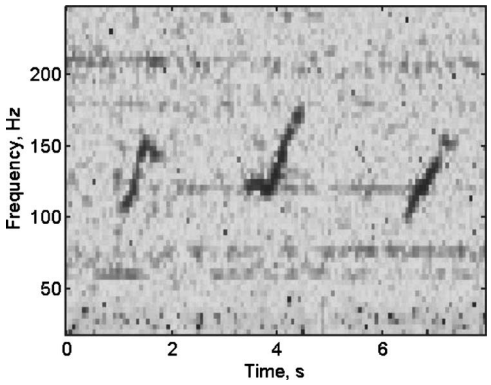


FIG. 1. A spectrogram of three NARW contact calls.

It is assumed that an EHO makes a decision regarding the presence or absence of a call by visually inspecting its spectrogram and listening to the call. An example of the spectrogram of three actual NARW contact calls is shown in Fig. 1. The spectrogram[17] was computed using a Hann window, a 256 point FFT and 50% overlap and represents a typical image used by the EHOs in their work.

The optimal detector is supposed to compute the minimally sufficient statistic in the form of the generalized likelihood ratio[15,16]

$$z(t) = \frac{\max_{A,\lambda} W(\boldsymbol{x}(t)|H)}{W(\boldsymbol{x}(t)|H_0)}, \tag{4}$$

where $\boldsymbol{x}(t) = (x(t), x(t+1), \ldots, x(t+N-1))^T$ is the N-dimensional vector of the data samples, $W(\boldsymbol{x}|H)$, $W(\boldsymbol{x}|H_0)$ are the probability density functions of the vector $\boldsymbol{x}$ under the hypotheses H (signal is present), and H_0 (signal is absent), respectively, and the superscript symbol "T" denotes the transpose. Many different detection schemes can be designed based on the general representation (4). To restrict our investigations, we use a scheme proposed by Urazghildiiev and Clark (2006) to detect NARW contact calls. The corresponding detector was implemented as a bank of P linear FIR filters such that the statistic was calculated in the form

$$z(t) = \max_p |z_p(t)|^2, \quad p = 1, 2, \ldots, P, \tag{5}$$

where $z_p(t) = \Sigma_{n=0}^{N-1} x(t-n) c(n, \lambda_p)$ is the output of the pth filter, $c(t, \lambda) = \exp[j\theta_\lambda(t)]$ is the impulse response of the filter, and P is the number of filters used in the detector. The statistic $z(t)$ is compared with a threshold C and the signal is considered detected if $z(t) \geqslant C$. The block-diagram of the GLRT-based detector implemented as a bank of P linear FIR filters is shown in Fig. 2. The threshold is determined by applying an optimality criterion, which is introduced based on the goals of the experiment as well as on *a priori* information. For instance, using the Neyman–Pearson criterion makes it possible to minimize a false alarm probability for a given probability of detection (or vice versa).[14–16]

The problem considered in this paper is to estimate the detection performances provided by the EHOs and to compare them with the GLRT-based detector specified by Eq. (5).

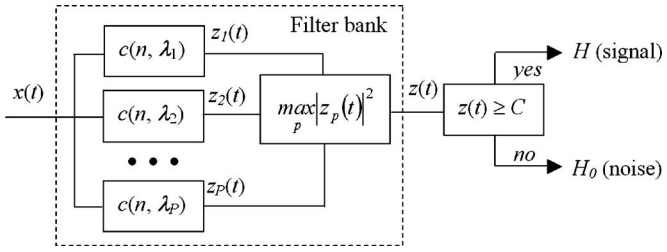


FIG. 2. Block-diagram of the GLRT-based detector implemented as a bank of P linear FIR filters.

We solve this problem by conducting a number of empirical tests using the synthetic data set [Eq. (1)]. Test results are presented in Sec. III.

III. TEST RESULTS

Five human operators with 2 to 5 years of experience in detecting NARW contact calls participated in the tests. One synthetic data set $x(t)$ [Eq. (1)], 24 h in length was created. The data set was composed of the additive sum of white Gaussian noise and four groups of synthetic calls, referred to as signals, where the groups of signals had SNRs of 18, 20, 22, and 24 dB. For the kth group of signals, $k=1,\ldots,4$, the SNR was defined as

$$\mathrm{SNR}_k = \frac{A_k^2 N}{2\sigma_0^2}. \tag{6}$$

Signal duration was 1 s and sampling frequency was 2 kHz so that $N=2000$. For each synthetic signal, the vector of signal parameters, $\lambda=(f_0,f_1,f_2)^T \in U_\lambda$, was composed from the random polynomial coefficients uniformly distributed over the corresponding intervals, $f_m \in |f_m^{\min}, f_m^{\max}|$, $m=0,1,2$, where the lower and the upper bounds of the intervals were obtained from an empirical distribution of the parameters of NARW contact calls, U_λ. The discrete set U_λ consisting of 271 vectors λ was used in the tests. This set was obtained from training data consisting of 721 NARW contact calls.[11] The spectrograms of three synthetic signals having different SNR are shown in Fig. 3. Signal TOAs were chosen as random values uniformly distributed throughout the 24-h data sample. The minimum time interval between any two signals was more than 3 s. A total of 400 signals (100 signals for each SNR) were present in the data. The number of signals, signal TOAs, and other parameters were unknown to the EHOs. EHOs visually inspected spectrograms and listened to synthetic calls, and indicated a detection and its TOA by annotating the signal on the spectrogram. The spectrogram was computed using a Hann window, a 256 point FFT and 50% overlap. A signal was considered to be correctly detected if the minimum time interval between its actual TOA and the TOA specified by the operator was less than 1 s.

The detection performances of the EHOs were evaluated in terms of the probability of detection, α_k (calculated for the group of signals with the same SNR_k), the total number of false alarms, N_{fa}, and the total probability of false alarm, β. The detection probability was calculated as $\alpha_k=M_k/M_\Sigma$ where M_k, M_Σ ($M_\Sigma=100$) are the number of signals with the SNR_k detected by the EHO and present in the data, respectively. A false alarm event occurred if the EHO annotated a segment in the spectrogram as a signal when no signal was present. Therefore, the probability of false alarm was computed as $\beta=N_{fa}/N_\Sigma$ where $N_\Sigma=86\,400$ is the maximum number of 1 s, non-overlapped noise segments existing in the 24 h data set. The detection performances provided by the EHOs are shown in Table I.

To evaluate the detection performance of the GLRT-based detector, the receiver operating characteristics (ROC)

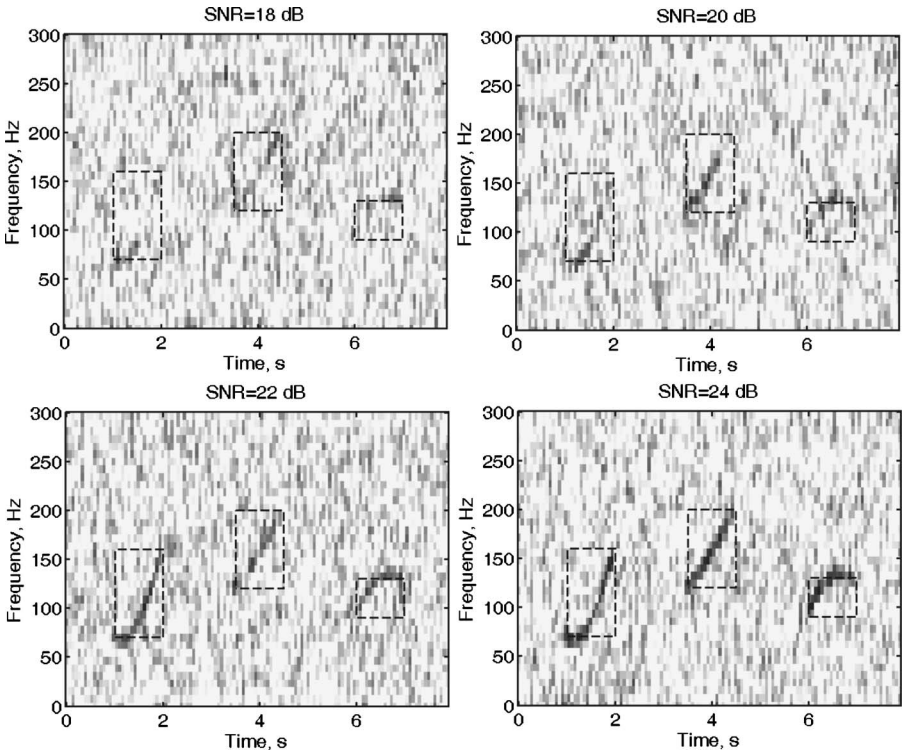


FIG. 3. Spectrograms of three synthetic NARW contact calls with different values of the SNR.

TABLE I. The detection performance provided by five experienced human operators (EHOs).

Parameter	Operator A	Operator B	Operator C	Operator D	Operator E
α_1 ($SNR_1=18$ dB)	0.34	0.34	0.41	0.32	0.06
α_2 ($SNR_2=20$ dB)	0.8	0.86	0.88	0.69	0.28
α_3 ($SNR_3=22$ dB)	1.0	0.99	1.0	0.99	0.81
α_4 ($SNR_4=24$ dB)	1.0	1.0	1.0	1.0	0.95
N_{fa}	74	60	49	85	5
β	8.5×10^{-4}	6.9×10^{-4}	5.6×10^{-4}	9.8×10^{-4}	5.8×10^{-5}

representing the detection probability as a function of the false alarm probability, $\alpha=\alpha(\beta)$,[15,16] were used (see Fig. 4). To calculate the detection probability, 1000 data segments composed of the additive sum of signal and noise,

$$x(t)=A_k s(t,\lambda_i)+w(t), \quad t=t_i,t_{i+1},\ldots,t_{i+N-1}, \tag{7}$$

and 1 s in length each were used. For the data segment, Eq. (7), the corresponding statistic value, $z_i=z(t_i)$, $i=1,2,\ldots,1000$, was calculated using Eq. (5). The vectors λ_p, $p=1,2,\ldots,P$, specifying the structure of the filter bank [Eq. (5)], were taken directly from the empirical *a priori* distribution of the parameters of NARW contact calls, U_λ, such that these vectors were different from the parameters used for simulation of signals. For a given threshold, C, the probability of detection was computed as $\alpha_k(C)=n_k(C)/1000$ where $n_k(C)$ is the number of the segments for which the statistic exceeds the threshold (i.e., the number of the events $z_i \geqslant C$). To calculate the false alarm probability, 86 400 data segments 1 s in length and composed of noise,

$$x(t)=w(t), \quad t=t_i,t_{i+1},\ldots,t_{i+N-1}, \tag{8}$$

were used. The false alarm probability was calculated as $\beta(C)=m(C)/86\,400$ where $m(C)$ is the number of the noise segments [Eq. (8)] for which the statistic exceeded the threshold. The plots of the pairs $\alpha_k=\alpha_k(C)$, $k=1,2,3,4$ and $\beta=\beta(C)$ over the range of thresholds $-\infty<C<\infty$ produce the ROC curves as a function $\alpha_k=\alpha_k(\beta)$.[15,16]

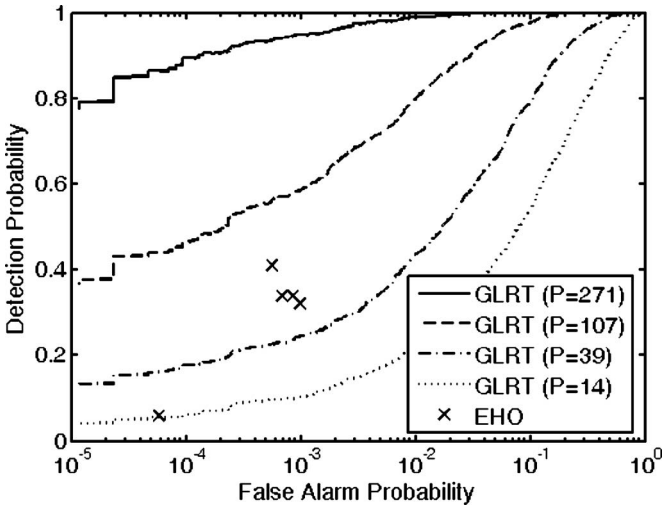


FIG. 4. ROC curves calculated for the GLRT-based detector with P linear FIR filters. Probabilities of detection and false alarm provided by the experienced human operators are also shown. SNR=18 dB.

It is important to note that in the region of SNR $\geqslant 20$ dB and in the region of false alarm probabilities $\beta \geqslant 10^{-5}$, the detection probabilities provided by the GLRT-based detector are very close to unity. Therefore, only the ROC curves computed for the $SNR_1=18$ dB are shown (see Fig. 4). The probabilities of detection and false alarm for each of the five operators are also shown in that figure.

Table I shows that signals with relatively high SNR (22 and 24 dB) were detected by almost all the operators. However, the abilities of operators to detect signals with low SNRs (<20 dB) are variable. For example, operator E had the lowest detection probabilities, but also the lowest false alarm probability. In fact, operator E was more deliberate and did not mark as many detections, so missed more signals but had fewer false detections.

The detection performance provided by the GLRT-based detector essentially depends on the mismatch between the model and signals parameters. As Fig. 4 shows, the detection performance increases as the number of filters increases. Asymptotically, as the number of filters tends to infinity, the GLRT-based detector can provide much higher detection probability for any given false alarm probability of an EHO. However, even if signals with low SNR are detected by the automatic detector, they will likely be rejected by human operators because these weak signals are not obvious in the spectrogram.

IV. CONCLUSION

Our study compared the performances of five experienced human operators and a statistically optimal GLRT-based detector in the task of detecting signals in different levels of SNR, when no impulsive noise was present. We found that human operators were less successful than the GLRT-based detector in the sense that for a given probability of false alarm provided by human operators, the GLRT-based detector provided higher probability of detection. However, for situations in which the human operator makes the final decision, weak signals will likely be rejected if they are neither visible in the spectrogram nor audible in the data. This leads us to two conclusions. First, there are no convincing reasons to design automatic detectors that enable signal detection with such a low SNR that they cannot be validated by a human operator. The performance of the automatic detector should be comparable to the ability of an EHO to detect signals masked by background noise. Second, the input for an automatic detector can be the same as that presented to a

human operator, namely a data spectrogram or other similar transformation of time-series data into a frequency-time representation.

The results of our tests confirmed that the conventional spectrogram is not an optimal transformation of the input data in the case of signals with frequency modulation. When the probability of impulsive noise is negligibly small, the transformation used by the GLRT detector, Eq. (5), can provide higher probability of detection for a given probability of false alarm. The main practical reason for using the spectrogram is that assists EHOs in the recognition of non-Gaussian, impulsive noise. However, more effective transformations suitable for visual detection of FM signals need to be designed and tested.

ACKNOWLEDGMENTS

The author wishes to thank M. Fowler, D. Ponirakis, A. Warde, E. Rowland, and I. Biedron for their assistance in marking right whale calls and for useful discussions. Thanks also to T. Krein for important comments on an earlier draft of this manuscript. Research funded by NOAA grant NA03NMF4720493.

[1]J. W. Bradbury and S. L. Vehrencamp, *Principles of Animal Communication* (Sinauer, Sunderland, MA, 1998).
[2]*Ecology and Evolution of Acoustic Communication in Birds*, edited by D. E. Kroodsma and E. H. Miller (Cornell University Press, Ithaca, NY, 1996).
[3]W. C. Stebbins, *The Acoustic Sense of Animals* (Harvard University Press, Cambridge, MA, 1983).
[4]K. Payne, M. Thompson, and L. Kramer, "Elephant calling patterns as indicators of group size and composition: The basis for an acoustic monitoring system," Afr. J. Ecol. **41**, 99–107 (2003).
[5]M. Garstang, "Long-distance, low-frequency elephant communication," J. Comp. Physiol. **190**, 791–805 (2004).
[6]K. M. Stafford, S. L. Nieukirk, and C. G. Fox, "Low-frequency whale sounds recorded on hydrophones moored in the eastern tropical Pacific," J. Acoust. Soc. Am. **106**, 3687–3698 (1999).
[7]T. F. Norris, M. A. McDonald, and J. Barlow, "Acoustic detections of singing humpback whales (*Megaptera novaeangliae*) in the eastern North Pacific during their northbound migration," J. Acoust. Soc. Am. **106**, 506–514 (1999).
[8]C. W. Clark and P. J. Clapham, "Acoustic monitoring on a humpback whale (*Megaptera novaeangliae*) feeding ground shows continual singing into late Spring," *Proceedings of the Royal Society London, Part B: Biological Sciences*, **271**, 1051–1057 (2004).
[9]C. W. Clark and W. T. Ellison, "Potential use of low-frequency sounds by baleen whales for probing the environment: Evidence from models and empirical measurements," in *Echolocation in Bats and Dolphins*, edited by J. Thomas, C. Moss, and M. Vater (The University of Chicago Press, Chicago, IL, 2000), pp. 564–582.
[10]K. M. Stafford, D. R. Bohnenstiehl, M. Tolstoy, E. Chapp, D. K. Mellinder, and S. E. Moore, "Antarctic-type blue whale calls recorded at low latitudes in the Indian and eastern Pacific Oceans," Deep-Sea Res., Part I **51**, 1337–1346 (2004).
[11]I. Urazghildiiev and C. Clark, "Acoustic detection of North Atlantic right whale contact calls using the generalized likelihood ratio test," J. Acoust. Soc. Am. **120**, 1956–1963 (2006).
[12]D. K. Mellinger and C. W. Clark, "Recognizing transient low-frequency whale sounds by spectrogram correlation," J. Acoust. Soc. Am. **107**, 3518–3529 (2000).
[13]D. Gillespie, "Detection and classification of right whale calls using an "edge" detector operating on a smoothed spectrogram," Can. Acoust. **32**, 39–47 (2004).
[14]E. L. Lehman, *Testing Statistical Hypotheses* (Wiley, New York, 1986).
[15]A. Hero, "Signal detection and classification," in *Digital Signal Processing Handbook*, edited by E. Madisetti and D. Williams (CRC Press, New York, 1999).
[16]H. L. Van Trees, *Detection, Estimation and Modulation Theory*, Part I (Wiley, New York, 2001).
[17]L. Cohen, *Time-Frequency Analysis* (Prentice Hall PTR, NJ, 1995).

Acoustic scattering in dispersions: Improvements in the calculation of single particle scattering coefficients

Valerie J. Pinfield[a)]
Procter Department of Food Science, University of Leeds, Leeds LS2 9JT, United Kingdom

(Received 28 November 2006; revised 26 March 2007; accepted 16 April 2007)

Measurements of ultrasound speed and attenuation can be related to the properties of dispersed systems by applying a scattering model. Rayleigh's method for scattering of sound by a spherical object, and its subsequent developments to include viscous, thermal, and other effects (known as the ECAH model) has been widely adopted. The ECAH method has difficulties, including numerical ill-conditioning, calculation of Bessel functions at large arguments, and inclusion of thermal effects in all cases. The present work develops techniques for improving the ECAH calculations to allow its use in instrumentation. It is shown that thermal terms can be neglected in some boundary equations up to ~100 GHz in water, and several simplified solutions result. An analytical solution for the zero-order coefficient is presented, with separate nonthermal and thermal parts, allowing estimation of the thermal contribution. Higher orders have been simplified by estimating the small shear contribution as the inertial limit is approached. The condition of the matrix solutions have been greatly improved by these techniques and by including appropriate scaling factors. A method is presented for calculating the required Bessel functions when the argument is large (high frequency or large particle size). The required number of partial wave orders is also considered.
 [DOI: 10.1121/1.2737745]

PACS number(s): 43.35.Bf [RR] Pages: 205–221

I. INTRODUCTION

Ultrasound spectroscopy is used with a wide range of materials in which particles, droplets, or molecules are present in a continuous medium, emulsions, dispersions, gels, and solutions of biomolecules. The propagation speed and attenuation are measured for an ultrasound signal transmitted or reflected through the system. The interaction of ultrasound with the material depends on the contrast between the constituent components. In the simplest form, the contrast is of density and compressibility, causing a change in the ultrasound speed and attenuation in the material. At a more complex level, contributions from thermal contrast, and viscous effects occur, and there may also be absorption due to other physical or chemical processes, such as relaxations of chemical equilibria. Sound speed and attenuation are also affected by the particle size distribution and concentration, and by their spatial distribution, e.g., aggregation.

A number of models exist which attempt to relate the sound speed and attenuation in a system to its physical properties. A substantial review of the approaches to the sound propagation problem for dispersed systems is given by Challis *et al.*,[1] including coupled phase models, and scattering models. The scattering model treats each particle as a scatterer of sound, and then determines the combined effect of scattering from an ensemble of such particles. The method was developed over a hundred years ago by Rayleigh,[2] who considered the problem of scattering of sound waves by a single spherical object. Later developments have extended the applicability of the model, for example, to include viscous effects, compressibility of the scatterer, and thermal effects. The scattering model for spherical objects developed by Epstein and Carhart[3] and the closely related Allegra and Hawley[4] model are now well established, and form the basis for the present work. The method is referred to as the ECAH method. This set of models includes viscous and thermal contributions and covers a range of values of the wavenumber-radius parameter ka (where k is the wave number and a the particle radius). The equivalence between the solutions for liquid/liquid (Epstein and Carhart[3]), solid/liquid (Allegra and Hawley[4]), and solid/solid systems (Ying and Truell[5]) has been established (Challis *et al.*[6]), implying that the formulation is of general applicability.

The ECAH method for the scattering of a plane wave by a single spherical particle uses Rayleigh expansions (spherical harmonics) of the sound field potentials, for each wave mode: propagational, thermal, and shear. The appropriate boundary conditions are then applied at the particle surface, resulting in a six square matrix equation to invert for each partial wave order n. However, the matrix equation is badly conditioned, causing inaccuracy with the matrix inversion unless a high degree of machine precision is used. The problem of the high precision required for solution was explored by Harlen *et al.*,[7] and has been documented by Challis and co-workers (O'Neill *et al.*[8]). Further difficulty arises from the need to calculate spherical Bessel and Hankel functions (Heine functions) at large complex arguments. Such numerical problems have limited the application of the method. Although the ECAH method is applicable to general values of ka, both Epstein and Carhart[3] and Allegra and Hawley[4] obtained analytical solutions for the lowest order scattering coefficients in the long wavelength limit, where $ka \ll 1$. These results have been more widely used, but restrict the applica-

a)Electronic mail: valerie.pinfield@Nottingham.ac.uk

tion to a limited frequency and particle size range. At higher values of ka, convergence of the sum of scattering coefficients over all wave orders (harmonics) is not monotonic, so that it is difficult to establish how many partial wave orders need to be included.[8] Some workers have also argued that the inclusion of thermal effects is unnecessary, or that it overcomplicates the solution.[9]

In the work reported here, the objective was to identify simplifications of the ECAH method in order to overcome the problems associated with it. This was achieved by establishing conditions under which certain contributions were small, and therefore causing near-singularity of the matrix. In particular, solutions were obtained for conditions in which the thermal contribution is small, and, independently, conditions for which the viscous contribution is small. In some cases, these contributions can be neglected altogether, whereas in other cases an estimate can be obtained for the small but non-negligible effect. The number of partial wave orders required for accurate velocity and attenuation calculation was also investigated. Additionally, a method was developed for determining the necessary Heine functions at large arguments. The work applies to the determination of the single particle scattering coefficients, and does not consider multiple scattering models. The developments presented here apply to fluid particles in a continuous fluid phase. The set of techniques presented in the paper overcomes many of the current difficulties with the ECAH method. This will permit the use of the scattering model calculations in instrumentation, rather than as a research tool requiring high precision arithmetic.

In Sec. II, the ECAH method is summarized. Section III considers the convergence problem, i.e., the number of partial wave orders needed to achieve a satisfactory result. Modifications to the ECAH method are presented in Sec. IV leading to improved conditioning and a method for calculating Bessel/Hankel functions. Further developments for calculation of the scattering coefficients are given in Sec. V, considering limits in which thermal effects are small, and viscous effects are small. Some numerical results are presented in Sec. VI.

II. THE ECAH METHOD

Full details of the ECAH method can be found in the respective publications, Epstein and Carhart[3] and Allegra and Hawley.[4] The method solves the single particle scattering coefficient for a single spherical particle in an infinite medium with an incident plane wave. Here, both particle and surrounding medium are assumed to be fluids.

A. Equations for solution

In summary, the solution is based on a series of steps:

(1) The equations of motion for sound propagation (including continuity of mass, linear Navier-Stokes, energy) are expressed using scalar and vector velocity potentials,

$$\mathbf{u} = -\nabla\phi + \nabla \times \boldsymbol{\chi}, \tag{1}$$

where $\mathbf{u}$ is the velocity of the fluid, ϕ is a scalar potential, and $\boldsymbol{\chi}$ is a vector potential satisfying the condition $\nabla \cdot \boldsymbol{\chi} = 0$.

(2) A solution with time dependence of the form $e^{-i\omega t}$ (where $\omega = 2\pi f$ is the angular frequency) is substituted in the equations, giving a Helmholtz equation for each of three wave modes with respective wave numbers

$$(\nabla^2 + k_c^2)\varphi = 0, \quad (\nabla^2 + k_T^2)\psi = 0, \quad (\nabla^2 + k_S^2)\boldsymbol{\chi} = \mathbf{0}. \tag{2}$$

The two scalar potentials correspond to longitudinal propagational and thermal wave modes, while the vector potential represents a transverse shear wave mode. The propagational mode is the "usual" mode by which sound travels in a fluid, whilst the thermal mode represents heat flow and the shear mode arises from viscous effects. Both the thermal and shear modes are highly localized because their amplitude decreases exponentially with each wavelength traveled. The total scalar potential is the sum of the propagational and thermal wave potentials. The axis of propagation of the plane wave is taken as the z axis so that the vector potential reduces to a single component χ_ψ denoted simply χ. The vector potential in Epstein and Carhart[3] used the symbol A; here χ has been used to avoid confusion with the scattering coefficient A_n.

(3) The solutions of the Helmholtz equation appropriate for a spherical boundary are partial wave solutions (or Rayleigh series), using spherical harmonic functions (in spherical coordinates), i.e., spherical Bessel or Hankel functions for the radial dependence, and Legendre polynomials (or associated Legendre polynomials) for the angular dependence. The solution is a sum over each partial wave order, so that, for example

$$\varphi = \sum_{n=0}^{\infty} i^n (2n+1) A_n h_n(k_c r) P_n(\cos\theta). \tag{3}$$

(4) Continuity conditions are applied at the boundary for velocity, stress, and thermal effects. A set of six boundary conditions must be solved for each partial wave order n to determine the scattering coefficient A_n.

B. The wave modes

The wave numbers, k_c for the propagational mode, k_T for the thermal mode, and k_S for the shear mode, are given to a very good approximation in fluids by

$$k_c = \frac{\omega}{\nu} + i\alpha, \quad k_T = \left(\frac{\omega}{2\sigma}\right)^{1/2}(1+i), \tag{4}$$

$$k_S = \left(\frac{\omega}{2\upsilon}\right)^{1/2}(1+i),$$

where ν is the speed of sound, α is the attenuation, σ the thermal diffusivity, such that $\sigma = \tau/\rho C_p$ where τ is the thermal conductivity, ρ is the density, C_p is the specific heat capacity at constant pressure, and $\upsilon = \eta/\rho$ is the kinematic viscosity, where η is shear viscosity.

The solution forms for the various wave potentials are given in the following, using the Hankel function h_n outside the particle, and the spherical Bessel function j_n inside the particle. Primed quantities refer to the inside of the particle,

and subscripts 0, c, T, and S refer to the incident wave, propagational, thermal, and shear modes, respectively, throughout the paper,

$$\varphi = \sum_{n=0}^{\infty} i^n (2n+1) A_n h_n(k_c r) P_n(\cos\theta),$$

$$\psi = \sum_{n=0}^{\infty} i^n (2n+1) B_n h_n(k_T r) P_n(\cos\theta),$$

$$\chi = \sum_{n=1}^{\infty} i^n (2n+1) C_n h_n(k_S r) P_n^1(\cos\theta),$$

$$\varphi' = \sum_{n=0}^{\infty} i^n (2n+1) A_n' j_n(k_c' r) P_n(\cos\theta),$$

$$\psi' = \sum_{n=0}^{\infty} i^n (2n+1) B_n' j_n(k_T' r) P_n(\cos\theta),$$

$$\chi' = \sum_{n=1}^{\infty} i^n (2n+1) C_n' j_n(k_S' r) P_n^1(\cos\theta). \quad (5)$$

Similarly the incident field (a plane wave) can be expressed as

$$\varphi_0 = \sum_{n=0}^{\infty} i^n (2n+1) j_n(k_c r) P_n(\cos\theta). \quad (6)$$

C. Boundary conditions

At the surface of the spherical particle, the boundary equations result from the requirement for continuity of fluid velocity components u_r and u_θ, the stress components P_{rr} and $P_{r\theta}$, temperature T, and heat flux $\tau\partial T/\partial r$. These conditions are applied at the spherical boundary $r=a$ where a is the radius of the particle.

For any scalar potential ϕ (propagational and thermal modes) the components are given by

$$u_r = -\frac{\partial\phi}{\partial r}, \quad u_\theta = -\frac{1}{r}\cdot\frac{\partial\phi}{\partial\theta}, \quad (7)$$

$$P_{rr} = (i\omega\rho - 2\eta k^2)\phi - 2\eta\frac{\partial^2\phi}{\partial r^2}, \quad (8)$$

$$P_{r\theta} = 2\eta\frac{\partial}{\partial\theta}\left(\frac{\phi}{r^2} - \frac{1}{r}\frac{\partial\phi}{\partial r}\right),$$

$$T = \Gamma\phi, \quad (9)$$

where the physical properties, wave number k, and thermal factor Γ are appropriate to the wave mode and the material.

The thermal factors are given by

$$\Gamma_c = \frac{-ik_c^2(\gamma-1)}{\beta(\omega + i\gamma\sigma k_c^2)} \approx \frac{k_c^2(\gamma-1)}{\beta\sigma k_T^2}, \quad (10)$$

$$\Gamma_T = \frac{-ik_T^2(\gamma-1)}{\beta(\omega + i\gamma\sigma k_T^2)} \approx -\frac{1}{\beta\sigma},$$

where $\gamma = C_p/C_v$ is the ratio of the specific heat capacities and β is the thermal expansivity. The subscripts c, T, and S refer throughout to the propagational, thermal, and shear modes, respectively. The approximate results on the right-hand side are valid under the condition

$$\left|\frac{k_c^2}{k_T^2}\right| \approx \frac{\omega\sigma}{\nu^2} \ll 1, \quad (11)$$

which is true over a very wide frequency range (up to $\sim 10^{11}$ Hz for water at 30 °C).

It is worth noting that the ratio of the temperature factors for the propagational and thermal modes is very small, showing that the temperature changes are dominated by the thermal mode,

$$\left|\frac{\Gamma_c}{\Gamma_T}\right| \approx \frac{k_c^2(\gamma-1)}{k_T^2}, \quad (12)$$

which is very small.

For the shear wave mode,

$$u_r = \frac{1}{r\sin\theta}\frac{\partial}{\partial\theta}(\chi\sin\theta), \quad u_\theta = -\frac{1}{r}\frac{\partial}{\partial r}(r\chi), \quad (13)$$

$$P_{rr} = \frac{2\eta}{\sin\theta}\frac{\partial}{\partial\theta}\left[\sin\theta\left(-\frac{\chi}{r^2} + \frac{1}{r}\frac{\partial\chi}{\partial r}\right)\right], \quad (14)$$

$$P_{r\theta} = \eta\left[\left(\frac{2\chi}{r^2} - \frac{\partial^2\chi}{\partial r^2}\right) + \frac{1}{r^2}\frac{\partial}{\partial\theta}\left(\frac{1}{\sin\theta}\frac{\partial}{\partial\theta}(\chi\sin\theta)\right)\right]. \quad (15)$$

Equations (14) and (15) for stress are appropriate for fluids. The P_{rr} component is usually divided by the factor $i\omega\rho$ in the boundary equations, and the $P_{r\theta}$ by a factor of η. Further details on the stress relations are given by Povey (Ref. 10, pp. 106–109), and also by Epstein and Carhart[3] and Allegra and Hawley.[4] Here, bulk viscosity has been neglected. The shear wave modes do not contribute to the temperature or heat flow.

When evaluating the boundary conditions, the following relations which were listed by Epstein and Carhart[3] are useful:

$$-\frac{dP_n(\cos\theta)}{d\theta} = P_n^1(\cos\theta), \quad (16)$$

$$\frac{1}{\sin\theta}\frac{d}{d\theta}(\sin\theta P_n^1(\cos\theta)) = n(n+1)P_n(\cos\theta). \quad (17)$$

D. Multiple scattering

In practice, ultrasound measurements are made for a system or ensemble of particles in a dispersion. Usually, the ultrasound velocity (or speed) and/or attenuation is measured, often over a range of frequencies. To relate the single particle scattering properties derived in the subsequent sec-

tions to the sound speed and attenuation in a dispersion, a scattering theory is used. The limiting far-field solution has the form

$$\varphi \rightarrow \frac{e^{ik_c r}}{r} f(\theta), \tag{18}$$

where

$$f(\theta) = \frac{1}{ik_c} \sum_{n=0}^{\infty} (2n+1) A_n P_n(\cos\theta). \tag{19}$$

The multiple scattering result for the wave number of the dispersion, K (Refs. 11–13) is

$$\left(\frac{K}{k_c}\right)^2 = 1 + \frac{3\phi}{k_c^2 a^3} f(0) + \frac{9\phi^2}{4k_c^4 a^6}\left(f^2(\pi) - f^2(0) - \int_0^{\pi} d\theta \frac{1}{\sin(\theta/2)}\left(\frac{d}{d\theta} f^2(\theta)\right)\right). \tag{20}$$

Note that here the symbol ϕ refers to the volume fraction of the dispersed particles.

The resulting dominant terms in the single scattering result are as follows:

$$\frac{1}{\nu_{\text{soln}}^2} = \frac{1}{\nu^2}\left[1 + \frac{3\phi}{k_c^3 a^3} \sum_{n=0} (2n+1)\text{Im}(A_n)\right], \tag{21}$$

$$\alpha_{\text{soln}} = \alpha - \frac{3\phi}{2k_c^2 a^3} \sum_{n=0} (2n+1)\text{Re}(A_n). \tag{22}$$

Hence it can be seen that the first-order term in the velocity and attenuation relates to an infinite sum of the scattering coefficients $(2n+1)A_n$.

Having summarized the ECAH method, the following sections examine a number of methods developed to improve the numerical calculation resulting from the formulation. First, the number of partial wave orders needed in the infinite sum is considered.

III. NUMBER OF PARTIAL WAVE ORDERS: CONVERGENCE

The scattered field for the propagational mode depends on an infinite sum of the scattering coefficients over all partial wave orders. It is this field which is detected experimentally and for which the velocity and attenuation are measured. When calculating the scattered field, the series must be truncated at a finite wave order. In the long wavelength region (small $k_c a$), where the model has been extensively applied, only the lowest orders make a significant contribution, so that the calculation may be limited to $n=0,1,2$. However, at larger values of $k_c a$, higher orders become more significant and must be included in the calculated scattered field. It is necessary to determine at what order the series can be terminated without significant loss of accuracy.

O'Neill *et al.*[8] explored the convergence of the series, demonstrating that the contribution of successive scattering coefficients does not decrease monotonically. The sum shows oscillatory behavior, and some plateaus at which convergence might be believed to have been achieved, although subsequent orders are significant. They demonstrated that where the product $k_c a$ is large, many tens or even over a hundred terms in the series may be necessary to obtain a sufficiently accurate solution. An empirical result was fitted to the data, relating the maximum number of partial wave orders required to the wave number parameter $k_c a$, thus

$$n\text{Max} = 1.05 k_c a + 4 \tag{23}$$

(limited to particle sizes above 1 μm in diameter).

O'Neill *et al.*[8] considered the solution for the scattering coefficients and their sum in the scattered field. However, the magnitude of the scattered wave field for each partial wave order will be limited by the corresponding magnitude of the incoming wave field for that same order. The scattering process, including thermal and shear effects, will modify the proportion of the incident field of a particular order which is scattered as propagational mode, but the magnitude is limited by the incoming wave. It should be noted that this is a linear formulation, and there is no conversion between modes of different orders. The incoming plane wave potential was given in Eq. (6). Each partial wave order has an incident amplitude at the particle surface proportional to the Bessel function $j_n(k_c a)$, and this is the limiting value of the outgoing "amplitude." The scattered propagational field [Eq. (5)] has an amplitude at the surface proportional to $h_n(k_c a)$ so a measure of the maximum coefficient value in heuristic terms is the ratio $j_n(k_c a)/h_n(k_c a)$.

The functions $j_n(z)$, $h_n(z)$, and the ratio $j_n(z)/h_n(z)$ are plotted against order n for a range of values of a real argument z (Fig. 1). For real arguments, $j_n(z)$ is real, but $h_n(z)$ is complex, and the ratio is plotted as the absolute value. It can be seen from the plots that all functions undergo a dramatic "termination" when the order n reaches a value just greater than the argument z. The Bessel function $j_n(z)$ stops its oscillatory behavior and becomes negligible above this order, the Hankel function $h_n(z)$ explodes to very large values, and the absolute value of the ratio $j_n(z)/h_n(z)$ suddenly stops being oscillatory and becomes negligible thereafter (i.e., for all higher orders). The point at which these changes occur relate closely to the empirical formula obtained by O'Neill *et al.*[8] That is around

$$n \sim k_c a + 4. \tag{24}$$

The plots demonstrate that the limit of partial wave orders is determined by the incident wave magnitude for the various orders which limits the amount of scattered field of the same order. Hence the number of partial wave orders required is defined by the number of wave orders needed to describe the incident field for a particle of a given radius. For a plane wave of very large wavelength compared with the particle radius, a small number of partial wave modes are sufficient to define the incident field at the particle surface, since it varies very little across the particle. At smaller wavelengths, more and more wave modes are needed to define the plane wave as partial wave modes at the particle surface. When the wavelength is shorter than the particle radius, for example $(k_c a > 1)$, many modes are needed since the wave amplitude is changing rapidly across the surface. The higher

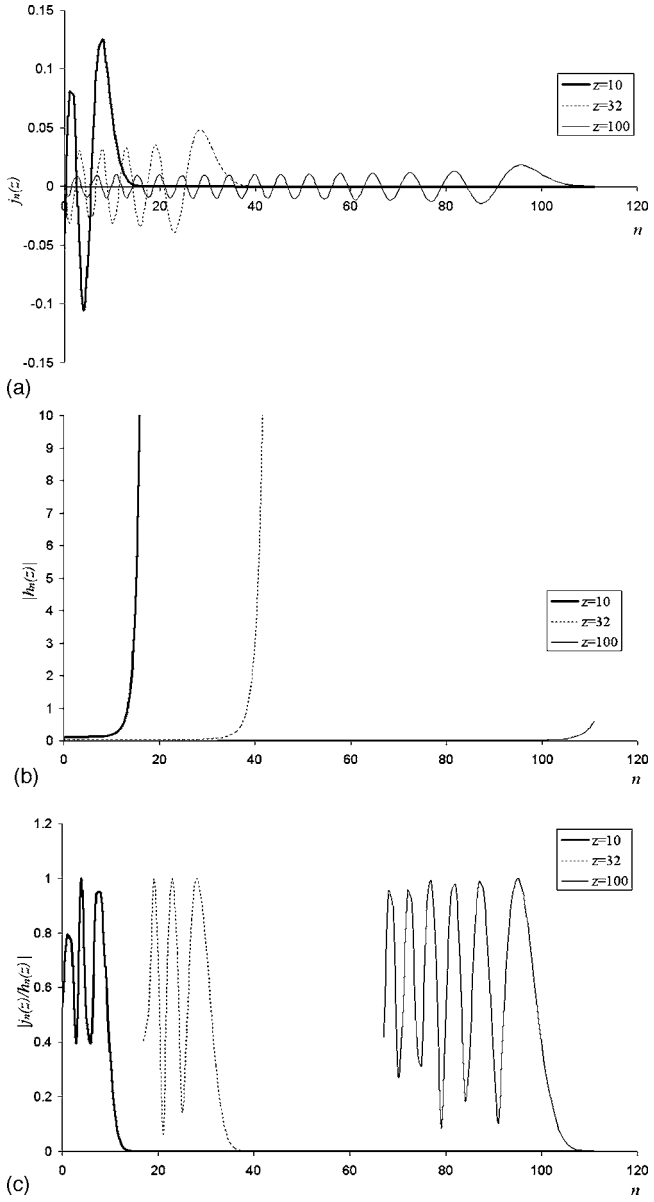


FIG. 1. Plots of the Bessel functions for the incoming and outgoing fields as a function of partial wave order n for a range of real arguments z. (a) $j_n(z)$, (b) $h_n(z)$, and (c) $j_n(z)/h_n(z)$. In (c) the oscillatory behavior over the full range of n has been omitted for clarity. The plots show that the number of orders for which there is a nonzero incoming partial wave is a little greater than the value of z; at higher orders n (on the x axis) the function remains zero. Hence the number of orders needed for a scattering calculation is a little more than $k_c a$. In (c) much of the oscillatory behavior for both $z=32$ and $z=100$ has been omitted for clarity.

wave orders allow greater "detail" in the variation of the incident field, which are required as the wavelength becomes shorter.

IV. MODIFICATIONS TO ECAH METHOD

One of the main difficulties with the ECAH method is that the set of boundary equations is not well suited to numerical solution. The equations form a 6×6 matrix equation for each order n which is inverted to obtain the six scattering coefficients. For the case $n=0$ the system is only a 4×4 matrix. The Hankel functions and Bessel functions used outside and inside the particle, respectively, diverge in magnitude, one becomes small, the other very large as the argument increases. Thus the matrix elements become similarly disparate in magnitude, rendering inversion inaccurate. The quality of the inversion is usually measured by the "condition number" of the matrix, which when multiplied by the machine precision gives the accuracy of the result. The condition numbers are typically very large, especially at large frequencies. A second difficulty is that determination of the Bessel functions is difficult and inaccurate for large arguments. These two elements are addressed in the following sections. The work concentrates on high frequencies or large particle sizes where these problems are most acute. In the long wavelength region (where $k_c a \ll 1$) analytical solutions are available and these problems do not arise.

A. Scaling of thermal and shear wave mode potentials

A straightforward scaling in the wave potential definitions has been used to address the problem of the widely differing magnitude of the components in the boundary equations. The problem occurs primarily for the thermal and shear wave modes, and it is these modes to which the scaling has been applied.

The wave potentials have been redefined with a factor of the relevant Bessel and Hankel function evaluated at the particle surface in each partial wave order. Bessel function zeros occur only for real arguments, so for arguments of the form $\varsigma(1+i)$ there are no zeros. Both thermal and shear wave numbers take this form, so that division by the Bessel function is appropriate. Thus the wave potentials are redefined as

$$\psi = \sum_{n=0}^{\infty} i^n (2n+1) B_n P_n(\cos\theta) h_n(k_T r)/h_n(k_T a),$$

$$\chi = \sum_{n=1}^{\infty} i^n (2n+1) C_n P_n^1(\cos\theta) h_n(k_S r)/h_n(k_S a),$$

$$\psi' = \sum_{n=0}^{\infty} i^n (2n+1) B_n' P_n(\cos\theta) j_n(k_T' r)/j_n(k_T' a),$$

$$\chi' = \sum_{n=1}^{\infty} i^n (2n+1) C_n' P_n^1(\cos\theta) j_n(k_S' r)/j_n(k_S' a). \quad (25)$$

The coefficients for the thermal and shear modes are not now the same as those in ECAH, since they include the appropriate factor.

B. Stress

The adjusted stress components are given here, including the scaling factors shown earlier for the thermal and shear modes. The stress terms then have ratios of the Bessel function derivatives with the scaling Bessel function. The stress components P_{rr} at the boundary have been divided by the factor $i\omega\rho$ (which is equal to ηk_S^2) to define the parameters in the following. For convenience, the inertial parts of the stress (the first terms in q) have been separated from the viscous parts, w,

$$q_{0n} = j_n(k_c a) + w_{0n},$$

$$q_{cn} = h_n(k_c a) + w_{cn},$$

$$q_{Tn} = 1 + w_{Tn},$$

$$q'_{cn} = \hat{\rho} j_n(k'_c a) + w'_{cn},$$

$$q'_{Tn} = \hat{\rho} + w'_{Tn}. \tag{26}$$

where the terms denoted w represent the shear or viscous contribution to the stress which are given by

$$w_{0n} = -2\frac{(k_c a)^2}{(k_S a)^2}[j_n(k_c a) + j''_n(k_c a)],$$

$$w_{cn} = -2\frac{(k_c a)^2}{(k_S a)^2}[h_n(k_c a) + h''_n(k_c a)],$$

$$w_{Tn} = -2\frac{(k_T a)^2}{(k_S a)^2}[1 + h''_n(k_T a)/h_n(k_T a)],$$

$$w'_{cn} = -2\hat{\rho}\frac{(k'_c a)^2}{(k'_S a)^2}[j_n(k'_c a) + j''_n(k'_c a)],$$

$$w'_{Tn} = -2\hat{\rho}\frac{(k'_T a)^2}{(k'_S a)^2}[1 + j''_n(k'_T a)/j_n(k'_T a)]. \tag{27}$$

For the shear waves,

$$q_{Sn} = \frac{2n(n+1)}{(k_S a)^2}\left[k_S a\frac{h'_n(k_S a)}{h_n(k_S a)} - 1\right],$$

$$q'_{Sn} = \frac{2n(n+1)\hat{\rho}}{(k'_S a)^2}\left[k'_S a\frac{j'_n(k'_S a)}{j_n(k'_S a)} - 1\right]. \tag{28}$$

Note that primed and double primed Bessel and Hankel functions denote first and second derivatives rather than the inside of the particle.

For the other stress component, $P_{r\theta}$, the stress is divided by a common factor of η, thus

$$d_{0n} = [k_c a j'_n(k_c a) - j_n(k_c a)],$$

$$d_{cn} = [k_c a h'_n(k_c a) - h_n(k_c a)],$$

$$d_{Tn} = [k_T a h'_n(k_T a)/h_n(k_T a) - 1],$$

$$d'_{cn} = \hat{\eta}[k'_c a j'_n(k'_c a) - j_n(k'_c a)],$$

$$d'_{Tn} = \hat{\eta}[k'_T a j'_n(k'_T a)/j_n(k'_T a) - 1] \tag{29}$$

and for the shear waves

$$d_{Sn} = \frac{1}{2}\left[(k_S a)^2\frac{h''_n(k_S a)}{h_n(k_S a)} + (n^2 + n - 2)\right],$$

$$d'_{Sn} = \hat{\eta}\cdot\frac{1}{2}\left[(k'_S a)^2\frac{j''_n(k'_S a)}{j_n(k'_S a)} + (n^2 + n - 2)\right]. \tag{30}$$

Special case $n=1$

In the particular case when $n=1$, some of these expressions become inaccurate, since they result in a small difference between two much larger quantities when the argument is small. Greater accuracy is achieved using the general Bessel function relation

$$zR'_n(z) - nR_n(z) = -zR_{n+1}(z),$$

where R_n denotes any (spherical) Bessel function. Thus the stress terms become

$$q_{S1} = -\frac{4h_2(k_S a)}{(k_S a)h_1(k_S a)},$$

$$q'_{S1} = -\frac{4\hat{\rho} j_2(k'_S a)}{(k'_S a)j_1(k'_S a)}, \tag{31}$$

$$d_{01} = -k_c a j_2(k_c a),$$

$$d_{c1} = -k_c a h_2(k_c a),$$

$$d'_{c1} = -\hat{\eta} k'_c a j_2(k'_c a),$$

$$d_{T1} = -k_T a h_2(k_T a)/h_1(k_T a),$$

$$d'_{T1} = -\hat{\eta} k'_T a j_2(k'_T a)/j_1(k_T a). \tag{32}$$

C. Bessel functions at large arguments

As has been stated previously, another difficulty with the ECAH method is the requirement to evaluate the spherical Bessel and Hankel functions with large complex arguments. In particular, the values of the thermal and shear parameters $k_T a, k'_T a, k_S a, k'_S a$ can be very large, and have equal real and imaginary parts. The imaginary term is especially a problem for the Bessel functions. The issue has been partly addressed by taking the ratio with the Bessel function itself in the shear and thermal wave modes, as seen in the preceding section. In so doing, the need to evaluate the Bessel functions themselves is removed, leaving only ratios of the derivatives of the Bessel function with its zero-order derivative at the boundary, e.g., $j'_n(k'_T a)/j_n(k_T a)$. At moderate values of the argument, this ratio would be evaluated numerically in the usual way, by determining the value of the derivative and the function itself. However, when the argument is large, it is appropriate to find an alternative way of calculating the ratio. For this reason, analytical approximations to the relevant ratios were sought for the case when the arguments are large.

The following recurrence relations apply to both types of Bessel function:

$$zR_n'(z) + (n+1)R_n(z) = zR_{n-1}(z) \tag{33}$$

and

$$z^2R_n''(z) + 2zR_{n-1}(z) = -z^2R_n(z) + (n+1)(n+2)R_n(z), \tag{34}$$

where R_n denotes the appropriate Bessel function, j_n or h_n. Note that the second of these equations [Eq. (34)] is given incorrectly in Epstein and Carhart.[3] Dividing throughout to obtain the required ratios leads to

$$\frac{R_n'(z)}{R_n(z)} = \frac{R_{n-1}(z)}{R_n(z)} - \frac{(n+1)}{z}, \tag{35}$$

$$\frac{R_n''(z)}{R_n(z)} = -1 + \frac{(n+1)(n+2)}{z^2} - \frac{2R_{n-1}(z)}{zR_n(z)}. \tag{36}$$

The series expansion of the spherical Hankel function gives, when z is large,

$$\frac{h_{n-1}(z)}{h_n(z)} = i + \frac{n}{z} + O\left(\frac{1}{z^2}\right) \tag{37}$$

and hence the derivative ratios are given by

$$\frac{h_n'(z)}{h_n(z)} = i - \frac{1}{z} + O\left(\frac{1}{z^2}\right),$$

$$\frac{h_n''(z)}{h_n(z)} = -1 - \frac{2i}{z} + O\left(\frac{1}{z^2}\right), \tag{38}$$

$$\frac{h_2(z)}{h_1(z)} = -i + \frac{2}{z} + O\left(\frac{1}{z^2}\right).$$

Inside the particle, the spherical Bessel function ratios are needed. The limiting form of the function is oscillatory, but the arguments are of the form

$$z = (1+i)|z|/\sqrt{2} \tag{39}$$

for the thermal and shear wave modes. Hence when the function is written in exponential form, the dominant term when $|z|$ is large is that including the factor $e^{|z|}$, so that

$$j_n(z) \approx -\frac{i^n}{2iz}e^{-iz}. \tag{40}$$

However, a greater accuracy is needed than is given by this limiting form, so that the result was determined numerically from calculated Bessel function values to obtain.

$$\frac{j_{n-1}(z)}{j_n(z)} = -i + \frac{n}{z} + O\left(\frac{1}{z^2}\right) \tag{41}$$

and thus the derivatives are given by

$$\frac{j_n'(z)}{j_n(z)} = -i - \frac{1}{z} + O\left(\frac{1}{z^2}\right),$$

$$\frac{j_n''(z)}{j_n(z)} = -1 + \frac{2i}{z} + O\left(\frac{1}{z^2}\right),$$

$$\frac{j_2(z)}{j_1(z)} = i + \frac{2}{z} + O\left(\frac{1}{z^2}\right).$$

These are valid also for $n=0$.

When the shear or thermal wave number parameters are large, the above presented forms of the relevant Bessel function ratios can be used, thus avoiding the need to calculate Bessel functions for large complex arguments. The method was made possible by scaling all shear and thermal terms by the relevant Bessel function evaluated at the particle boundary.

D. Equations for solution

The resulting six boundary equations for each partial wave order are given in the following, using the notation defined in the previous sections. These equations are in the order $u_r, P_{rr}, u_\theta, P_{r\theta}$, temperature, and heat flow. The thermal and shear wave modes have the Bessel function scaling factors as included in the potential definitions presented earlier. These differ from the ECAH method,

$$k_c a j_n'(k_c a) + k_c a h_n'(k_c a)A_n + k_T a \frac{h_n'(k_T a)}{h_n(k_T a)} B_n - n(n+1)C_n = k_c' a j_n'(k_c' a)A_n' + k_T' a \frac{j_n'(k_T' a)}{j_n(k_T' a)} B_n' - n(n+1)C_n', \tag{42}$$

$$q_{0n} + q_{cn}A_n + q_{Tn}B_n + q_{Sn}C_n = q_{cn}'A_n' + q_{Tn}'B_n' + q_{Sn}'C_n', \tag{43}$$

$$j_n(k_c a) + h_n(k_c a)A_n + B_n - \left[1 + k_S a \frac{h_n'(k_S a)}{h_n(k_S a)}\right] C_n = j_n(k_c' a)A_n' + B_n' - \left[1 + k_S' a \frac{j_n'(k_S' a)}{j_n(k_S' a)}\right] C_n', \tag{44}$$

$$d_{0n} + d_{cn}A_n + d_{Tn}B_n - d_{Sn}C_n = d_{cn}'A_n' + d_{Tn}'B_n' - d_{Sn}'C_n', \tag{45}$$

$$\Gamma_c j_n(k_c a) + \Gamma_c h_n(k_c a)A_n + \Gamma_T B_n = \Gamma_c' j_n(k_c' a)A_n' + \Gamma_T' B_n', \tag{46}$$

$$\Gamma_c k_c a j_n'(k_c a) + \Gamma_c k_c a h_n'(k_c a)A_n + \Gamma_T k_T a \frac{h_n'(k_T a)}{h_n(k_T a)} B_n = \hat{\tau}\left\{\Gamma_c' k_c' a j_n'(k_c' a)A_n' + \Gamma_T' k_T' a \frac{j_n'(k_T' a)}{j_n(k_T' a)} B_n'\right\}. \tag{47}$$

The solution has been made numerically more stable, and a method has been established to avoid determining Bessel functions for large arguments. Further improvements are possible in certain conditions, in which some wave modes make a small or negligible contribution to the scattering. These simplifications are considered in the following sections.

V. SIMPLIFICATIONS FOR THE ECAH METHOD

A. Solution for A_0

1. Equations for the limit $k_c^2/k_T^2 \ll 1$

When $k_c^2/k_T^2 \ll 1$ (up to 10^{11} Hz in water) it is also true that the ratio of the thermal factors for the propagational and thermal modes is small (see Sec. II C), $\Gamma_c/\Gamma_T \ll 1$. Physically, this means that the temperature change caused by the propagational mode is much smaller than that produced by a thermal mode of the same displacement amplitude. Since the thermal modes are produced at the boundary of the particle in order to ensure continuity of temperature and heat flux, only a small "amount" of thermal mode is needed to compensate for a relatively "large" temperature difference caused by the propagational modes inside and outside the particle.

For the zero-order solution, only four boundary conditions are applicable, Eqs. (42), (43), (46), and (47). Comparing Eqs. (42) and (47), the propagational mode terms in the heat flux equation (47) are a ratio Γ_c/Γ_T smaller than the thermal terms. Hence Eq. (47) is satisfied almost entirely by the thermal modes alone, leading to the condition

$$B_0' = \frac{\tau\Gamma_T}{\tau'\Gamma_T'} \cdot \frac{k_T a}{k_T' a} \cdot \frac{h_0'(k_T a)\, j_0(k_T' a)}{h_0(k_T a)\, j_0'(k_T' a)} B_0. \tag{48}$$

This approximation was identified by Epstein and Carhart[3] when obtaining limiting solutions for the long wavelength limit. However, their conclusions are also valid for general values of $k_c a$ and therefore can be applied generally, depending only on the smallness of the ratio k_c^2/k_T^2. Here the solution is continued for the general case, rather than taking the long wavelength limit. Equation (48) can be substituted into the other three boundary equations (42), (43), and (46) to obtain:

$$k_c a h_0'(k_c a) A_0 + b_1 B_0 - k_c' a j_0'(k_c' a) A_0' = -k_c a j_0'(k_c a), \tag{49}$$

$$\Gamma_c h_0(k_c a) A_0 + b_2 B_0 - \Gamma_c' j_0(k_c' a) A_0' = -\Gamma_c j_0(k_c a), \tag{50}$$

$$q_{cn} A_0 + b_3 B_0 - q_{cn}' A_0' = -q_{0n}, \tag{51}$$

where some terms have been grouped together to define new parameters

$$b_1 = k_T a \frac{h_0'(k_T a)}{h_0(k_T a)} \left(1 - \frac{\tau\Gamma_T}{\tau'\Gamma_T'}\right), \tag{52}$$

$$b_2 = \Gamma_T \left(1 - \frac{\tau}{\tau'} \cdot \frac{k_T a}{k_T' a} \cdot \frac{h_0'(k_T a)}{h_0(k_T a)} \cdot \frac{j_0(k_T' a)}{j_0'(k_T' a)}\right), \tag{53}$$

$$b_3 = q_{Tn} - \frac{\tau\Gamma_T}{\tau'\Gamma_T'} \cdot \frac{k_T a}{k_T' a} \cdot \frac{h_0'(k_T a)}{h_0(k_T a)} \cdot \frac{j_0(k_T' a)}{j_0'(k_T' a)} \cdot q_{Tn}'. \tag{54}$$

The solution of Eqs. (49)–(51) is the scattering coefficient A_0. Since it is now only a 3×3 set of equations, an analytical solution becomes more feasible.

2. Separating thermal effects

The zero-order coefficient includes effects due to differences in compressibility and attenuation between the two materials, and those due to thermal differences. Since under some conditions the thermal contribution may be negligible, it is helpful to separate the thermal effects from the rest. When the thermal modes become very small, the matrix solution would be ill-conditioned since the equations will be nearly identical and the matrix nearly singular. This has clearly been a problem where a solution for the matrix boundary equations is forced where the thermal contribution is negligible.

The coefficient is expressed as a sum of nonthermal and thermal terms,

$$A_0 = A_{0N} + A_{0T}, \quad A_0' = A_{0N}' + A_{0T}'. \tag{55}$$

The subscripts nN and nT denote the nonthermal and thermal part of the nth order coefficient. The nonthermal terms can be found by neglecting the thermal wave modes in the boundary equations and solving only the boundary conditions for u_r and P_{rr}, Eqs. (49) and (51), leading to the results

$$A_{0N} = \frac{[k_c a q_{cn}' j_0'(k_c a) - k_c' a q_{0n} j_0'(k_c' a)]}{[k_c' a q_{cn} j_0'(k_c' a) - k_c a q_{cn}' h_0'(k_c a)]}, \tag{56}$$

$$A_{0N}' = \frac{[k_c a q_{cn} j_0'(k_c a) - k_c a q_{0n} h_0'(k_c a)]}{[k_c' a q_{cn} j_0'(k_c' a) - k_c a q_{cn}' h_0'(k_c a)]}. \tag{57}$$

The radial stress parameters q_{cn} and q_{cn}' are dominated by viscous terms at low frequency when $|k_S a|$ is small [see Eqs. (26) and (27)]. In that region the nonthermal part of the zero-order coefficient is imaginary and thus the effect is almost entirely on velocity, with little contribution to attenuation. Epstein and Carhart's[3] limiting analytical result in the long wavelength region is also imaginary. At higher frequencies, where $|k_S a|$ is large, the viscous terms are less significant in the radial stress parameter q_{cn}, which is dominated by the inertial contribution [the first terms in Eq. (26)]. Here there is some contribution to attenuation from the nonthermal coefficient. Equation (56) is the standard result for the scattering coefficient when thermal terms are neglected. Thermal effects are an additional contribution to this, but in some cases can be much larger than this nonthermal component. In other cases, the thermal term is negligible and the nonthermal result is the dominant part of the zero-order coefficient.

The remaining terms in the equations are solved by the thermal parts of the coefficients,

$$k_c a h_0'(k_c a) A_{0T} + b_1 B_0 - k_c' a j_0'(k_c' a) A_{0T}' = 0, \tag{58}$$

$$\Gamma_c h_0(k_c a) A_{0T} + b_2 B_0 - \Gamma_c' j_0(k_c' a) A_{0T}' = c_1, \tag{59}$$

$$q_{cn} A_{0T} + b_3 B_0 - q_{cn}' A_{0T}' = 0, \tag{60}$$

where

$$c_1 = -\Gamma_c j_0(k_c a) - \Gamma_c h_0(k_c a) A_{0N} + \Gamma_c' j_0(k_c' a) A_{0N}'.$$

The solution to this set of equations is straightforward

$$A_{0T} = \frac{c_1 \left[\frac{b_1}{b_3} q_{cn}' - k_c' a j_0'(k_c' a)\right]}{D}, \tag{61}$$

$$D = \left[\left(\frac{b_1}{b_3} q_{cn} - k_c a h_0'(k_c a) \right) \left(\frac{b_2}{b_3} q_{cn}' - \Gamma_c' j_0(k_c' a) \right) - \left(\frac{b_1}{b_3} q_{cn}' - k_c' a j_0'(k_c' a) \right) \left(\frac{b_2}{b_3} q_{cn} - \Gamma_c h_0(k_c a) \right) \right]. \quad (62)$$

In summary, Eqs. (56), (61), and (62) are the solution to the ECAH scattering problem for the zeroth partial wave order in the limit $k_c^2/k_T^2 \ll 1$ for any value of $k_c a$. The solution demonstrates the contribution from nonthermal effects through A_{0N} and thermal effects through A_{0T} independently. The nonthermal term is dependent only on $k_c a$ and $k_c' a$ and includes no thermal factors. The thermal term is a function of the propagational wave numbers and the parameters $k_T a$ and $k_T' a$ which relate to the thermal wave modes.

The new method avoids the use of a matrix solution which is nearly singular when the thermal contribution is small. The matrix solution requires higher and higher accuracy at large $k_T a$ when the thermal modes contribute ever smaller effects. At large frequency, where $k_T a$ is large, the thermal term decreases as $1/(k_T a)$. However, the limiting analytical solution in this case is not of great use, other than to elucidate the nature of the decrease in magnitude of the thermal contribution at large frequency. Having included the Bessel function scaling factor in the thermal wave modes, it is possible to evaluate the thermal contribution for any $k_T a$ however large, by using the asymptotic values of the ratios such as $h_0'(k_T a)/h_0(k_T a)$ as given in Sec. IV C. The problem of evaluating Bessel functions at large arguments is avoided. The separation of the thermal and nonthermal parts of the zero-order coefficient allows an estimate to be made of the significance of the thermal term. Previously, it was only possible to invert the boundary equation matrix to obtain the full coefficient, and the magnitude of the thermal effect was not separated. Here, an initial estimate of thermal effects can be made in order to determine whether or not thermal effects should be included in calculations for a particular system.

B. Solution for higher orders A_n, $n>0$

1. Thermal contribution

For higher orders, $n>0$, all of the six boundary equations must be satisfied. However, when $k_c^2/k_T^2 \ll 1$, the thermal terms can be neglected in the velocity and stress equations. Epstein and Carhart[3] pointed this out in Appendix A, when developing an approximate analytical solution for the long wavelength region for the first-order coefficient A_1. However, the assumption is also valid for any $k_c a$ and any order. These wider applications were not explored by Epstein and Carhart, and the simplifications resulting from the assumption are developed here.

The case for neglecting the thermal wave modes for orders $n>0$ is based on a comparison of Eqs. (42) and (47), and Eqs. (46) and (44). Within the frequency limit given, the ratio of the thermal factors for propagational and thermal modes is again small, Sec. II C. The pairs of similar equations imply that the contribution of the thermal modes in the nonthermal boundary equations (in velocity and stress) is negligible. Hence, for $n>0$ there remain only four boundary equations, those in $u_r, P_{rr}, u_\theta, P_{r\theta}$, with the thermal modes neglected. These can be written in the following form:

$$k_c a h_n'(k_c a) A_n - n(n+1) C_n - k_c' a j_n'(k_c' a) A_n' + n(n+1) C_n' = -k_c a j_n'(k_c a) - k_T a \frac{h_n'(k_T a)}{h_n(k_T a)} B_n + k_T' a \frac{j_n'(k_T' a)}{j_n(k_T' a)} B_n', \quad (63)$$

$$q_{cn} A_n + q_{Sn} C_n - q_{cn}' A_n' - q_{Sn}' C_n' = -q_{0n} - q_{Tn} B_n + q_{Tn}' B_n', \quad (64)$$

$$h_n(k_c a) A_n - \left[1 + k_S a \frac{h_n'(k_S a)}{h_n(k_S a)} \right] C_n - j_n(k_c' a) A_n' + \left[1 + k_S' a \frac{j_n'(k_S' a)}{j_n(k_S' a)} \right] C_n' = -j_n(k_c a) - B_n + B_n', \quad (65)$$

$$d_{cn} A_n - d_{Sn} C_n - d_{cn}' A_n' + d_{Sn}' C_n' = -d_{0n} - d_{Tn} B_n + d_{Tn}' B_n'. \quad (66)$$

The full solution for the viscous and inertial coefficient is obtained by the matrix solution of these equations.

If the thermal terms are neglected entirely, $B_n = B_n' = 0$ the resulting matrix equation for the nonthermal coefficient is as follows:

$$\bar{\bar{L}}_{nN} \begin{pmatrix} A_{nN} \\ C_{nN} \\ A_{nN}' \\ C_{nN}' \end{pmatrix} = \bar{\bar{R}}_{nN}, \quad (67)$$

where

$$\bar{\bar{L}}_{nN} = \begin{pmatrix} k_c a h_n'(k_c a) & -n(n+1) & -k_c' a j_n'(k_c' a) & n(n+1) \\ q_{cn} & q_{Sn} & -q_{cn}' & -q_{Sn}' \\ h_n(k_c a) & -\left[1 + k_S a \frac{h_n'(k_S a)}{h_n(k_S a)} \right] & -j_n(k_c' a) & \left[1 + k_S' a \frac{j_n'(k_S' a)}{j_n(k_S' a)} \right] \\ d_{cn} & -d_{Sn} & -d_{cn}' & d_{S'n}' \end{pmatrix} \quad (68)$$

and

$$\bar{\bar{R}}_{nN} = \begin{pmatrix} -k_c a j_n'(k_c a) \\ -q_{0n} \\ -j_n(k_c a) \\ -d_{0n} \end{pmatrix}. \tag{69}$$

If thermal terms are neglected (setting B_n and B_n' to zero), only four Eqs. (63)–(66) are solved, and the equations are much better conditioned numerically. Since the thermal terms were making a very small contribution, the full matrix equation was nearly singular. Removing the thermal terms has improved the conditioning on the matrix dramatically. In one example, the condition improved from 10^{10} to 10^4, although the improvement depends on all parameters and in some cases, the matrix can still be badly conditioned after removing thermal terms. These conditions are discussed later.

In order to estimate the (initially neglected) thermal contribution, the calculated coefficients A_{nN} and A_{nN}' (which have been obtained by neglecting thermal effects) can be substituted into the thermal boundary equations [Eqs. (46) and (47)], to estimate the thermal coefficients B_n and B_n'. The contribution of these thermal waves to the propagational mode coefficients can then be estimated by solving Eqs. (63)–(66) again, but with the estimated thermal contribution. Since the solution has already been obtained in the absence of thermal terms, the additional contribution to the coefficients, ΔA_{nT}, etc., results simply from the extra terms on the right-hand side, so that

$$\bar{\bar{L}}_{nN} \begin{pmatrix} \Delta A_{nT} \\ \Delta C_{nT} \\ \Delta A_{nT}' \\ \Delta C_{nT}' \end{pmatrix} = \Delta \bar{\bar{R}}_{nT}, \tag{70}$$

where

$$\Delta \bar{\bar{R}}_{nT} = \begin{pmatrix} -k_T a h_n'(k_T a)/h_n(k_T a) B_n + k_T' a j_n'(k_T' a)/j_n(k_T' a) B_n' \\ -q_{Tn} B_n + q_{Tn}' B_n' \\ -B_n + B_n' \\ -d_{Tn} B_n + d_{Tn}' B_n' \end{pmatrix}. \tag{71}$$

Thus the error in neglecting thermal effects can be easily estimated, to establish the validity of the approximation. The cycle could be repeated to obtain a more accurate result including thermal effects if necessary. First substitute the new estimate for the coefficients $A_n = A_{nN} + \Delta A_{nT}$ (and similarly for A_n') into the thermal equations (46) and (47), to obtain an improved estimate of the thermal coefficients, then solve Eq. (70) again to obtain a new estimate of the thermal contribution ΔA_{nT}. The difference between consecutive estimates of the thermal contribution, ΔA_{nT}, demonstrates whether the thermal effects are significant. When consecutive estimates are unchanged, the estimate is accurate. Note that the matrix inversion need only be carried out once, since it occurs in both equations with only the right-hand side being different.

The above-demonstrated method is of general validity where some terms in the equations make a small contribution. The neglect of thermal terms in the boundary equations for orders $n > 0$ was noted by Epstein and Carhart but applied only in a very limited set of conditions. Here, it is applied in the general case, and leads to a significant improvement in the condition of the matrix equation. In addition it is possible to estimate the error or thermal contribution, in order to establish confidence in the approximation, or to improve the calculation where necessary.

2. Shear contribution

At high frequencies, or at very low viscosity, when $|k_S a|^2 \gg 1$, the shear mode terms make a negligible contribution and the viscous contribution to the stress components are negligible. These conditions are referred to as the *inertial limit*. When the shear modes contribute only a small amount in the boundary equations, the boundary matrix equation will again become nearly singular, and badly conditioned. It is therefore necessary under these conditions to separate the effect of the shear wave modes from the solution in a similar way to the separation of thermal terms in Sec. V B.1

The nth order coefficient can be expressed as

$$A_{nN} = A_{nNI} + A_{nNV}, \tag{72}$$

where A_{nNI} is the inertial limit and A_{nNV} is the viscous contribution. The subscript N denotes the fact that this is the nonthermal part of the solution, in which the thermal modes have been neglected. For the inertial limit, only the equations in radial velocity u_r and stress P_{rr} [Eqs. (42) and (43)] are relevant, with both the thermal and shear mode coefficients set to zero. The "density" or inertia terms dominate the stress components, and hence the viscous contributions w_{0n}, w_{cn}. and w_{cn}' are neglected (Sec. IV B). Thus

$$A_{nNI} = -\frac{[\hat{\rho} k_c a j_n'(k_c a) j_n(k_c' a) - k_c' a j_n(k_c a) j_n'(k_c' a)]}{[\hat{\rho} k_c a h_n'(k_c a) j_n(k_c' a) - k_c' a h_n(k_c a) j_n'(k_c' a)]} \tag{73}$$

and inside the particle

$$A_{nNI}' = \frac{k_c a [j_n(k_c a) h_n'(k_c a) - h_n(k_c a) j_n'(k_c a)]}{[\hat{\rho} k_c a h_n'(k_c a) j_n(k_c' a) - k_c' a h_n(k_c a) j_n'(k_c' a)]}. \tag{74}$$

This is the true inertial limit. Although A_{nNI} is a function of $k_c a$, the real part of $k_c a$ varies linearly with frequency, but the imaginary part varies with frequency in the same way as the attenuation, which could be second order or lower. Hence, when the attenuation is significant there may be some frequency dependence of A_{nNI} which does not scale with fa.

Now it is necessary to determine the viscous or shear contribution to the coefficient A_{nN} which is denoted A_{nNV} [Eq. (72)]. In a similar way to the isolation of the remaining thermal terms in Sec. V B.1, the shear terms result from the matrix $\bar{\bar{L}}_{nN}$ with the extra components on the right-hand side. Thus

$$\bar{\bar{L}}_{nN} \begin{pmatrix} A_{nNV} \\ C_n \\ A_{nNV}' \\ C_n' \end{pmatrix} = \bar{\bar{R}}_{nNV}, \tag{75}$$

$$\bar{R}_{nNV} = \begin{pmatrix} 0 \\ -w_{0n} - w_{cn}A_{nNI} + w'_{cn}A'_{nNI} \\ -(\hat{\rho}-1)j_n(k'_c a)A'_{nNI} \\ -d_{0n} - d_{cn}A_{nNI} + d'_{cn}A'_{nNI} \end{pmatrix}. \quad (76)$$

The resulting solution is exact (within the nonthermal approximation) using Eq. (72). However, the matrix equation which remains to be solved relies on inversion of the same matrix $\bar{L}_{nN}$ as did the full (nonthermal) solution. Hence it does not achieve any improvement in conditioning numerically, and the shear contribution A_{nNV} becomes difficult to calculate at large values of $k_S a$. The only benefit in defining the problem in this way is to identify the contributions of inertial and viscous effects to the coefficient. The inertial part is the limiting solution at large frequencies.

In order to identify the behavior of the shear contribution as the inertial limit is approached, the solution for A_{nNV} was obtained analytically and expanded as a series in powers of $1/(k_S a)$. The appropriate limiting expressions for the Bessel functions at large arguments were substituted for $|k_S a|^2 \gg 1$ and the parameter inside the particle defined as

$$k'_S a = \sqrt{\frac{\hat{\rho}}{\hat{\eta}}} k_S a.$$

The algebraic solution was determined using MAPLE (version 10). The leading order term at large $|k_S a|$ was found to be

$$A_{nNV} = -\frac{n(n+1)}{k_c a \cdot k_S a} \frac{\sqrt{\hat{\rho}\hat{\eta}}}{(1+\sqrt{\hat{\rho}\hat{\eta}})} \times \left[\frac{(\hat{\rho}-1)j_n(k'_c a)}{[\hat{\rho} k_c a h'_n(k_c a) j_n(k'_c a) - k'_c a h_n(k_c a) j'_n(k'_c a)]} \right]^2. \quad (77)$$

However, the numerical results for all the terms in the series showed that the leading order term did not dominate under the conditions selected, and therefore should not be generally adopted as an approximate solution at large $|k_S a|$.

An alternative method is now proposed which is applicable when the shear wave modes cannot be neglected entirely, but their effect is so small that the matrix is becoming difficult to invert, such as happens at large $|k_S a|$ but before the inertial limit is reached. The method is similar to the approach taken to obtain an estimate of the thermal contribution, giving an approximate solution, with an estimate of the error.

In this case, a pseudoinertial limit is defined, by solving the u_r and P_{rr} boundary equations (42) and (43) with zero shear wave mode coefficients, but including the viscous parts of the stress components w_{0n}, w_{cn}, w'_{cn}. These are negligible in the true inertial limit. The coefficients are:

$$A_{nNIS} = -\frac{[k_c a j'_n(k_c a) q'_{cn} - k'_c a j'_n(k'_c a) q_{0n}]}{[k_c a h'_n(k_c a) q'_{cn} - k'_c a j'_n(k'_c a) q_{cn}]}, \quad (78)$$

$$A'_{nNIS} = -\frac{[k_c a j'_n(k_c a) q_{cn} - k_c a h'_n(k_c a) q_{0n}]}{[k_c a h'_n(k_c a) q'_{cn} - k'_c a j'_n(k'_c a) q_{cn}]}. \quad (79)$$

In some cases, this value is a much better estimate of the full coefficient than is A_{nNI} at frequencies below the true inertial limit. However, this does not appear to be generally true, but the pseudoinertial coefficient was used in this method as the initial estimate.

The contribution of the shear wave modes can be estimated by calculating C_n, C'_n by solving the boundary equations for tangential velocity u_θ and the $P_{r\theta}$ stress component [Eqs. (44) and (45)], given the values for $A_n \approx A_{nNIS}$ and $A'_n \approx A'_{nNIS}$ presented earlier. Note that the thermal contribution is still neglected here. These values for C_n, C'_n can then be substituted back into the u_r and P_{rr} boundary equations (42) and (43) to determine the adjustment to A_n caused by the shear wave modes. Using Eqs. (42) and (43) the estimated shear contribution to A_n is as follows:

$$\Delta R_{nNS1} = n(n+1)(C_n - C'_n), \quad (80)$$

$$\Delta R_{nNS2} = -q_{Sn}C_n + q'_{Sn}C'_n, \quad (81)$$

$$\Delta A_{nNS} = \frac{[\Delta R_{nNS1} q'_{cn} - k'_c a j'_n(k'_c a) \Delta R_{nNS2}]}{[k_c a h'_n(k_c a) q'_{cn} - k'_c a j'_n(k'_c a) q_{cn}]}, \quad (82)$$

$$\Delta A'_{nNS} = \frac{[q_{cn} \Delta R_{nNS1} - k_c a h'_n(k_c a) \Delta R_{nNS2}]}{[k_c a h'_n(k_c a) q'_{cn} - k'_c a j'_n(k'_c a) q_{cn}]}, \quad (83)$$

and the estimated coefficient is given by

$$A_{nN} = A_{nNIS} + \Delta A_{nNS}. \quad (84)$$

The process can be reiterated to obtain an improved coefficient, by substituting the new coefficient back into Eqs. (44) and (45) to get a better estimate of the shear wave mode coefficients, followed by Eqs. (80)–(82) to find a more accurate value of ΔA_{nNS}. When the matrix becomes difficult to invert, this method can be used to obtain an estimate for the shear terms, and iterated if necessary to obtain greater accuracy. Hence this is a valid and useful solution when the matrix becomes numerically badly conditioned or singular. The method removes the difficulty of ill-conditioned matrix equations, allowing solutions to be obtained under all conditions without resorting to extremely high precision programming. The use of the scaling with the Bessel function value at the particle surface has also removed the need to calculated Bessel functions at large values of $k_S a$ or $k'_S a$ since only the ratios of derivatives are required. Limiting values for these have been presented in Sec. IV C.

VI. NUMERICAL RESULTS

The solutions detailed in the previous sections have been programmed using MATLAB (version 7.1) under Windows XP and 64 bit precision, an accuracy of 2×10^{-16}. Calculations of the coefficients were carried out for a model emulsion of hexadecane in water (with 1% Tween20). The physical prop-

TABLE I. Physical properties of *n*-hexadecane in water with 1% Tween20 at 20 °C (Ref. 14).

	Water+1% Tween20	*n*-hexadecane
Ultrasound velocity ($m\,s^{-1}$)	1485.5	1357.9
Density ($kg\,m^{-3}$)	999.5	773
Thermal expansivity (K^{-1})	0.00022	0.00091
Viscosity (Pa s)	0.00111	0.00334
Specific heat capacity ($J\,kg^{-1}\,K^{-1}$)	4182	2093
Thermal conductivity ($W\,m^{-1}\,K^{-1}$)	0.59	0.14
Attenuation exponent p	2	2
Attenuation factor ($Np\,m^{-1}\,MHz^{-p}$)	0.025	0.101

erties of the two components are given in Table I.[14] Where relevant, the volume fraction was 0.1.

A. Zero-order coefficient A_0

The zero-order coefficient includes nonthermal and thermal contributions (see Sec. V A). The new solution has been compared with the ECAH solution for different particle sizes 0.1–10 μm and frequency 0.01–1000 MHz. The match between the full ECAH matrix solution and the analytical results presented in Sec. V A was extremely accurate. The contribution of thermal effects is shown in Fig. 2 as a function of the thermal parameter k_Ta. The plot covers a wide range of k_ca values. For $a=0.1$ μm, $k_ca=4\times10^{-6}-4$ and $\mathrm{Re}(k_Ta)=0.047-47$. For $a=10$ μm, $k_ca=4\times10^{-4}-4$, $\mathrm{Re}(k_Ta)=4.7-470$. In this case, the coefficient is plotted in the form $A_{0T}/(k_ca)^3$, which is the parameter appearing in velocity and attenuation. The plots for the two different particle sizes overlay for a large proportion of the range. However, they diverge at the higher frequency (large $|k_Ta|$) end. In the long wavelength region, the parameter plotted $(A_{0T}/(k_ca)^3)$ is purely a function of the thermal parameter k_Ta, and therefore

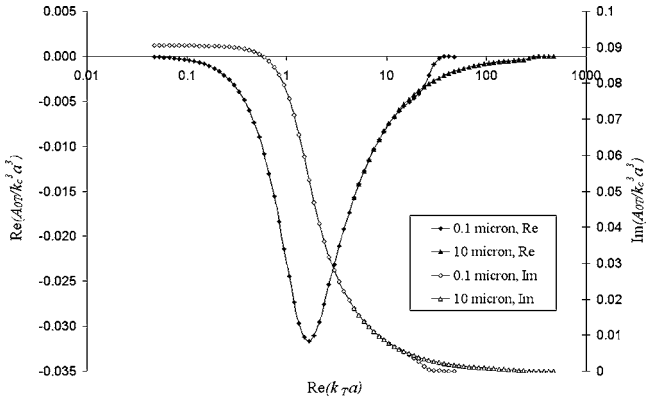


FIG. 2. Plot of the thermal part of the zero-order coefficient as a function of the thermal wave number radius product. This demonstrates the frequency dependence for two different particle sizes. The coefficient is scaled by $(k_ca)^3$ which is the term appearing in the wave number equation (20) for velocity and attenuation, and which removes the predominant frequency dependence in this range of $|k_ca|$ from $4\times10^{-6}-4$. The plots for different particle sizes overlay in the long wavelength region where $|k_ca|\ll1$. The real part relates predominantly to the attenuation, and the thermal contribution decreases to zero at both low and high frequencies. The imaginary part relates to velocity and although the thermal contribution decreases to zero at high frequencies, it reaches a constant value at the low frequency limit, giving an offset in velocity.

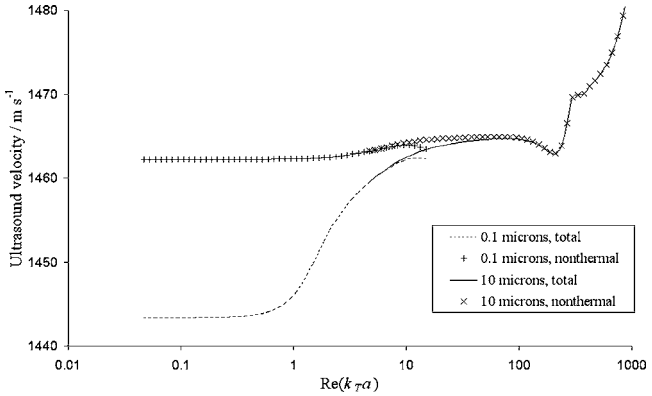


FIG. 3. Ultrasound velocity for a system of 10% (v/v) hexadecane in water including or excluding the thermal contribution in the zero-order coefficient A_0. Higher order coefficients are included and have been calculated by the standard method, but the velocity is calculated using single scattering only (terms up to first order in volume fraction). The velocity is plotted as a function of the thermal parameter $\mathrm{Re}(k_Ta)$ which demonstrates the frequency dependence for the different particle sizes. The plot covers the range $|k_ca|$ from $4\times10^{-6}-20$, and plots for the different particle sizes overlay where $|k_ca|\ll1$. The thermal contribution to velocity becomes negligible at high frequencies and reaches a constant value at low frequency.

the results for different particle sizes (and therefore different k_ca values) are the same. Outside that region, the value of k_ca affects the result, and therefore the plots do not overlay. The point where the plots differ is the limit of the long wavelength region for the smallest particle size plotted.

Figure 2 shows that the real part of the thermal part of the coefficient decreases to zero at both ends of the frequency range, with a maximum around $|k_Ta|\sim1$. The real part of the scattering coefficient A_n is the dominant contribution to attenuation, and the thermal effects are therefore negligible at both low and high frequency limits. However, the imaginary part of the thermal coefficient decreases to zero at the high frequency limit (large $|k_Ta|$) but reaches a constant nonzero value as the frequency decreases (low $|k_Ta|$). The imaginary part of the coefficient is the dominant contribution to the ultrasound velocity, resulting in an offset in velocity in the low frequency (or small particle size) limit. At high frequency, the thermal contribution to velocity is again negligible. It should be noted that the thermal coefficient is affected by the thermal wave number inside the particle as well as outside.

Solutions for the large values of $|k_Ta|$ can easily be calculated in the new model. With the ECAH model, problems would be encountered for systems in which the particle size is large, and therefore $|k_Ta|$ is large at experimental frequencies. Although the thermal contribution becomes negligible in the high frequency limit, the present method allows a means of estimating the thermal contribution without attempting to evaluate Bessel functions with large arguments, or solve an ill-conditioned matrix equation. As the thermal contribution becomes negligible the full matrix solution becomes nearly singular. It is now possible to estimate the thermal effect to *demonstrate* that it can be neglected.

The contribution of thermal effects to the ultrasound velocity and attenuation is shown in Figs. 3 and 4 for the radii of 0.1, 1, and 10 μm, and a frequency range of 0.01–1000 MHz, covering the range $|k_ca|$ from $4\times10^{-6}-20$. The result

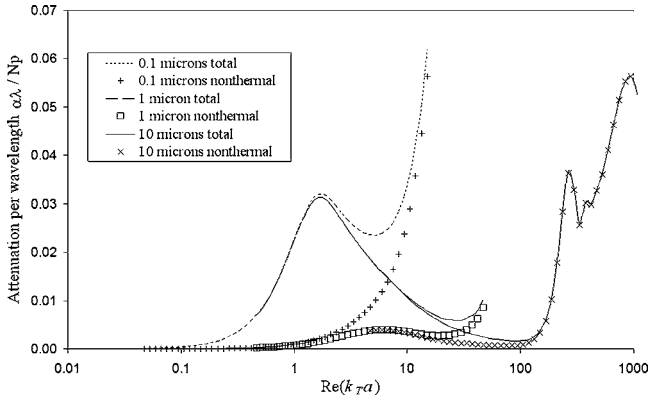


FIG. 4. Ultrasound attenuation per wavelength for a system of 10% hexadecane in water including or excluding the thermal contribution. Higher order coefficients are included and have been calculated by the standard method, but the attenuation is calculated using single scattering only (terms up to first order in volume fraction). The attenuation is plotted as a function of the thermal parameter $\mathrm{Re}(k_T a)$, which demonstrates the frequency dependence for the different particle sizes. The plot covers the range $|k_c a|$ from 4×10^{-6}–20. Plots for the different particle sizes overlay only where $|k_c a| \ll 1$ and the intrinsic attenuation is negligible. The thermal contribution to attenuation gives a peak in $\alpha\lambda$ and at large frequency [large $\mathrm{Re}(k_T a)$] the thermal contribution becomes negligible so that the nonthermal results approach the full result. The sharp increase in attenuation per wavelength at high frequencies is caused by the intrinsic attenuation (which varies as f^2).

for velocity and attenuation per wavelength ($\alpha\lambda$) are again plotted as a function of $\mathrm{Re}(k_T a)$. Note that the attenuation per wavelength is the attenuation of the solution (using single scattering only) multiplied by the wavelength of sound in water. The plots include the results including or excluding the thermal part of the zero-order coefficient A_0. Higher order coefficients are included and have been calculated by the standard method, but the velocity is calculated using single scattering only (terms up to first order in volume fraction). The plots of velocity for the two particle sizes overlay in the long wavelength region as before. The features in the plot at high $|k_T a|$ for the large particle size occur when $k_c a \approx 1$ and the variation with frequency becomes complicated. Figure 3 shows that the thermal component causes a large shift in ultrasound velocity in the low frequency limit, but as the frequency (and $k_T a$) increases, the thermal contribution becomes negligible. The attenuation per wavelength plots overlay in the long wavelength region when the intrinsic attenuation contribution is negligible. An additional plot for $a = 1$ μm has been included to demonstrate the agreement of the lower part of the curves (small $|k_T a|$) for all particle sizes. However, as the frequency increases there are two reasons for the plots to differ for different particle sizes. First, the limit of the long wavelength region occurs at smaller values of $|k_T a|$ for smaller particle sizes, and outside this region there is dependence on $k_c a$ as well as the thermal parameter $|k_T a|$. Second, as the frequency increases, the intrinsic attenuation becomes significant causing the large increase in attenuation seen at around $\mathrm{Re}(k_T a) \approx 5$ for $a = 0.1$ μm, and $\mathrm{Re}(k_T a) \approx 200$ for $a = 10$ μm. Hence these attenuation plots do not overlay over the full range plotted here.

The thermal contribution to attenuation is negligible at both low and high frequencies (small or large $|k_T a|$), with a peak in the midrange. The peak occurs at around $\mathrm{Re}(k_T a) \approx 1$. The strong increase in attenuation due to intrinsic attenuation truncates the thermal peak to a differing degree for each particle size. The nonthermal contribution to the attenuation per wavelength also has a peak, occurring at around $\mathrm{Re}(k_T a) \approx 6$, for the two larger particle sizes. This cannot be caused by thermal effects because it is in the nonthermal contribution to the scattering coefficient, but it varies with the thermal wave parameter $k_T a$. The peak is related to the maximum loss due to viscous and inertial effects, appearing through the stress parameter q_{cn} and corresponding terms. These are determined by the shear wave parameter $k_S a$ which has the same frequency and particle size dependence as $k_T a$, and hence the two particle sizes show overlaying peaks.

B. A_n $n>0$

Calculations for orders $n>0$ were carried out for a system similar to hexadecane in water. It was required to determine the behavior of the shear contribution as a function of the parameter $k_S a$. This was carried out first for the same sets of parameters as used in the previous section, for hexadecane in water. However, a broader range of values of $k_S a$ and $k_c a$ was required in order to establish the general conditions under which the shear contribution can be neglected, or estimated. Calculations in which the frequency is varied for a given particle size result in large $k_S a$ when $k_c a$ is large, and vice versa, and the general case is not easily determined. Therefore a set of data was calculated for fixed values of $k_c a$, for each of which a range of values of $k_S a$ were used.

For each selected value of $k_c a$ a set of particle radius and frequency values were chosen to produce the relevant $\mathrm{Re}(k_c a)$ value, but for a range of $k_S a$ values. In order to achieve this, the viscosity in the two phases was adjusted, so that an appropriate range of $k_S a$ values could be obtained for very different $k_c a$ values. The parameters are limited by a number of conditions:

(1) The attenuation must be small compared with the real part of the wave number and
(2) The thermal limit is satisfied $|k_c^2/k_T^2| \ll 1$.

These are approximately satisfied for frequencies

$$f \leqslant f_{\max} \approx 10 \text{ GHz} \tag{85}$$

in water, and calculations were limited to this region. The smallest particle size corresponding to a given frequency with a set value of $k_c a$ is therefore

$$a = \frac{(k_c a)v}{2\pi f_{\max}}. \tag{86}$$

Equation (86) defines a minimum particle size which can be used at the maximum frequency [Eq. (85)] to obtain a fixed value of $k_c a$. The shear wave parameter $k_S a$ is related by

TABLE II. Viscosity values used for various values of $k_c a$.

$k_c a$	η (Pa s)	η' (Pa s)
1	0.0111	0.0334
5	1.11	3.34
20	1.11	1.2
50	30	35
100	30	35

$$|k_S a| = \sqrt{\frac{\omega\rho}{\eta}} a = k_c a v \sqrt{\frac{\rho}{2\pi f \eta}} \tag{87}$$

so at larger values of $k_c a$ the smallest value of $k_S a$ obtainable is limited by the maximum frequency. In order to explore lower $k_S a$ values, the viscosity has been increased appropriately. The values of viscosity used are given in Table II. All other properties for the system are those given in Table I.

1. Thermal contribution

In most cases the thermal contribution to the coefficients for orders $n>0$ can be neglected. The percentage error in the real and imaginary parts of the coefficients A_n were calculated over a wide range of conditions. The percentage errors were found to be mostly less than 1%, although in some cases a much larger error was found. These related to coefficients which were themselves small compared with the coefficients for other orders under the same conditions. The coefficients for different orders with the same parameters can be very different in magnitude, so that a large relative error in an order which has a very small coefficient can be tolerated. Conversely, a small relative error is required for the most significant coefficients. Hence the percentage error of individual coefficients was not a good measure of accuracy.

A measure of the significance of the coefficients to velocity and attenuation is obtained by comparison with the single scattering contribution

$$\sum_n (2n+1)A_n.$$

The thermal contribution to this term is more relevant to velocity and attenuation calculations. The percentage error in the summation caused by neglecting thermal terms is mostly small $<1\%$ in both real and imaginary parts, but there are instances where the error increases to a significant value, up to 5%. The sum consists, in some cases, of a sum of large numbers of oscillating sign, which largely cancel out. This proves to be rather inaccurate. When there are a large number of orders to be included (when $k_c a$ is large) the factor $(2n+1)$ amplifies any error in the largest orders so that the accuracy of these coefficients must be higher. Thus it is advisable to use the estimate of the thermal effect rather than neglecting the thermal effect altogether, to ensure the accuracy of the solution. Two iterations of the estimate are usually sufficient, but further iterations can be used until the required accuracy is achieved for the particular coefficient. Using the estimated thermal contribution, the solution was indistinguishable from the ECAH solution over the full range where calculation was possible.

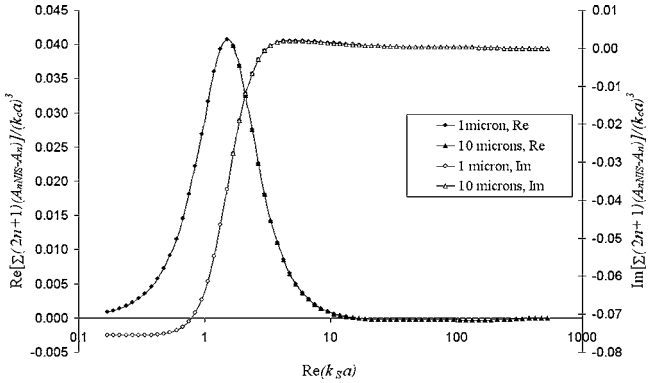


FIG. 5. The relative contribution of shear waves to the velocity and attenuation as a function of the shear parameter $\mathrm{Re}(k_S a)$, which shows frequency dependence for different particle sizes. The shear contribution is quantified by the difference in the sum $\Sigma_n(2n+1)A_n/(kc)^3$ including or excluding shear waves. The imaginary part of this sum appears in the equation for the velocity of the dispersion and its real in the attenuation [Eqs. (21) and (22)]. The effect of shear waves on attenuation becomes negligible at both low and high frequency limits (the real part of the plotted difference becomes zero). At high frequency, the shear waves also make a negligible contribution to velocity (the imaginary part in the plot), but their effect reaches a constant value at the low frequency limit. The plot covers the range $|k_c a|$ from 4×10^{-6}–40 and plots for the different particle sizes overlay in the long wavelength region.

2. Shear contribution

The contribution made by the shear wave modes to the scattering coefficient was first investigated for the standard hexadecane in water system, whose properties are listed in Table I. Scattering coefficients were calculated for a number of particle sizes over a frequency range 0.01–100 MHz, either including or excluding shear wave modes from the calculation. The parameter used to demonstrate the significance of the shear wave contribution was again the sum $\Sigma_n(2n+1)A_n$, taking the difference between the sum for the full coefficient A_n and for the coefficient A_{nNIS} which excludes shear wave modes. The real and imaginary parts of this shear wave contribution are plotted in Fig. 5 as a function of the shear wave parameter $\mathrm{Re}(k_S a)$. The real part of the shear wave contribution, which predominantly affects attenuation, falls to zero at both ends of the frequency range. The imaginary part, which predominantly affects ultrasound velocity, reaches a constant value as the frequency decreases, but at high frequencies its value falls to zero.

At large values of $|k_S a|$ the shear wave mode contribution, therefore, decreases to zero in both the real and imaginary parts of the coefficient (i.e., its effect on both velocity and attenuation vanishes at high frequency or for large particles). Hence the coefficient approaches the pseudo-inertial limit A_{nNIS} at high frequencies. The true inertial coefficient A_{nNI} is also approached when the viscous contribution to the stress terms [Eq. (26)] become negligible. As an example, Figs. 6(a) and 6(b) show the real and imaginary parts of a coefficient for $k_c a=20$, $n=5$, demonstrating the approach to the inertial limit. Also shown is the calculation estimating the shear wave contribution, using Eqs. (78) and (82), well before the inertial limit is reached, but when the shear effect is

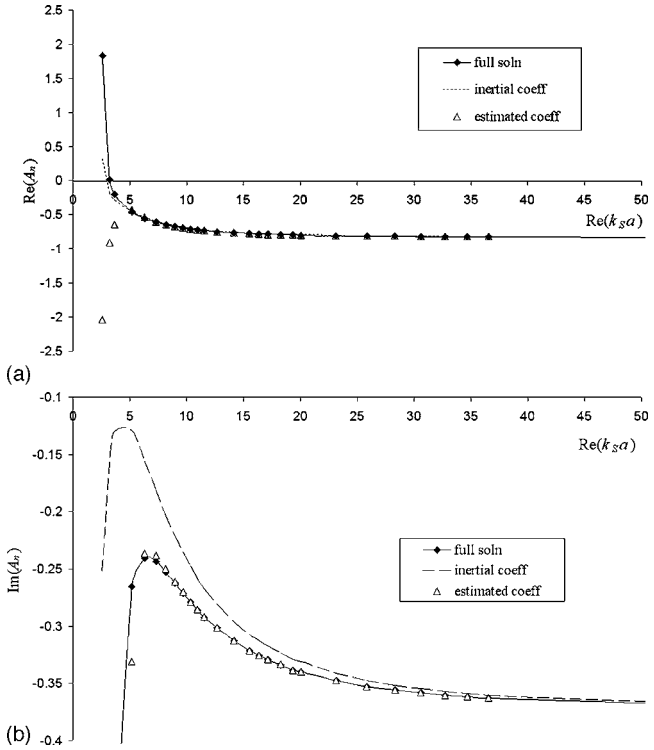


FIG. 6. The coefficient A_n plotted as a function of $\mathrm{Re}(k_S a)$ for $k_c a = 20$, $n = 5$, showing the full calculations, the inertial limit and the coefficient calculated by estimating the small shear contribution when $\mathrm{Re}(k_S a)$ is large; (a) the real part (which predominantly affects attenuation) and (b) the imaginary part (which predominantly affects velocity). The plots demonstrate that at very large values of $\mathrm{Re}(k_S a)$ the coefficient approaches the inertial limit. However, the estimated shear contribution provides an accurate estimate of the coefficient at much lower values of $\mathrm{Re}(k_S a)$. The plot covers a range of particle size and frequency which combine to give a fixed value $k_c a = 20$. Having fixed $k_c a$, an increasing value of $\mathrm{Re}(k_S a)$ corresponds to larger particle size, and lower frequency. The viscosity was set to be approximately 1000× its usual value in both phases (see the text for details) in order to cover a wide range of parameter space.

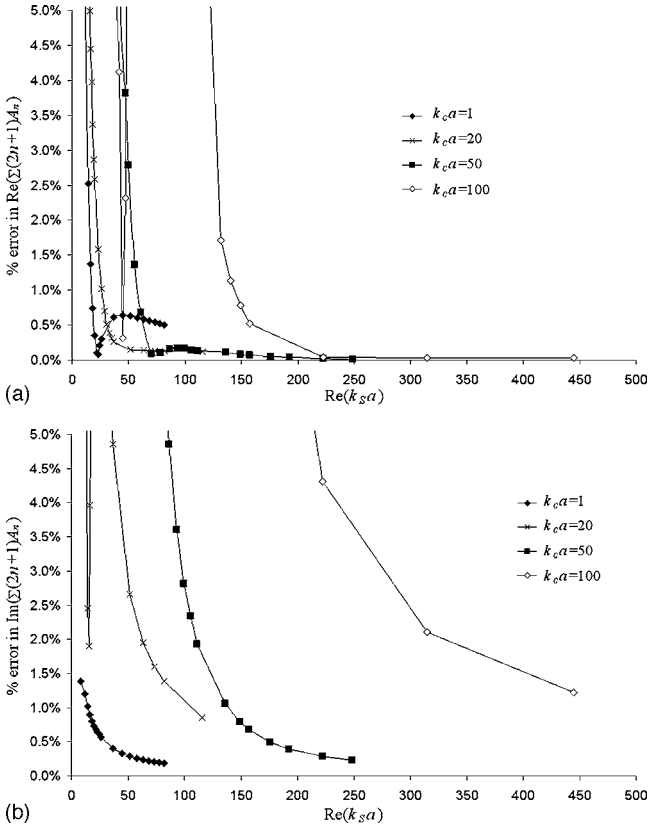


FIG. 7. The percentage error in the (a) real and (b) imaginary parts of the sum over coefficients $\Sigma_n(2n+1)A_n$ calculated for the inertial limit and for the full coefficients. The real part of this sum relates to attenuation and its imaginary part to velocity. The purpose of the plot is to determine the limit at large $\mathrm{Re}(k_S a)$ at which the inertial coefficients become a satisfactory estimate for the full coefficients, i.e., when the plotted error becomes small. This limit is reached at larger values of $\mathrm{Re}(k_S a)$ when $k_c a$ is also large. The viscosity of the components was varied significantly to explore the parameter space.

small. The estimate is poor at low values of $|k_S a|$ but shows very accurate agreement with the full coefficient at higher $|k_S a|$.

It is now necessary to determine the conditions under which the inertial limit is reached. For this purpose a set of calculations were carried out for a set of fixed values of $k_c a$, for each of which a range of values of $k_S a$ were calculated. The parameter $\Sigma_n(2n+1)A_n$ was calculated, taking the difference between the full coefficient and the pseudoinertial limit A_{nNIS}, as for Fig. 5. This time, the difference is expressed as percentage error in the sum for the full coefficient. Figure 7 therefore shows the percentage error in the real and imaginary parts of the parameter $\Sigma_n(2n+1)A_n$ caused by neglecting the shear wave modes. The percentage error approaches zero as $|k_S a|$ increases, but as $k_c a$ gets larger, the inertial limit is reached at a larger value of $|k_S a|$. It is also worth noting that the error is not monotonically decreasing, and a deceptively low value of the percentage error can be achieved at much smaller values of $|k_S a|$ than the true inertial limit, but the error subsequently rises again before falling to zero. This can be seen, for example in Fig. 7(a) for $k_c a = 100$, which has a minimum when $\mathrm{Re}(k_S a) \approx 50$. However, the error does not finally fall below 1% until $\mathrm{Re}(k_S a) \geqslant 150$.

An approximate relationship was obtained to define the inertial limit, where the error in both real and imaginary parts were found to be less than 1% for

$$|k_S a| \geqslant 5 k_c a,$$

when $5 < k_c a < 100$. The condition may be related to the relative contribution of inertial and viscous terms to the stress, Eqs. (26) and (27). This is a very approximate result, since only a few data points were obtained. Each data point requires the calculation of coefficients for a fixed value of $k_c a$ over a large range of $|k_S a|$ to identify the condition. When $k_c a$ is large, the shear effects continue to make a significant contribution for very large values of the shear parameter $|k_S a|$. Hence the methods presented here both for calculating the Bessel function ratios at large arguments and for estimating the shear contribution when it is very small are extremely valuable.

VII. CONCLUSIONS

A number of methods have been presented which ease the numerical calculation of scattering coefficients in the ECAH model. First, a straightforward scaling of the thermal and shear wave modes by the respective Bessel or Hankel

function (Heine function) evaluated at the boundary caused a significant improvement in the conditioning of the boundary equations. The scaling also removed the need to calculate directly the Heine functions at large arguments, since only ratios of the function derivatives with the function itself then appeared in the boundary conditions. Limiting values of these ratios at large arguments have been evaluated.

Further improvement of the solution was achieved by noting that under the condition $|k_c^2/k_T^2| \ll 1$ the thermal terms can be neglected in some of the boundary equations. The approximation is valid for frequencies up to around 100 GHz in water. The fact that the thermal terms are negligible in some boundary equations implies their inclusion would cause near-singularity. For the zero-order coefficient, the use of this approximation, and the separation of the coefficient into thermal and nonthermal contributions produced an analytical result for the coefficient. Thus the thermal contribution can be readily estimated to determine whether it is significant for the system of interest. For the higher order coefficients the neglect of thermal terms improves matrix conditioning dramatically because the thermal contribution is small. A method is also presented for estimating the magnitude of the neglected thermal contribution either to improve accuracy or to confirm the validity of the approximation.

For the nonzero orders, even after thermal terms are neglected, the boundary equations can still become badly conditioned at large frequencies. In this region, the shear wave modes become insignificant, and this causes near-singularity of the matrix. The coefficient was separated into the inertial part (which is the limiting high-frequency value) and an additional shear contribution. This contribution can be estimated in a similar way to the thermal estimate. Thus the coefficient can be obtained in cases where previously the matrix solution was badly conditioned. It has been demonstrated that the inertial limit is reached at higher values of $|k_S a|$ (the shear wave parameter) when the value of $k_c a$ is large. Thus it may be necessary to evaluate the shear contribution for large values of $|k_S a|$ when $k_c a$ is also large since the shear contribution is still non-negligible. The method for calculating the Bessel/Hankel functions at large arguments are particularly relevant here.

In summary, the work presented here has achieved a number of improvements in the calculations for the ECAH scattering model. The conditioning of the matrix equations has been improved by introducing scaling factors for the thermal and shear wave potentials, and by neglecting thermal and shear mode contributions where appropriate. Further improvement has been obtained by developing a method for calculating the required Bessel functions. For the zero-order coefficient an analytical result has been obtained with separate nonthermal and thermal parts. These developments enable the model to be adopted on standard precision computers, rather than as a research tool on high precision machines, and potentially to be incorporated within instrumentation. Previously, such implementation has been limited to the long wavelength region where analytical results were available.

ACKNOWLEDGMENT

The author would like to thank Professor Malcolm Povey of the University of Leeds for his support of this work and for helpful comments on the manuscript.

NOMENCLATURE

A, B, C = scattering coefficients of compressional, thermal, and shear modes, respectively
a = particle radius
b_1, b_2, b_3, c_1 = parameters for the analytical solution of the zero-order scattering coefficient
C_p = specific heat capacity at constant pressure
D = denominator for the analytical solution for the zero-order scattering coefficient
d = scaled stress components $P_{r\theta}$
f = frequency
$f(\theta)$ = far-field angular distribution of sound field
j_n, h_n = spherical Bessel and Hankel functions, respectively, of the nth order
k_c, k_T, k_S = wave numbers of compressional, thermal, and shear wave modes, respectively
K = wave number of dispersion
$\bar{\bar{L}}$ = matrix to be inverted to obtain scattering coefficients
$n\mathrm{Max}$ = maximum number of partial wave orders required for convergence
P_n, P_n^1 = Legendre polynomials, associated Legendre polynomials
P_{ij} = components of stress tensor
q = scaled stress components P_{rr}
$\bar{\bar{R}}$ = matrix for the right-hand side of the scattering coefficient equation
T = temperature
t = time
$\mathbf{u}$ = (fluid) velocity
v = sound speed (of compressional mode)
w = scaled viscous contributions to stress component P_{rr}
α = attenuation (of compressional mode)
β = thermal expansivity
ϕ = a scalar velocity potential; volume fraction
φ, ψ, χ = wave potential for compressional, thermal and shear modes respectively
Γ = thermal factor
$\gamma = C_p/C_v$ = ratio of specific heat capacities
η = shear viscosity
$\hat{\eta} = \eta'/\eta$ = ratio of viscosities
λ = wavelength
ρ = density
$\hat{\rho} = \rho'/\rho$ = ratio of densities
$\sigma = \tau/(\rho C_p)$ = thermal diffusivity
τ = thermal conductivity
$\hat{\tau} = \tau'/\tau$ = ratio of conductivities
$\upsilon = \eta/\rho$ = kinematic viscosity
ω = angular frequency

Subscripts and superscripts

0 = (on wave potentials) incident wave

c, T, S = compressional, thermal, shear wave mode, respectively
$'$ = dispersed phase (particle)
j_n', j_n'', h_n', h_n'' = first and second-order derivatives of Bessel or Hankel function, respectively
n = partial wave order
$\hat{x}$ = ratio of dispersed phase to continuous phase for property x, e.g., $\hat{\rho}=\rho'/\rho$
I = inertial part (neglecting shear waves and viscous contributions to stress)
IS = pseudoinertial part (neglecting shear waves but including viscous stress)
N, T = nonthermal and thermal parts
S = viscous contribution (in addition to pseudoinertial limit)
V = viscous contribution (in addition to inertial limit)
soln = relating to the solution

[1]R. E. Challis, M. J. W. Povey, M. L. Mather, and A. K. Holmes, "Ultrasound techniques for characterizing colloidal dispersions," Rep. Prog. Phys. **68**, 1541–1637 (2005).
[2]J. W. Strutt (Baron Rayleigh), *The Theory of Sound*, 2nd ed. (Macmillan, London, 1896).
[3]P. S. Epstein and R. R. Carhart, "The absorption of sound in suspensions and emulsions. I. Water fog in air," J. Acoust. Soc. Am. **25**, 553–565 (1953).
[4]J. R. Allegra, and S. A. Hawley, "Attenuation of sound in suspensions and emulsions: Theory and experiments," J. Acoust. Soc. Am. **51**, 1545–1564 (1972).
[5]C. F. Ying and R. Truell, "Scattering of a plane longitudinal wave by a spherical obstacle in an isotropically elastic solid," J. Appl. Phys. **27**, 1086–1097 (1956).
[6]R. E. Challis, J. S. Tebbutt, and A. K. Holmes, "Equivalence between three scattering formulations for ultrasonic wave propagation in particulate mixtures," J. Phys. D **31**, 3481–3497 (1998).
[7]O. G. Harlen, M. J. Holmes, M. J. W. Povey, Y. Qiu, and B. D. Sleeman, "A low frequency potential scattering description of acoustic propagation in dispersions," SIAM J. Appl. Math. **61**, 1906–1931 (2001).
[8]T. J. O'Neill, J. S. Tebbutt, and R. E. Challis, "Convergence criteria for scattering models of ultrasonic wave propagation in suspensions of particles," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **48**, 419–424 (2001).
[9]A. S. Dukhin and P. J. Goetz, *Ultrasound for Characterizing Colloids: Particle Sizing, Zeta Potential, Rheology* (Elsevier, Amsterdam, 2002).
[10]M. J. W. Povey, *Ultrasonic Techniques for Fluids Characterization* (Academic, San Diego, 1997).
[11]P. Lloyd and M. V. Berry, "Wave propagation through an assembly of spheres. IV. Relations between different multiple scattering theories," Proc. Phys. Soc. London **91**, 678–688 (1967).
[12]P. C. Waterman and R. Truell, "Multiple scattering of waves," J. Math. Phys. **2**, 512–537 (1961).
[13]J. G. Fikioris and P. C. Waterman, "Multiple scattering of waves. II. 'Hole correction' in the scalar case," J. Math. Phys. **5**, 1413–1420 (1964).
[14]D. J. McClements and J. N. Coupland, "Theory of droplet size distribution measurements in emulsions using ultrasonic spectroscopy," Colloids Surf., A **117**, 161–170 (1996).

Is the Kramers-Kronig relationship between ultrasonic attenuation and dispersion maintained in the presence of apparent losses due to phase cancellation?

Adam Q. Bauer, Karen R. Marutyan, Mark R. Holland, and James G. Miller
Department of Physics, Washington University in Saint Louis, Saint Louis, Missouri 63130

(Received 25 January 2007; revised 29 March 2007; accepted 29 March 2007)

Phase cancellation effects can compromise the integrity of ultrasonic measurements performed with phase sensitive receiving apertures. A lack of spatial coherence of the ultrasonic field incident on a phase sensitive receiving array can produce inaccuracies of the measured attenuation coefficient and phase velocity. The causal (Kramers-Kronig) link between these two quantities in the presence of phase distortion is investigated using two plastic polymer materials, Plexiglas™ and Lexan™, that exhibit attenuation coefficients that increase linearly with frequency, in a fashion analogous to that of soft tissue. Flat and parallel plates were machined to have a step of a thickness corresponding to an integer number of half wavelengths within the bandwidth investigated, 3 to 7 MHz. Insonification of the stepped portion of each plate produces phase cancellation artifacts at the receiving aperture and, therefore, in the measured frequency dependent attenuation coefficient. Dispersion predictions using two different forms of the Kramers-Kronig relations were performed for the flat and the stepped regions of each plastic plate. Despite significant phase distortion and a detection system sensitive to these aberrations, the Kramers-Kronig link between the apparent attenuation coefficient and apparent phase velocity dispersion remains intact. © *2007 Acoustical Society of America.* [DOI: 10.1121/1.2735803]

PACS number(s): 43.35.Cg, 43.80.Cs, 43.20.Hq, 43.35.Bf [FD] Pages: 222–228

I. INTRODUCTION

Artifacts arising from phase cancellation effects at the face of phase sensitive piezoelectric receiving transducers were well known to the physical acoustics community by the 1960s.[1–3] From the 1960s through the 1980s, phase aberration artifacts in medical ultrasound were investigated by a number of laboratories[4–10] and remain a topic of research today. Distinct from interference effects in the ultrasonic field, which represent lossless redistributions of energy, signal loss at a phase sensitive receiving aperture is an irreversible loss of information.

The causality-induced relationship between the frequency dependence of attenuation and phase velocity represented by the Kramers-Kronig relations have been investigated in the context of biomedical ultrasound for a number of years.[11–23] The purpose of the present paper is to examine the validity of the Kramers-Kronig relations under conditions in which phase cancellation at a phase sensitive aperture introduces an apparent attenuation. In the context of medical imaging and tissue characterization, aberrations arising from tissue inhomogeneities that represent local variations in phase velocity are often the source of such artifacts.[24–34]

Information lost as a consequence of residual phasing errors remaining after focusing the signals arising from individual array elements is the source of the apparent signal loss of interest here. We explore the question of whether the causality-induced link between attenuation and dispersion remains valid when the signal loss arises as a consequence of inadequacies in the process of receiving the distorted ultrasonic field rather than from intrinsic losses within the medium under study.

In order to explore this question, we measured the frequency-dependent apparent attenuation and phase velocity in media that exhibit an approximately linear with frequency attenuation coefficient and logarithmic with frequency phase velocity. Flat and parallel plates of Plexiglas™ (polymethyl methacrylate) and Lexan™ (polycarbonate resin thermoplastic) were machined to have a small step discontinuity of a thickness corresponding to an integer number of half wavelengths on one of their flat and parallel sides. The thicker, thinner, and stepped regions of these plates were interrogated using a matched pair of single element planar transducers. The resulting apparent attenuation coefficient and phase velocity values were compared with corresponding predictions from the Kramers-Kronig relations.

II. EXPERIMENTAL PROCEDURE

A. Sample preparation

The samples investigated were machined from pieces of polymethyl methacrylate (Plexiglas™, Arkema Inc., Philadelphia, PA) and polycarbonate resin thermoplastic (Lexan™, AIN Plastics, Madison Heights, MI). Both plastics were measured to verify that they exhibit approximately linear with frequency attenuation coefficients. These two materials were chosen because their coefficients of attenuation differ by nearly an order of magnitude.[19,21,35] Each sample was machined into a flat and parallel plate 2-in. long, 1-in. wide,

 0001-4966/2007/122(1)/222/7/$23.00

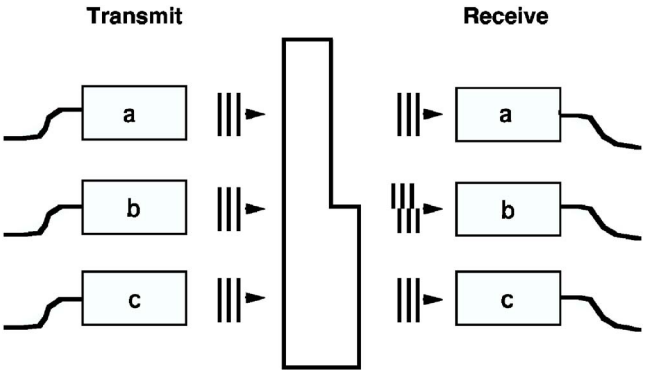


FIG. 1. The three regions of the stepped plastic plate interrogated to demonstrate the effects of phase cancellation at the face of a phase sensitive piezoelectric receiver. Measurements (a) and (c) represent the flat and parallel regions, whereas (b) represents the stepped region.

and 0.5-in. thick. The resulting plates then had a small step discontinuity milled into one of their surfaces, as seen in Fig. 1. The step sizes in each plate were chosen using the criteria described below.

For a bounded plane wave normally incident and transmitted through a flat and parallel plate except for a small step discontinuity present on one of its surfaces, the portion of the plane wave transmitted through the thinner part of the plate will be out of phase with the portion of the beam transmitted through the thicker part of the plate. The phase difference between the two portions of the beam is expressed by the equation

$$\Delta\varphi = (k_w - k_s)\Delta x_s \tag{1}$$

where k_w and k_s are the wave numbers of the water and sample, respectively, and Δx_s is the step thickness. If this phase shift is an odd multiple of pi radians, the two portions of the incident ultrasonic field will approximately cancel when summed across the face of a phase sensitive receiving aperture. Therefore, detection with a receiving transducer sensitive to these phase differences will result in a loss of information from the ultrasonic field. By selecting the step size, the frequencies at which these cancellations occur, f_c, are chosen by using Eq. (1) with the condition that $\Delta\varphi = (2n-1)\pi$,

$$f_c = \frac{(2n-1)\pi}{2\pi\left(\frac{1}{c_w} - \frac{1}{c_s}\right)\Delta x_s}, \tag{2}$$

where $(2n-1)$ is an odd integer number of half wavelengths, and c_w and c_s are the speeds of sound in the water and sample, respectively, and $n=1,2,3$, etc. For phase cancellation at the face of a phase sensitive receiver to occur at approximately 5 MHz, one Plexiglas plate was milled to have a 0.3 mm step on one of its flat surfaces, whereas the Lexan plate was milled with a 0.4 mm step. To observe multiple frequencies cancelled within the experimental bandwidth, a separate Plexiglas sample was milled to have a step size of approximately 3 mm.

B. Equipment and experimental setup

Through-transmission measurements were performed in a water tank ($T \approx 21\ ^\circ C$) containing separate transmitting and receiving 5 MHz center frequency, 0.25-in. diam single element planar immersion transducers (Panametrics V310, Waltham, MA). These two transducers were secured to the bottom of the water tank at twice the calculated natural focal length, $D^2 f/4c_w$, of each transducer at 5 MHz, the natural (diffractive) focus of the transducer, where D is the aperture diameter, f is the frequency of interest, and c_w is the speed of sound in water.[36] For a sample exhibiting a higher speed of sound than the host medium placed between the transmitting and receiving transducers, the diffraction pattern is shifted axially towards the transmitting transducer relative to that of the reference path on receive. To minimize these phase distortion effects, the receiving transducer could be moved towards the transmitting transducer to compensate for the resulting change in the diffraction pattern. However, because such phase aberration errors are the subject of this investigation, these translations were not carried out.

The transmitting transducer was excited by a broadband radio frequency (RF) pulse from a pulser/receiver (Panametrics 5800, Waltham, MA) in through-transmission mode. The output of the receiving transducer was passed through the receiving circuitry of the Panametrics 5800 and fed into the 50 Ω input of an eight-bit digitizing oscilloscope (Tektronix 5052b, Beaverton, OR). The resulting time-domain broadband signals were digitized at a 2.5 GS/s digitization rate (400 ps/pt) and averaged 256 times in the time domain to improve the signal to noise ratio. Each plastic plate was fastened in succession to a custom designed mount attached to a four-axis motion controller (Aerotek Unidex 12, Pittsburgh, PA) and placed between the overlapping focal zones of the transmitting and receiving transducers. Motion control was operated via GPIB interface using custom-built software (Igor Pro, WaveMetrics, Inc., Lake Oswego, OR) operated by a personal computer (PowerMac G4, Apple Computer, Inc., Cupertino, CA).

C. Measurements

The phase velocity and attenuation properties of each plastic were measured by utilizing a substitution method. A water path reference trace was acquired prior to every sequence of sample path traces in order to provide a method of compensating for the electromechanical responses of the transducers and associated electronics.[37] Figure 1 illustrates the three regions of each plate inspected for this study. The acquired data were transferred via GPIB to a PowerMac G4 computer for offline analysis.

III. DATA ANALYSIS

The frequency dependence of the attenuation and phase velocity for each plastic specimen was determined by a broadband magnitude and phase spectroscopy RF data reduction technique.[38–40]

A. Phase velocity

The frequency response of the reference signal at the receiving aperture can be expressed as

$$\tilde{U}_w(x=L,\omega)=[A_0(\omega)\cdot\exp(-\alpha_w(\omega)L)]\exp\left(i\frac{\omega}{c_w(\omega)}L\right)$$

$$=|\tilde{U}_w(x=L,\omega)|\exp(i\varphi_w(\omega)), \tag{3}$$

where $A_0(\omega)$ is the initial pressure amplitude of each frequency component, α_w is the ultrasonic attenuation coefficient within the water bath, L is the length of the reference path, and $c_w(\omega)$ is the phase velocity within the water. The magnitude of the response is related to the attenuation coefficient of the water. The reference phase spectrum, $\varphi_w(\omega)$, contains information about the phase velocity and, therefore, the dispersion properties of the water.

For a path containing the sample material, the complex response can be written as

$$\tilde{U}_s(x=L,\omega)=A_0(\omega)T^P_{ws}(\omega)T^P_{sw}(\omega)H(\omega)\cdot\exp(-\alpha_w\cdot(L-d))$$
$$\times\exp\left(i\frac{\omega}{c_w(\omega)}(L-d)\right), \tag{4}$$

where $T^P_{ws}(\omega)$ and $T^P_{sw}(\omega)$ represent the pressure transmission coefficient from the water-sample and sample-water interfaces, respectively, d is the thickness of the sample, and $H(\omega)$ is the transfer function of the sample:

$$H(\omega)=\exp(-\alpha_s(\omega)d)\exp\left(i\frac{\omega}{c_s(\omega)}d\right). \tag{5}$$

It can be seen from (5) that the magnitude and phase of the transfer function contain information about the attenuation properties and the speed of propagation of planes of constant phase within the sample, respectively. Because the thicknesses of each sample were much larger than the length of the interrogating ultrasonic pulse, the confounding effects of the frequency dependences of the transmission coefficients commonly associated with very thin plates were not significant.[41,42]

The phase spectra for the reference and through-sample RF signals were unwrapped and compensated for the time shifts associated with the trigger delay and positive peak symmetrization.[17] The phase velocity of the sample was calculated by comparing the phase spectrum of the sample trace with that of the reference. The difference in the phase between these two spectra, $\Delta\varphi$, determines the frequency dependent phase velocity,

$$c_s(\omega)=c_w\cdot\left[\frac{\omega d}{\omega d+c_w\cdot\Delta\varphi(\omega)}\right], \tag{6}$$

where c_w is the known speed of sound in the water.

For the uniform thin or thick regions of the plate, the sample thickness traversed by the ultrasonic field is well defined. However, the effective sample thickness of the stepped portions of each plate is less clear, and was taken to be the average thickness of the thin and thick regions of the sample for approximately half of the incident ultrasonic field traversed the thinner part of that plate, while the other half of the ultrasonic field traversed the thicker portion of the plate. Although the choice of sample thickness minimally affects the frequency dependence of the phase velocity, this thickness significantly determines the overall level of the phase velocity shift, as seen in Sec. IV B.

B. Attenuation

The attenuation coefficient of the sample was obtained by a log-spectral subtraction technique.[43] Although the power spectra of the reference and through-sample RF signals are influenced by the frequency responses of the associated electronics and transducers, log spectral subtraction provides a somewhat system independent measurement of the attenuation properties of a material.

The magnitudes of the Fourier transforms of the reference and through-sample signals are

$$|\tilde{U}_w(x=L,\omega)|=|A_0(\omega)|\exp(-\alpha_w(\omega)L) \tag{7}$$

and

$$|\tilde{U}_s(x=L,\omega)|=|T^P_{ws}(\omega)||T^P_{sw}(\omega)||A_0(\omega)|$$
$$\times\exp(-\alpha_w(L-d)-\alpha_s(\omega)d), \tag{8}$$

respectively. The attenuation coefficient of water was neglected.[44] The logarithm of the ratio of (7) and (8), divided by the thickness of the sample, yields the amplitude attenuation coefficient of the sample (in dB/cm),

$$\alpha_s(\omega)=\frac{20\log(|\tilde{U}_w(\omega)|)-20\log(|\tilde{U}_s(\omega)|)+10\log(T^I_{ws}T^I_{sw})}{d}, \tag{9}$$

where T^I_{ws} and T^I_{sw} are the intensity transmission coefficients of the water-sample and sample-water interfaces, respectively.

C. Kramers-Kronig analysis

Because the attenuation coefficients of the plastics obey an approximately linear relationship with frequency, the predictions of phase velocity from the measured attenuation were calculated using the once-subtracted nearly-local approximation of the Kramers-Kronig equations[16,19,21] for the flat and parallel sides of the plates:

$$\frac{1}{c_s(\omega)}-\frac{1}{c_s(\omega_0)}=-\frac{2}{\pi}\frac{\beta}{2\pi}\ln\left(\frac{\omega}{\omega_0}\right), \tag{10}$$

where ω_0 represents a reference angular frequency within the usable bandwidth and β is the slope of a least squares line fit to the measured attenuation coefficient plotted as a function of frequency.

For the step discontinuities on each plate, the once-subtracted forms of the Kramers-Kronig relations were used to predict the apparent phase velocity from the apparent attenuation coefficient over the usable bandwidth (ω_a to ω_b):

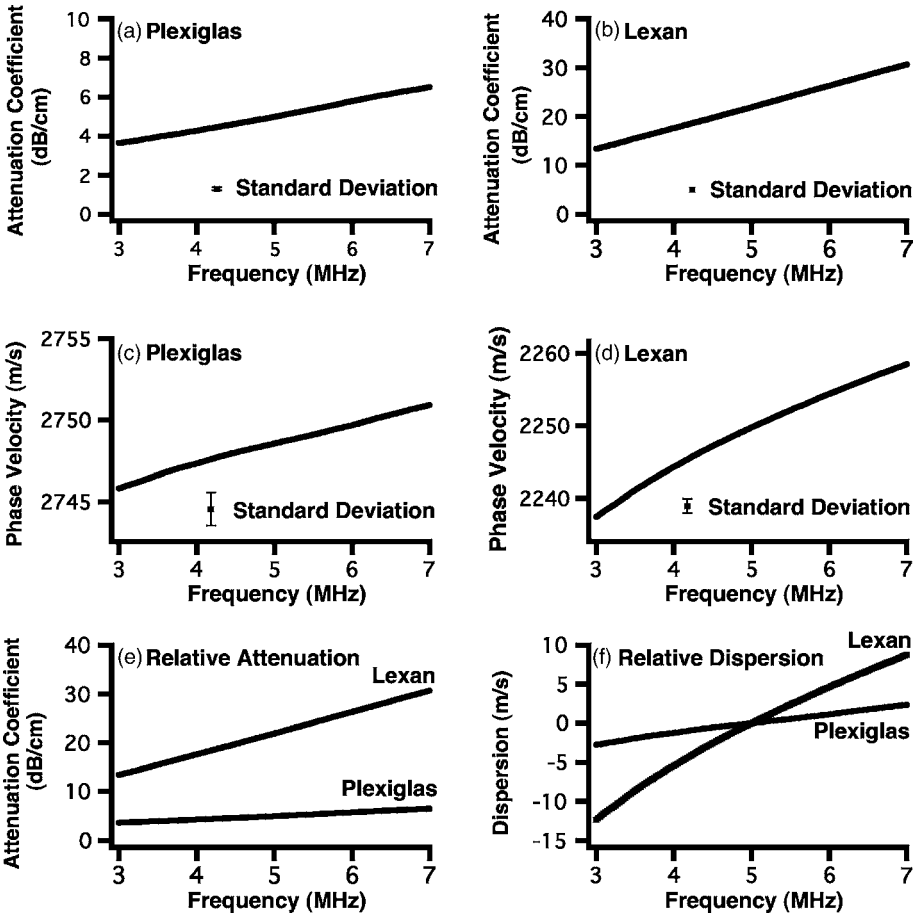


FIG. 2. Measured properties of the flat and parallel regions of the two polymer plates. (a) Attenuation coefficient of Plexiglas, (b) attenuation coefficient of Lexan, (c) phase velocity of Plexiglas, (d) phase velocity of Lexan, (e) compared attenuation coefficients of Plexiglas and Lexan, (f) relative dispersion of Plexiglas and Lexan.

$$\frac{1}{c_p(\omega)} - \frac{1}{c_0} = \frac{2}{\pi}\int_{\omega_a}^{\omega_b} \frac{\alpha(\omega') - \alpha(\omega)}{\omega'^2 - \omega^2} \mathrm{d}\omega', \tag{11}$$

where c_0 is manually chosen to set the absolute level of the predicted phase velocity. The impetus for this approach is described below and elsewhere.[16,18,21]

IV. RESULTS AND DISCUSSION

A. Measurements

Measurements of the attenuation coefficient over the useful bandwidth of the measurement system for the flat and parallel sides of each plastic are shown in Figs. 2(a) and 2(b) and compared in Fig. 2(e). As anticipated, both materials exhibit an approximately linear with frequency attenuation coefficient, but with markedly different slopes. The slopes of the frequency dependence of the attenuation of Lexan was found to be (mean ± standard deviation) 4.33 ± 0.07 dB/cm MHz whereas that of the Plexiglas was measured to be 0.73 ± 0.02 dB/cm MHz. Figures 2(c) and 2(d) show the measured phase velocity of each sample, with their relative dispersions displayed in Fig. 2(f). The phase velocity at 5 MHz of the Plexiglas and Lexan samples was measured to be 2749 ± 1 m/s and 2250 ± 1 m/s, respectively. As anticipated[16] a logarithmic dispersion with frequency was observed over the flat regions.

Measurements from the stepped regions of each plate show a markedly different behavior with frequency. Figures 3(a) and 3(b) show the apparent frequency dependence of the attenuation coefficient for insonification over the small step discontinuity in the surfaces of the Lexan and Plexiglas plates. These data demonstrate a peak value near the calculated phase cancellation frequency, f_c. The apparent phase velocity curves, shown in Figs. 3(c) and 3(d), display a dramatic increase in the apparent phase velocity as the frequency approaches and then passes through f_c. A comparison of the trends of Figs. 3(a) and 3(b) with those of Figs. 3(c) and 3(d) indicates that the peak value in the experimentally measured frequency dependent apparent attenuation coefficient of a given plastic (at f_c) corresponds to the point of maximum change in the corresponding measured frequency dependence of the apparent phase velocity. It is this "derivative-like" relationship that is explored (via the Kramers-Kronig predictions) in the following section.

B. Kramers-Kronig predictions

Figures 4(a) and 4(b) show the predicted phase velocity based on the measured attenuation data for the flat and parallel sides of the plastic plates obtained using the nearly-local approximation to the Kramers-Kronig relations. Because this form of the Kramers-Kronig equations provides predictions of the dispersion but not the absolute velocities, the measured values of the phase velocity at band center for the flat and parallel regions were used to set the overall value of phase velocity for each plastic, namely 2749 m/s for the Plexiglas plate, and 2250 m/s for the Lexan plate. Good

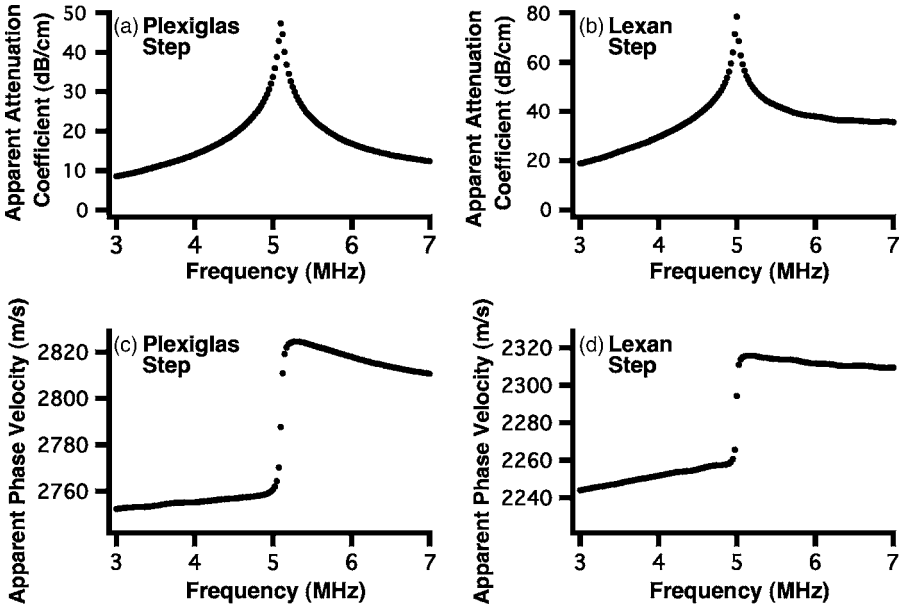


FIG. 3. Measured apparent properties of the stepped region of the two polymer Plates. (a) Apparent attenuation coefficient of Plexiglas (0.3 mm step), (b) apparent attenuation coefficient of Lexan (0.4 mm step), (c) apparent phase velocity of Plexiglas, (d) apparent phase velocity of Lexan.

agreement between the predicted and measured phase velocity was observed for both the Plexiglas and the Lexan samples.

Figures 4(c) and 4(d) display the experimentally measured apparent phase velocity with the predicted apparent phase velocity obtained from the measured apparent attenuation data for insonification through the stepped region of the plastic plates. The overall level of these predictions was set using the same value as those used for the nearly local approximation, i.e., the previously measured phase velocity of each plastic at 5 MHz. The overall shapes of the predictions seem to match those of the measured quantities well. However, the absolute levels of the predicted curves are shifted relative to the absolute level of the measurements. The predicted velocities more closely match the measured values if the offset values of c_0 in (12) are set to 2800 and 2305 m/s

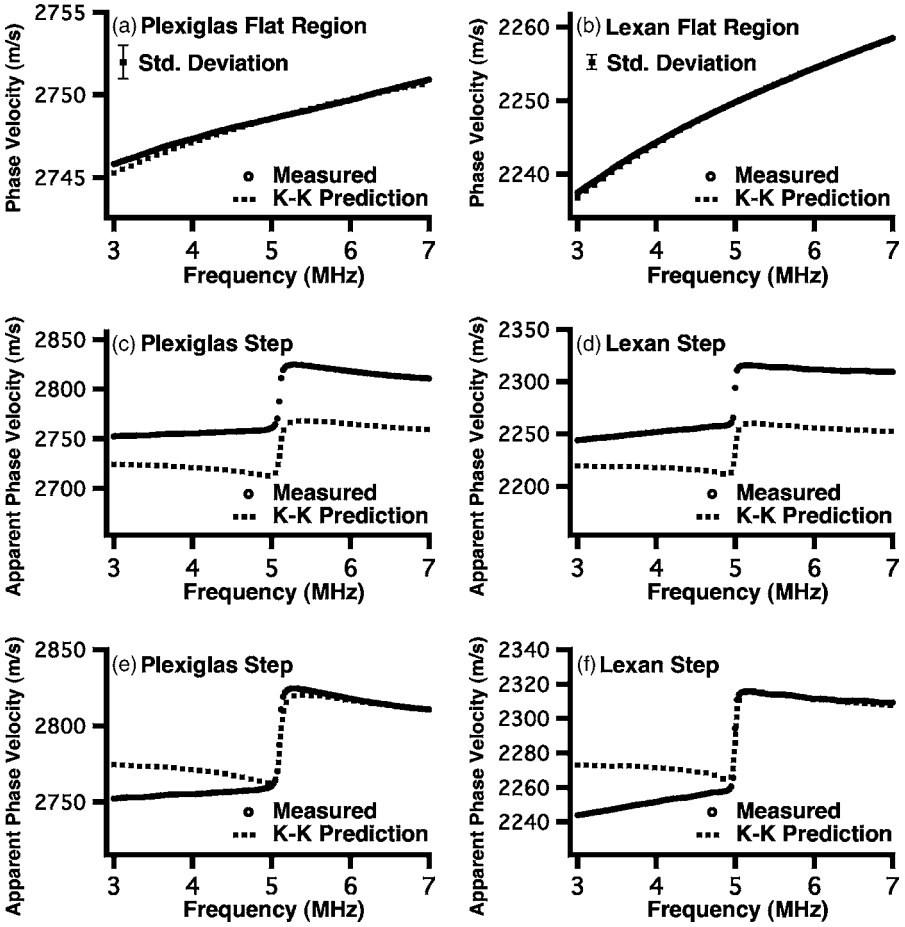


FIG. 4. (a) and (b) The measured and Kramers-Kronig predictions of the phase velocity from the measure attenuation coefficient for the flat and parallel regions of the two polymer plates, (a) Plexiglas and (b) Lexan. (c)–(f) The measured and Kramers-Kronig predictions of the apparent phase velocity from the measured apparent attenuation coefficient for the stepped regions of the plates using (c) and (d) the phase velocity at 5 MHz for the flat and parallel regions and (e) and (f) manually chosen values for the absolute level of the predictions.

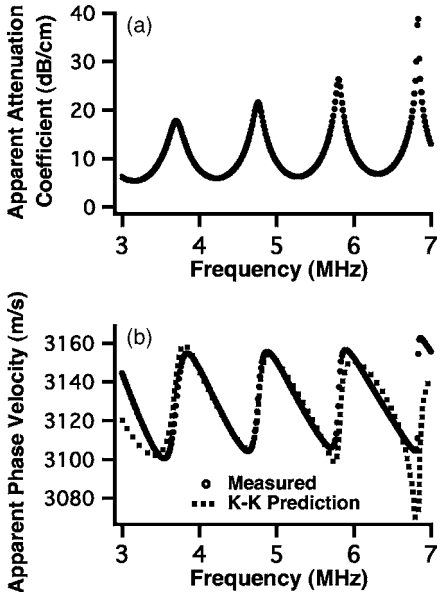


FIG. 5. (a) The measured apparent attenuation coefficient and (b) the measured and predicted apparent phase velocity of the Plexiglas sample containing the 3 mm step.

for the Plexiglas and Lexan, respectively, as depicted in Figs. 4(e) and 4(f). Although these values are higher than those experimentally measured for the flat and parallel regions, they are in the range of the higher apparent phase velocity values measured for the two stepped plates, as demonstrated by comparing Figs. 3(c) and 3(d) with Figs. 2(c) and 2(d). For those values of c_0 in Eq. (11), overall agreement appears quite good between the predictions and the experimental measurements. The Kramers-Kronig relationship accurately predicts the magnitude of the increase of the measured apparent phase velocity in the rapidly changing range from 4.8 to 5.2 MHz, as well as the trend from this point to the upper end of the bandwidth for both plastics investigated.

For additional confirmation of the validity of the Kramers-Kronig relations in the presence of phase cancellation at the surface of a phase sensitive piezoelectric receiver, the Plexiglas sample with a 3 mm step milled into one of its surfaces was analyzed in exactly the same manner described above as the other two plates. Figures 5(a) and 5(b) display the measured apparent attenuation, the measured apparent phase velocity, and the predicted apparent phase velocity for the 3 mm step. The overall offset, c_0, for the prediction to lie on top of the experimental data is 3135 m/s. Again, the agreement found between the frequency dependence of the predicted apparent phase velocity and the frequency dependence of the measured apparent phase velocity appears quite good. The overall level of the phase velocity of both plastics increases when the stepped regions are interrogated. The physical significance of c_0 and how it relates to this observed behavior in the measured phase velocity is the focus of ongoing investigation.

The results of this study suggest that the Kramers-Kronig relationship between the apparent phase velocity and apparent attenuation coefficient remains approximately valid, even in the presence of phase-distorted data. This initially surprising result suggests a potential approach for compensating for phase and amplitude distortion artifacts arising from ultrasonic field propagation in inhomogeneous and anisotropic media. Data obtained from the individual elements of a two dimensional array could, in principle, be compensated in an iterative fashion in which frequency dependent apparent attenuation would be used to improve estimates of frequency dependent velocity, which could in turn be used to improve estimates of attenuation, and so on.

V. CONCLUSION

The Kramers-Kronig relations were used to predict the (apparent) phase velocity from the (apparent) attenuation for three stepped plastic plates measured by phase sensitive methods. The Kramers-Kronig predictions of the apparent phase velocities were compared with the experimentally determined apparent phase velocities using the measured apparent attenuation coefficients. Data reported indicate that a causally consistent relationship between apparent phase velocity and apparent attenuation exists even in the presence of severe phase distortion of the ultrasonic field and despite the substantial artifacts incurred by measuring with an experimental setup sensitive to these aberrations.

ACKNOWLEDGMENT

This study was supported, in part, by NIH Grant R37 HL40302.

[1]R. Truell, C. Elbaum, and B. B. Chick, *Ultrasonic methods in solid state physics* (Academic Press, New York, 1969), pp. 107–108.
[2]H. Seki, A. Granato, and R. Truell, "Diffraction effects in the ultrasonic field of a piston source and their importance in the accurate measurement of attenuation," J. Acoust. Soc. Am. **28**, 230–238 (1956).
[3]R. Truell and W. Oates, "Effect of lack of parallelism of sample faces on the measurement of ultrasonic attenuation," J. Acoust. Soc. Am. **35**, 1382–1386 (1963).
[4]L. J. Busse, J. G. Miller, D. E. Yuhas, J. W. Mimbs, A. N. Weiss, and B. E. Sobel, "Phase cancellation effects: A source of attenuation artifact eliminated by a CdS acoustoelectric receiver," Ultrasound Med. Biol. **3**, 1519–l535 (1977).
[5]L. J. Busse and J. G. Miller, "Detection of spatially nonuniform ultrasonic radiation with phase sensitive (piezoelectric) and phase insensitive (acoustoelectric) receivers," J. Acoust. Soc. Am. **70**, 1377–1386 (1981).
[6]L. J. Busse and J. G. Miller, "A comparison of finite aperture phase sensitive and phase insensitive detection in the near field of inhomogeneous material," *Proceedings of IEEE Ultrasonics Symposium 81CH1689-9*, pp. 617–626 (1981).
[7]P. H. Johnston and J. G. Miller, "Phase-insensitive detection for measurement of backscattered ultrasound," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **33**, 713–721 (1986).
[8]P. W. Marcus and E. L. Carstensen, "Problems with absorption measurements of inhomogeneous solids," J. Acoust. Soc. Am. **58**, 1334–1335 (1975).
[9]O. T. von Ramm, F. L. Thurstone, and J. Kisslo, "Cardiovascular diagnosis with real time ultrasound imaging," in *Acoustical Holography*, edited by N. Booth (Plenum Press, New York, 1975), Vol. 6, pp. 91–102.
[10]F. L. Thurstone and O. T. von Ramm, "A new ultrasound technique employing two-dimensional, electronic beam steering," in *Acoustic Holography*, edited by P. S. Green (Plenum Press, New York, 1974), Vol. 5, pp. 249–259.
[11]M. O'Donnell, E. T. Jaynes, and J. G. Miller, "General relationships between ultrasonic attenuation and dispersion," J. Acoust. Soc. Am. **63**, 1935–1937 (1978).
[12]M. O'Donnell, E. T. Jaynes, and J. G. Miller, "Kramers-Kronig relationship between ultrasonic attenuation and phase velocity," J. Acoust. Soc. Am. **69**, 696–701 (1981).
[13]J. Mobley, K. R. Waters, M. S. Hughes, C. S. Hall, J. N. Marsh, G. H.

Brandenburger, and J. G. Miller, "Kramers-Kronig relations applied to finite bandwidth data from suspensions of encapsulated microbubbles," J. Acoust. Soc. Am. **108**, 2091–2106 (2000).
[14]J. Mobley, K. R. Waters, M. S. Hughes, C. S. Hall, J. N. Marsh, G. H. Brandenburger, and J. G. Miller, Erratum: "Kramers-Kronig relations applied to finite bandwidth data from suspensions of encapsulated microbubbles," J. Acoust. Soc. Am. **112**, 760–761 (2002).
[15]K. R. Waters, M. S. Hughes, G. H. Brandenburger, and J. G. Miller, "On a time-domain representation of the Kramers-Kronig dispersion relations," J. Acoust. Soc. Am. **108**, 2114–2119 (2000).
[16]K. R. Waters, M. S. Hughes, J. Mobley, G. H. Brandenburger, and J. G. Miller, "On the applicability of Kramers-Kronig relations for ultrasonic attenuation obeying a frequency power law," J. Acoust. Soc. Am. **108**, 556–563 (2000).
[17]R. L. Trousil, K. R. Waters, and J. G. Miller, "Experimental validation of the use of Kramers-Kronig relations to eliminate the phase sheet ambiguity in broadband phase spectroscopy," J. Acoust. Soc. Am. **109**, 2236–2244 (2001).
[18]J. Mobley, K. R. Waters, and J. G. Miller, "Finite bandwidth effects on the casual prediction of ultrasonic attenuation of the power-law form," J. Acoust. Soc. Am. **114**, 2782–2790 (2003).
[19]K. R. Waters, M. S. Hughes, J. Mobley, and J. G. Miller, "Differential forms of the Kramers-Kronig dispersion relations," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **50**, 68–76 (2003).
[20]J. Mobley, K. R. Waters, and J. G. Miller, "Causal determination of acoustic group velocity and frequency derivative of attenuation with finite-bandwidth Kramers-Kronig relations," Phys. Rev. E **72**, 016604-1–5 (2005).
[21]K. R. Waters, J. Mobley, and J. G. Miller, "Causality-imposed (Kramers-Kronig) relationships between attenuation and dispersion," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **52**, 822–833 (2005).
[22]N. Akashi, J. Kushibiki, and F. Dunn, "Acoustic properties of egg yolk and albumen in the frequency range 20–400 MHz," J. Acoust. Soc. Am. **102**, 3774–3778 (1997).
[23]R. M. Arthur and K. V. Gurumurthy, "A single-pole model for the propagation on ultrasound in soft tissue," J. Acoust. Soc. Am. **77**, 1589–1597 (1985).
[24]P. P. Antich, J. A. Anderson, R. B. Ashman, J. E. Dowdey, J. Gonzales, R. C. Murry, J. E. Zerwekh, and C. Y. C. Pak, "Measurement of mechanical-properties of bone material invitro by ultrasound reflection—methodology and comparison with ultrasound transmission," J. Bone Miner. Res. **6**, 417–426 (1991).
[25]J. F. Aubry, M. Tanter, J. Gerber, J. L. Thomas, and M. Fink, "Optimal focusing by spatio-temporal inverse filter. II. Experiments. Application to focusing through absorbing and reverberating media," J. Acoust. Soc. Am. **110**, 48–58 (2001).
[26]V. Behar, "Techniques for phase correction in coherent ultrasound imaging systems," Ultrasonics **39** 603–610 (2002).
[27]D. H. Huang and J. Tsao, "Analysis and correction of ultrasonic wavefront distortion based on a multilayer phase-screen model," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **49**, 1686–1703 (2002).
[28]J. C. Lacefield, W. C. Pilkington, and R. C. Waag, "Comparisons of lesion detectability in ultrasound images acquired using time-shift compensation and spatial compounding," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **51**, 1649–1659 (2004).
[29]J. C. Lacefield and R. C. Waag, "Time-shift estimation and focusing through distributed aberration using multirow arrays," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **48**, 1606–1624 (2001).
[30]Y. Li, "The influences of ambiguity phase aberration profiles on focusing quality in the very near field—Part II: Dynamic range focusing on reception," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **49**, 72–84 (2002).
[31]S. E. Masoy, T. F. Johansen, and B. Angelsen, "Correction of ultrasonic wave aberration with a time delay and amplitude filter," J. Acoust. Soc. Am. **113**, 2009–2020 (2003).
[32]A. I. Nachman, I. James, F. Smith, and R. C. Waag, "An equation for acoustic propagation in inhomogeneous media with relaxation losses," J. Acoust. Soc. Am. **88**, 1584–1595 (1990).
[33]N. M. Ivancevich, J. J. Dahl, G. E. Trahey, and S. W. Smith, "Phase-aberration correction with a 3-D ultrasound scanner: Feasibility study," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **53**, 1432–1439 (2006).
[34]K. D. Wallace, M. R. Holland, B. S. Robinson, R. J. Fedewa, C. W. Lloyd, and J. G. Miller, "Impact of propagation through an aberrating medium on the linear effective apodization of a nonlinearly generated second harmonic field," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **53**, 1260–1268 (2006).
[35]T. L. Szabo and J. R. Wu, "A model for longitudinal and shear wave propagation in viscoelastic media," J. Acoust. Soc. Am. **107**, 2437–2446 (2000).
[36]L. E. Kinsler, A. R. Frey, A. B. Coppens, and J. V. Sanders, *Fundamentals of Acoustics*, 4th ed. (Wiley, New York, 2000), pp. 179–184.
[37]J. Mobley, C. S. Hall, J. N. Marsh, M. S. Hughes, K. R. Waters, G. H. Brandenburger, and J. G. Miller, "Measurements and predictions of the phase velocity and attenuation coefficient in suspensions of elastic microspheres," J. Acoust. Soc. Am. **106**, 652–659 (1999).
[38]W. Sachse and Y. H. Pao, "On the determination of phase and group velocities of dispersive waves in solids," J. Appl. Phys. **49**, 4320–4327 (1978).
[39]J. E. Perrin, G. Berger, A. Salesses, and M. Agneray, "Methods of tissue characterization with ultrasonic spectroscopy, analysis of the results obtained, prospects for clinical application in a short time," C. R. Seances Soc. Biol. Fil. **173**, 345–58 (1979).
[40]H. Hachiya, S. Ohtsuki, M. Tanaka, and F. Dunn, "Determination of sound speed in biological tissues based on frequency analysis of pulse response," J. Acoust. Soc. Am. **92**, 1564–1568 (1992).
[41]J. Blitz, *Fundamentals of Ultrasonics*, 1st ed. (Butterworth & Co. LTD., London, 1963), 24–28.
[42]K. A. Wear, T. A. Stiles, G. R. Frank, E. L. Madsen, F. Cheng, E. J. Feleppa, C. S. Hall, B. S. Kim, P. Lee, W. D. O'Brien, Jr., M. L. Oelze, B. I. Raju, K. K. Shung, T. A. Wilson, and J. R. Yuan, "Interlaboratory comparison of ultrasonic backscatter coefficient measurements from 2 to 9 MHz," J. Ultrasound Med. **24**, 1235–1250.
[43]S. L. Baldwin, K. R. Marutyan, M. Yang, K. D. Wallace, M. R. Holland, and J. G. Miller, "Measurements of the anisotropy of ultrasonic attenuation in freshly excised myocardium," J. Acoust. Soc. Am. **119**, 3130–3139 (2006).
[44]J. J. Markham, "Absorption of sounds in fluids," Rev. Mod. Phys. **23**, 353–411 (1953).

Effects of acoustic parameters on bubble cloud dynamics in ultrasound tissue erosion (histotripsy)

Zhen Xu[a] and Timothy L. Hall
Department of Biomedical Engineering, University of Michigan, Ann Arbor, Michigan 48109

J. Brian Fowlkes
Department of Radiology and Department of Biomedical Engineering, University of Michigan, Ann Arbor, Michigan 48109

Charles A. Cain
Department of Biomedical Engineering and Department of Electrical Engineering and Computer Science, University of Michigan, Ann Arbor, Michigan 48109

(Received 12 September 2006; revised 3 April 2007; accepted 3 April 2007)

High intensity pulsed ultrasound can produce significant mechanical tissue fractionation with sharp boundaries ("histotripsy"). At a tissue-fluid interface, histotripsy produces clearly demarcated tissue erosion and the erosion efficiency depends on pulse parameters. Acoustic cavitation is believed to be the primary mechanism for the histotripsy process. To investigate the physical basis of the dependence of tissue erosion on pulse parameters, an optical method was used to monitor the effects of pulse parameters on the cavitating bubble cloud generated by histotripsy pulses at a tissue-water interface. The pulse parameters studied include pulse duration, peak rarefactional pressure, and pulse repetition frequency (PRF). Results show that the duration of growth and collapse (collapse cycle) of the bubble cloud increased with increasing pulse duration, peak rarefactional pressure, and PRF when the next pulse arrived after the collapse of the previous bubble cloud. When the PRF was too high such that the next pulse arrived before the collapse of the previous bubble cloud, only a portion of histotripsy pulses could effectively create and collapse the bubble cloud. The collapse cycle of the bubble cloud also increased with increasing gas concentration. These results may explain previous *in vitro* results on effects of pulse parameters on tissue erosion. © *2007 Acoustical Society of America.* [DOI: 10.1121/1.2735110]

PACS number(s): 43.35.Ei, 43.80.Gx, 43.80.Sh [FD] Pages: 229–236

I. INTRODUCTION

Tissue disintegration using ultrasound induced cavitation[1–9] and shockwave generated tissue destruction[10,11] have been observed by many researchers. Our recent studies have shown that short, high-intensity pulses delivered at certain pulse repetition frequencies (PRF) can achieve complete mechanical tissue fragmentation.[12–15] This technique can be considered a form of soft tissue lithotripsy, which we call "histotripsy." At a tissue-fluid interface, histotripsy results in tissue erosion with sharply demarcated boundaries.[12] Histotripsy induced tissue erosion may be applied to many clinical applications where tissue removal is needed. For example, we are currently developing the technique as a noninvasive procedure to perforate the atrial septum (the tissue separating the two atria of the heart) in the treatment of a congenital heart disease called hypoplastic left heart syndrome.[12,15]

Acoustic cavitation is believed to be the primary mechanism for the histotripsy process. Our previous studies have shown that histotripsy pulses can generate a dynamically changing bubble cluster, which has been observed by high speed imaging.[16] The bubbles appear to behave as one dynamic unit, growing and shrinking together,[17] and therefore this bubble cluster can be called a "bubble cloud" as previously defined by cavitation researchers.[18–21] The bubble cloud formation corresponds to the initiation of a temporally changing acoustic backscatter,[16] which is an excellent indicator of tissue erosion.[22] These concordant results suggest that the cavitating bubble cloud is essential for the erosion process.

The acoustic pressures effective for histotripsy are similar to those found in lithotripter shockwave pulses. In comparison to the one cycle pulses commonly used in lithotripsy, the histotripsy pulses are several acoustic cycles in duration. Shockwave lithotripsy studies have shown that positive pressure can compress existing bubbles, while following negative pressure can cause bubble growth and collapse. The durations of growth and collapse are long (hundreds of microseconds) compared to the lithotripsy pulse length (several microseconds). The bubble radius-time curve has been modeled for lithotripsy[23] and confirmed experimentally *in vitro*[24–26] and *in vivo*.[27–29] In this paper, we used transmitted light signals to trace the growth and collapse of the bubble cloud generated by histotripsy pulses. As this optical method detects bubbles by monitoring the attenuation of the transmitted light intensity, we refer to it as an optical attenuation method. Our previous studies have shown that the optical attenuation was only detected when a bubble cloud was observed using high speed imaging.[16]

[a]Electronic mail: zhenx@umich.edu

TABLE I. Pulse parameters used in Fig. 1 and corresponding light attenuation results at a tissue-water interface.

Row number in Fig. 4	Gas (PO_2)	Acoustic parameters: PD (cycles)	P^- (MPa)	P^+ (MPa)	I_{SPPA} (W/cm^2)	PRF (kHz)	Light attenuation results (initiated): Attenuation duration (mean±s.d.) (μs)	Peak attenuation level (mean±s.d.)	Number of pulses: Total	Initiated
First	33%–40%	**3 (4 μs)**	13.9	25.1	**9.5 k**	2	23.8±13.1	0.18±0.12	910	267
		6 (8 μs)					32.2±11.9	0.30±0.13	910	622
		12 (16 μs)					37.3±9.4	0.48±0.18	910	754
Second	33%–40%	3	**13.9**	**25.1**	**9.5** k	2	23.8±13.1	0.18±0.12	910	267
			15.5	**31.9**	**12.4 k**		42.5±13.5	0.40±0.20	910	746
			17.1	**39.7**	**15.6 k**		64.4±11.5	0.74±0.17	910	902
Third	22%–24%	3	15.5	31.9	12.4 k	**0.5**	N/A	N/A	200	None
						2	42.9±6.0	0.68±0.15	200	171
						5	80.4±24.3	0.76±0.21	98	90
Fourth						**10**	67.6±11.1	0.71±0.23	98	90
Fifth						**20**	82.2±20.1[a]	0.82±0.13[a]	98	90

At PRF of 20 kHz, there was a consistent pattern between pairs of pulses with one pulse producing a bubble cloud and the subsequent pulse prolonging and collapsing it. So the attenuation duration was calculated by the duration of light attenuation generated by every two pulses.

The selection of pulse parameters including pressure amplitude or intensity levels, pulse duration, and PRF affects the extent and efficiency of the mechanical tissue disruption induced by histotripsy.[12,13,30,31] For example, the axial erosion rate is almost an order of magnitude greater using shorter pulses (three cycles) as opposed to longer pulses (twenty-four cycles) with the same energy.[12] Erosion also appears more energy efficient at certain PRFs than others.[12] The axial erosion rate is slower and the erosion area is significantly larger at high intensity ($I_{SPPA} > 5000$ W/cm^2).[30] Furthermore, the erosion rate depends on the gas content of the fluid in which the tissue is submerged. Erosion is harder to initiate at lower dissolved gas concentrations, but the erosion rate is faster.[12] The dependence of other cavitational effects on pulse parameters and gas content has also been observed by many other researchers.[8,32–37]

In this paper, we used the optical attenuation method to study the effects of pulse parameters and dissolved gas concentration on the dynamics of a bubble cloud generated by histotripsy pulses at a tissue-water interface. Pulse parameters studied include pulse duration, peak rarefactional pressure, and PRF. Furthermore, we compared bubble cloud dynamics at a tissue-water interface to that in free water to evaluate the influence of a soft tissue boundary on the bubble cloud. We report the duration of growth and collapse (collapse cycle) of the bubble cloud, as measured by the duration of the light attenuation. The results may explain the effects of pulse parameters on tissue erosion and serve as a guide for future parameter optimization.

II. METHODS

A. Ultrasound generation and calibration

Ultrasound pulses were generated by an 18-element piezocomposite spherical-shell therapeutic array (Imasonic, S.A., Besançon, France) with a center frequency of 750 kHz and a geometric focal length of 100 mm. The therapy array had an annular configuration with outer and inner diameters of 145 and 68 mm, respectively. All the array elements were excited together in phase. The array driving system, maintained under PC control, consists of channel driving circuitry, associated power supplies (Model 6030A, HP, Palo Alto, CA), and a software platform to synthesize driving patterns. A PC console also provided control of a motorized three-dimensional positioning system (Parker Hannifin, Rohnert Park, CA) to position the array at each exposure site.

The pressure wave form at the focus of the 18-element array in the acoustic field was measured in degassed water (12%–25% of normal saturation) (i.e., free-field conditions) using a fiber-optic probe hydrophone developed in-house.[38] The lateral and axial pressure profiles of the focused beam were measured to be 2.2 mm $\times$ 12.6 mm in width (full width at half maximum) at peak rarefactional pressure of 14 MPa and 1.8 $\times$ 11.9 mm at 19 MPa. The beamwidth decreased with increasing pressure. The peak rarefactional and compressional pressures and spatial-peak pulse-average intensity (I_{SPPA})[39] used in experiments were measured for free-field conditions and reported in Table I. In Fig. 1, multiple pulses were used to generate a bubble cloud. While in other figures (Figs. 2 and 3; Tables II and III), a single pulse was used to create a bubble cloud. The pressure levels in the single pulse experiments could not be calibrated successfully due to the instantaneous cavitation. A peak rarefactional pressure of 21 MPa and a peak compressional pressure of 76 MPa were measured at a lower power level without generation of bubbles during measurement. This pressure wave form is shown in Fig. 4.

B. Tissue sample preparation

Fresh porcine atrial wall tissue (1 to 2 mm thick) was obtained from a local abattoir and used within 24 h of harvesting. All tissue specimens were preserved in a 0.9% sodium chloride solution at 4 °C. Tissue was wrapped over ring-shaped tube fitting (2 cm in diameter), so that no tissue interfered with the laser beam. The tissue was submerged in

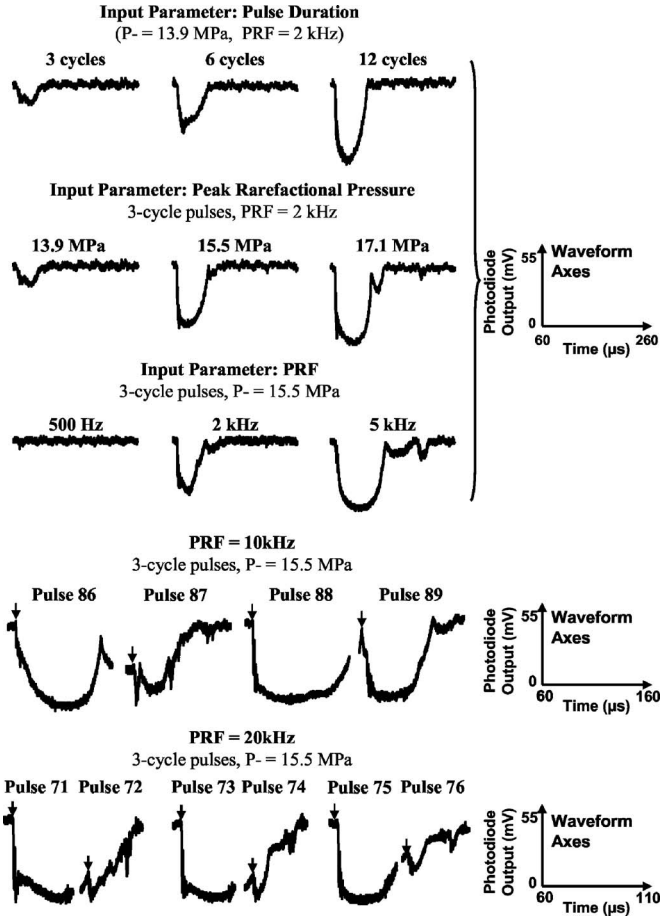


FIG. 1. Examples of light attenuation signals caused by the bubble cloud generated by histotripsy pulses using different pulse parameters. The axes for the wave form are the same for each row and shown on the right. Arrows indicate the arrival of the pulse. The pulse parameters and corresponding light attenuation results are listed in Table I.

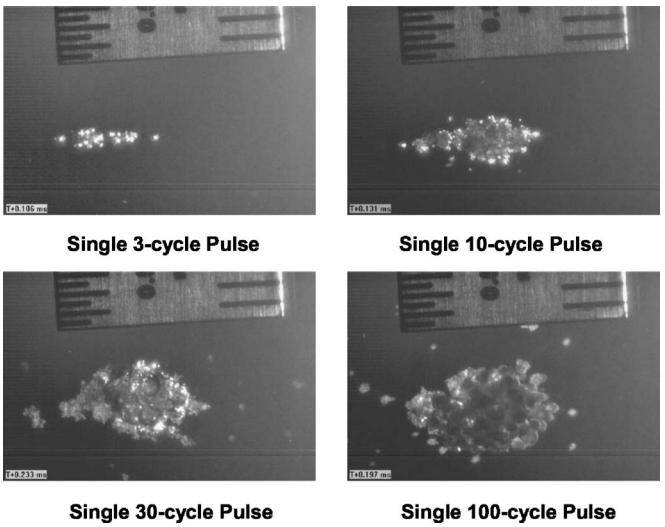


FIG. 2. Images of the bubble cloud generated in free water created by a single histotripsy pulse at different pulse durations. The histotripsy pulse was delivered from the left to the right of each image. The overall size of the bubble cloud increased with increasing pulse duration. The ruler on the top of each image has markings of 1 mm on the right side and 0.5 mm on the left.

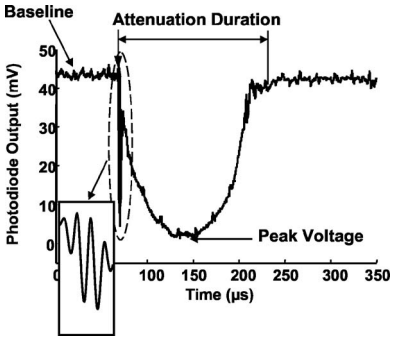


FIG. 3. Example of the light attenuation signal caused by a bubble cloud, recorded as the photodiode voltage output. The bubble cloud was generated by a three-cycle (4-μs) pulse at a tissue-water interface with 98%–100% gas concentration. The left arrow below "Attenuation Duration" indicates the arrival of the histotripsy pulse at the transducer focus where the laser beam was projected. The insert is an expanded view (expanded in the horizontal direction and compressed in the vertical) of the artifact in the light attenuation signal during the histotripsy pulse, which tracks the pulse wave form.

water of desired gas concentration in room temperature (~22 °C) for an hour prior to experimentation. The partial pressure of oxygen (PO_2) in the water was used as our metric for gas concentration and was measured with an YSI dissolved oxygen meter (Model 5000, YSI, Yellow Springs, OH).

C. Optical attenuation detection

The optical attenuation method monitors the light transmission and detects the light beam reduction caused by bubbles. The schematic diagram of the experimental setup is shown in Fig. 5. Bubble clouds were produced in a 30-cm-wide × 60-cm-long × 30-cm-high water tank designed to enable optical observations. A 1-mW helium-neon gas laser (Model. 79245, Oriel, Stratford, CT) was placed on one side of the tank to emit a laser beam which traveled through the ultrasound focus (and in front of the tissue at a tissue-water interface). The light intensity was monitored continuously by a photodiode (Model DET100, ThorLabs, Newton, NJ) aligned with the laser beam at the other side of the tank.

To align the laser beam through the ultrasound focus, the therapy transducer was first pulsed in free water to create a visible bubble cloud at its focus. Photos of the bubble cloud taken by a high speed camera are shown in Fig. 2. The position of the transducer was adjusted to direct the laser beam

TABLE II. Attenuation durations (mean ± s.d.) generated by a single pulse to study the effects of the gas concentrations and a tissue boundary.

		Pulse Duration			
Environment	Gas concentration	Three cycles (μs)	Six cycles (μm)	Twelve cycles (μs)	Twenty four cycles (μs)
Free water	24%–26%	None[a]	139±66	173±33	219±60
	98%–100%	76±55	173±53	744±243	936±375
Tissue-water interface	24%–26%	111±34	148±69	210±15	226±14
	98%–100%	149±48	409±172	1066±400	2017±276

[a]Only the artifact during the histotripsy pulse was observed, no light attenuation was detected after the pulse.

TABLE III. Peak attenuation levels (mean±s.d.) generated by a single pulse to study the effects of the gas concentrations and a tissue boundary.

		Pulse duration			
Environment	Gas concentration	Three cycles	Six cycles	Twelve cycles	Twenty cycles
Free water	24%–26%	None	0.58±0.41	0.81±0.21	0.93±0.13
	98%–100%	0.16±0.09	0.79±0.27	1±0	0.96±0.06
Tissue-water interface	24%–26%	0.40±0.29	0.70±0.24	0.96±0.05	0.99±0.02
	98%–100%	0.84±0.22	1±0[a]	0.98±0.04	1±0[a]

A peak attenuation level of 1 corresponds to the complete blockage of the light beam by bubbles.

through the center of the bubble cloud by visual alignment. To form a tissue-water interface, a piece of porcine atrial wall was placed parallel and immediately behind (<1 mm) the laser beam. The laser beam width (0.48 mm at $1/e^2$) was smaller than the bubble cloud, so the photodiode measured the light transmitted through a portion of the bubble cloud and not the whole cloud. However, since the bubbles within the cloud appeared to grow and shrink together,[17] the dynamics of a portion of the bubble cloud is considered here as indicative of the dynamics of the overall bubble cloud.

The attenuated light signal was recorded as the voltage output of the photodiode. The photodiode output was connected to a digital oscilloscope (Model 9384L, LeCroy Chestnut, NY) using a 1-MΩ dc coupling in parallel with a 250-Ω resistor. An impedance of 250 Ω was chosen to achieve a good signal to noise ratio (30–35 dB) and a wide enough dynamic range (60 dB) for attenuation detection, while still maintaining good temporal resolution (−3-dB width response time of 15 ns). The acquisition of the optical signals was synchronized with the acoustic therapy pulse.

An example of a light attenuation signal is demonstrated in Fig. 3, in which the light intensity began to decrease at the arrival of a single 4-μs histotripsy pulse (produced by a three-cycle pulse driving signal at 750 kHz) at a tissue-water interface. The light intensity continued to decrease for 78 μs after the end of the histotripsy pulse before it recovered to the baseline (164 μs after the arrival of the histotripsy pulse). During the time window when the histotripsy pulse propagated through the laser beam, the light attenuation signal seemed to track the ultrasound pulse wave form (insert in Fig. 3). This acousto-optic effect is most likely due to

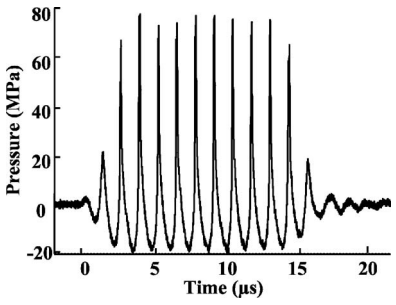


FIG. 4. Acoustic pressure wave form of a ten-cycle (14-μs) histotripsy pulse in water at the transducer focus. For this pulse, the peak rarefactional pressure was 21 MPa and the peak compressional pressure was 76 MPa.

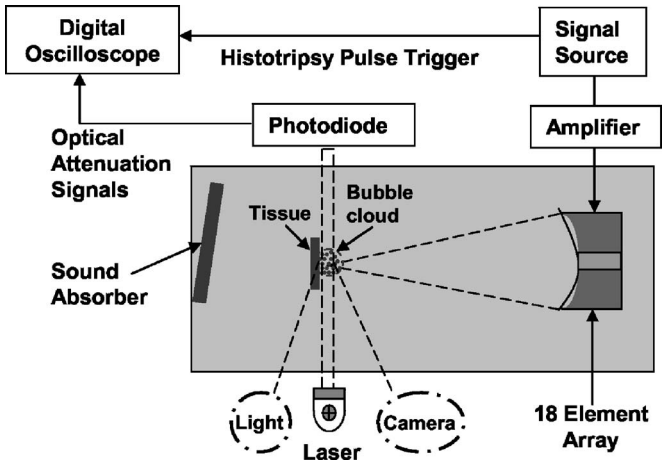


FIG. 5. Diagram of the experimental arrangement for bubble cloud monitoring at a tissue-water interface using an optical attenuation method. Light source and camera (in dashed circle) are setup for high speed imaging in water. However, at a tissue-water interface, the light source was blocked by the tissue, and the imaging could not be used with the optical attenuation detection system.

changes in the index of refraction of water during the histotripsy pulse. It provides a convenient timing indicator for locating the histotripsy pulse with respect to the generated bubble cloud and was discussed in detail in our previous paper on optical monitoring of histotripsy generated bubbles.[16]

Researchers have used optical scattering and reflection signals from bubbles to effectively trace radius-time curves generated by a lithotripsy pulse.[25,27,40] Based on these studies, duration of the light attenuation (attenuation duration) is believed to indicate the duration of growth and collapse of a bubble cloud. The attenuation duration is defined as the period of time when the light intensity (photodiode output) falls below a threshold of baseline minus 3 times the noise level. The baseline and noise level are mean and standard deviation (s.d.) values, respectively, of the photodiode output receiving the laser light before the arrival of the histotripsy pulse (i.e., when no bubbles are present). The peak amplitude of the light attenuation signal is affected by the size and number of the bubbles within the laser beam. The peak attenuation level is defined as the difference between the baseline and the peak voltage divided by the baseline level, ranging between 0 and 1. The peak voltage is calculated excluding the light attenuation signal which tracks the ultrasound wave form during the histotripsy pulse.

Our previous optical monitoring and high speed imaging results have shown that initiation and extinction of the light attenuation correspond well to formation and disappearance of the bubble cloud.[16] Therefore, the light attenuation duration and peak attenuation level were only calculated when the light attenuation was initiated, i.e., when the bubbles were generated. Initiation of light attenuation occurs when the light attenuation duration exceeds the pulse duration for five consecutive pulses. Extinction of light attenuation occurs when, after initiation, the light attenuation duration drops below the pulse duration for five consecutive pulses.

The purpose of using pulse duration as a threshold is to overcome the light attenuation increase due to the ultrasound induced water index of refraction change.

D. Optical imaging

Images of bubble clouds were taken by a high speed digital imaging system (Model Phantom V9, Vision Research, Wayne, NJ) at a frame rate of 7 kHz and a shutter speed of 10 μs. The imaging system was placed outside the water tank approximately 100 mm away from the bubble cloud. The bubble cloud was illuminated by a strong light source (Carousel 4400 Projector light, Kodak, Rochester NY) at a $\sim$30° angle with respect to the camera (Fig. 5). An optical lens with a focal length of 50–100 mm (Zoom Nikkor, Nikon, Japan) was mounted in front of the imaging system to increase the magnification. The imaging system was only used to image the bubble cloud in free water. At a tissue-water interface, the light source was blocked by the tissue and the imaging could not be used with the optical attenuation detection system (Fig. 5).

E. Specific studies

We investigated the dependence of the bubble cloud dynamics on pulse parameters and dissolved gas concentration in the fluid where tissue erosion occurs. Pulse parameters studied included pulse duration, peak rarefactional pressure, and PRF. We also compared the dynamics of the bubble cloud created at a tissue-water interface and in free water to determine the influence of the soft tissue boundary on bubble activity.

To investigate the dependence of the bubble cloud on pulse duration, peak rarefactional pressure, and PRF, multiple (100-910) pulses were used. We recorded optical light attenuation signals and calculated the attenuation duration and peak attenuation levels of initiated optical attenuation signals. Table I lists the pulse parameters and gas concentration ranges used in each specific study and corresponding light attenuation duration and peak attenuation level results (mean±s.d.). Pulse duration in the following refers to the number of cycles in the pulse driving signal applied to the transducer at 750 kHz.

To study the dependence of the bubble cloud on gas concentration and tissue boundary, a single pulse was used to generate the bubble cloud. The bubble cloud was created in free water and at a tissue-water interface. The purpose of using a single pulse was to monitor the bubble cloud without the influence from adjacent pulses. Gas concentration ranges of 24%–26% and 98%–100% were tested. The same peak rarefactional pressure was used, but the focal pressure field could not be successfully measured due to the rapid onset of cavitation. Between three and eight data points made up the sample for each combination of parameters.

III. RESULTS

A. Effects of pulse duration

The duration of light attenuation results show that the collapse cycle of bubble clouds generated by histotripsy pulses was much longer than the pulse duration itself. The attenuation duration increased with increasing pulse duration at a tissue-water interface (Table I and the first row of Fig. 1) such that the attenuation durations (mean±s.d.) were 23.8±13.1, 32.2±11.9, and 37.3±9.4 μs for pulse duration of three cycles, six cycles, and twelve cycles, respectively (T-test p value$<$0.001 for each pair). A peak rarefactional pressure of 13.9 MPa, a PRF of 2 kHz, a total of 910 pulses, and a gas concentration of 33%–40% were used for all pulse durations. Out of the 910 pulses, the number of pulses which initiated the optical attenuation signal also increased with increasing pulse duration (Table I).

The increasing trend of attenuation duration with pulse duration was even more evident in bubble clouds produced at a higher peak rarefactional pressure ($>$21 MPa) and higher gas concentration (98%–100%) (Table II). The attenuation durations (mean±s.d.) were 149±48 and 1066±400 μs for pulse durations of three cycles and twelve cycles, respectively (T-test p value$<$0.001). This indicates a seven-fold increase in the attenuation duration for a four-fold increase in pulse duration.

The peak attenuation level is affected by the size and number of the bubbles within the laser beam and this attenuation measure increased with increasing pulse durations (Table I, the first row of Fig. 1). The mean±s.d. of the peak attenuation level was 0.18±0.12 for three-cycle pulses, 0.30±0.13 for six-cycle pulses, and 0.48±0.18 for twelve-cycle pulses (T-test p value$<$0.001 for each pair).

Figure 2 depicts the high-speed camera images of bubble clouds created by a single histotripsy pulse at different pulse durations in gas saturated water. Images of bubble clouds showed that the bubble cloud dimensions were larger with longer pulse duration. The bubble cloud diameters along the lateral/axial acoustic beam direction were approximately 1.1/5.5, 2.7/6.9, 4.5/7.9, and 4.6/8.2 mm for pulse durations of three cycles, ten cycles, thirty cycles, and one-hundred cycles, respectively. This finding is consistent with the optical attenuation result showing that the peak attenuation level increased with increasing pulse duration (Table III).

B. Effects of peak rarefactional pressure

The attenuation duration lengthened as the peak rarefactional pressure increased (Table I and the second row of Fig. 1). The resulting increase in the attenuation duration was greater than the increase in peak rarefactional pressure. The attenuation durations (mean±s.d.) were 23.8±13.1, 42.5±13.5, and 64.4±11.5 μs for peak rarefactional pressures of 13.9, 15.5, and 17.1 MPa, respectively (T-test p value$<$0.001 for each pair). A nearly three-fold increase in attenuation duration resulted from a 23% increase in peak rarefactional pressure (e.g., at peak rarefactional pressures of 13.9 and 17.1 MPa). A pulse duration of three cycles, a PRF of 2 kHz and a total of 910 pulses, and a gas concentration of 33%–40% were used for all variations of peak rarefactional pressures. The number of initiated pulses also increased with increasing peak rarefactional pressure (Table I).

The peak attenuation level was greater with higher peak

rarefactional pressure (Table I and the second row of Fig. 1). The peak attenuation levels (mean±s.d.) were 0.18±0.12, 0.40±0.20, and 0.74±0.17 for peak rarefactional pressures of 13.9, 15.5, and 17.1 MPa, respectively (T-test p value <0.001 for each pair). Again this is approximately three-fold increase in peak attenuation level when changing peak rarefactional pressure by only 23%.

C. Effects of PRF

PRF values ranging from 500 Hz to 20 kHz were tested. A pulse duration of three cycles, a peak rarefactional pressure of 15.5 MPa, and gas concentration of 22%–24% were used for all PRFs. When the PRF was too low, the light attenuation could not be initiated and the light attenuation was never initiated at the 500-Hz PRF (Table I and Fig. 1).

When the PRF was above the threshold to initiate the light attenuation signal, the attenuation duration increased with increasing PRF. The attenuation durations (mean±s.d.) were 42.9±6.0 μs for PRF of 2 kHz and 80.4±24.3 μs for PRF of 5 kHz (Table I and the third row of Fig. 1; T-test p value <0.001). This remained true when the pulse repetition period was longer than the attenuation duration, i.e., when the next pulse arrived after the collapse of the bubble cloud created by the previous pulse. After the primary light attenuation, rebounds of the bubble cloud were detected as additional light attenuation for some pulses. The number of pulses that produced rebounds was greater with higher PRF.

When the PRF was too high and the next pulse arrived before the collapse of the previous bubble cloud or the rebound of the previous cloud, the collapse of the previous cloud was disrupted, and the generation of the new cloud was also inhibited. As a result, only a portion of the pulses effectively generated and collapsed the bubble cloud. At a PRF of 10 kHz, approximately one-quarter of the pulses arrived before the collapse of the previous bubble cloud or the previous cloud rebound. In these cases, the light attenuation signals never fully recovered to the baseline (i.e., the collapse of previous bubble cloud did not occur completely or did not occur at all) (the fourth row of Fig. 1). The light attenuation did increase again after the arrival of the next pulse. At a PRF of 20 kHz, the next pulse always arrived before the collapse of the previous bubble cloud. There was a relatively consistent pattern within pairs of pulses with one pulse producing a bubble cloud and the subsequent pulse prolonging and collapsing it (the fifth row of Fig. 1).

There was no discernible trend of the peak attenuation level with PRF when the bubble cloud was initiated. No significant statistical difference was observed in the initiated attenuation level at PRFs of 2, 5, 10, and 20 kHz (T-test p value >0.05 for each peak attenuation level pair).

D. Effects of gas concentration

The light attenuation duration increased with increasing gas concentration. This trend was observed in free water and at a tissue-water interface (Table II). At a tissue-water interface, the attenuation durations (mean±s.d.) generated by the same single twenty-four-cycle pulse were 226±14 μs and 2017±276 at gas concentration ranges of 24%–26% and 98%–100%, respectively (Table II). This shows almost a full order of magnitude difference (T-test p value=0.003).

The peak attenuation level was also greater with higher gas concentration. This observation was consistent in free water and at a tissue-water interface (Table III). At a tissue-water interface, peak attenuation levels (mean±s.d.) generated by a single three-cycle pulse were 0.40±0.29 and 0.84±0.22 for gas concentration ranges of 24%–26% and 98%–100%, respectively (T-test p value=0.005).

E. Effects of tissue boundary (tissue-water interface versus free water)

The attenuation duration increased with increasing pulse duration and gas concentration both at a tissue-water interface and in free water using a single pulse (Table II). However, the collapse cycle of the bubble cloud was longer at a tissue-water interface than in free water, as indicated by the longer attenuation duration (Table II). For example, the attenuation durations (mean±s.d.) produced by a single six-cycle pulse with a gas concentration of 98%–100% was 173±53 μs in free water and 409±172 μs at a tissue-water interface (T-test p value=0.009).

The peak attenuation level also increased with increasing pulse duration and gas concentration both at a tissue-water interface and in free water (Table III). The peak attenuation level was greater at a tissue-water interface than in free water (Table III). The peak attenuation levels (mean±s.d.) produced by a single three-cycle pulse with a gas concentration of 98%–100% were 0.16±0.09 and 0.84±0.22 in free water and at a tissue-water interface, respectively (T-test p value=0.0001).

IV. DISCUSSION

The optical monitoring results show that the dynamics of the bubble cloud generated by histotripsy pulses depend on the pulse parameters. The light attenuation duration results suggest that the collapse cycle of the bubble cloud increased with increasing pulse duration and peak rarefactional pressure. The collapse cycle of the bubble cloud was greater with higher PRF when the next pulse arrived after the collapse of the previous bubble cloud. When the PRF was too high and the next pulse arrived before the collapse of the previous cloud, only a portion of the pulses effectively generated and collapsed the bubble cloud. Further, the peak attenuation level results suggest that the size and number of the bubbles in the cloud increased with increasing pulse duration and peak rarefactional pressure.

These results shed light on the dependence of histotripsy induced tissue erosion on pulse parameters observed in our *in vitro* tissue experiments.[12,22,30] One important observation is that erosion is more energy efficient at a certain PRF[12] than others. Based on the optical monitoring results, when the pulse repetition period is longer than the bubble cloud collapse cycle, the collapse cycle of the bubble cloud increased with increasing PRF. Consequently, at higher PRF, more cavitational nuclei from the previous collapse are available to provide seeds for the next pulse, resulting in greater erosion efficiency. When the PRF is too high and the next

pulse arrives before the collapse of the previous bubble cloud, the collapse of the cloud is adversely affected, leading to reduced erosion efficiency. Therefore, the most energy efficient erosion only occurs when the timing of the next pulse (i.e., PRF) is "just right." Based on results from this paper, the most efficient PRF should be between 5 and 10 kHz at a pulse duration of three cycles and a peak rarefactional pressure of 15.5 MPa. However, the most energy efficient PRF from our previous *in vitro* experiment[12] at a pulse duration of three cycles is ~14 kHz. This difference could be attributed to the lower peak rarefactional pressure (11.6 MPa) used in the *in vitro* experiment. As the collapse cycle of the bubble cloud decreases with decreasing peak rarefactional pressure, the most efficient PRF is expected to be higher at a lower pressure. We used a greater peak rarefactional pressure for optical monitoring as the pressure level used in the *in vitro* study could not generate a large enough bubble cloud to be detected by the optical system with a good signal-to-noise ratio.

Another *in vitro* result shows that the axial erosion rate is slower and the erosion area is significantly larger at high pulse intensity ($I_{SPPA} > 5000$ W/cm^2).[30] This may be explained by the effects of pulse pressure/intensity on the optical attenuation signals. The peak attenuation level increases with increasing peak rarefactional pressure, which suggests a greater number and/or size of bubbles in the cloud. The cloud may induce a "shadowing" effect where ultrasound energy is scattered by bubbles, slowing down the erosion in the center. But more and/or larger bubbles may increase local scattering, thereby increasing the peripheral erosion.

Our optical monitoring results of the dynamics of bubble clouds generated by histotripsy pulses are generally consistent with published studies on bubble clouds generated by lithotripsy pulses. Researchers have shown an increasing dependence of the bubble cloud collapse cycle on peak rarefactional pressure in water[41] and in tissue.[27] Previous PRF studies[42,43] have demonstrated that the number of cavitation bubbles increases with increasing PRF (below 1 kHz), as more residual bubbles from the previous pulse can serve as cavitation nuclei for the subsequent pulse.

In addition to pulse parameters, the dissolved gas content in the fluid also affects the bubble cloud dynamics. At higher gas concentration, the collapse cycle of the bubble cloud is longer and the number and/or size of bubbles are greater, which is likely due to the greater availability of gas to form bubbles. This result is consistent with our *in vitro* observation of a shorter time to initiate the bubble cloud at higher gas concentration.[22] When the gas concentration is high with pulse duration $\geqslant$6 cycles, the peak attenuation levels were close to 1 and the standard deviations were low (close to zero). The peak attenuation level is the fractional value of the maximum light attenuation. A peak attenuation level of 1 indicated the complete blockage of light beam transmission by bubbles. When high gas concentration and long pulses were used, the size and number of bubbles within the generated bubble cloud could be high enough to block most of the light beam or the entire light beam.

The tissue boundary also affects the bubble cloud dynamics. For a bubble cloud generated by a single histotripsy pulse, the collapse cycle is longer at a tissue-water interface than in free water. The soft tissue boundary restricted the movement, expansion, and other dynamic changes of the bubbles, which may explain the collapse cycle difference. When multiple pulses are used, some additional factors also need to be considered. The tissue could reduce the effects of acoustic streaming and micro-streaming caused by the bubbles. The tissue surface may trap microbubbles and provide additional cavitation nuclei for subsequent pulses. The influence of a tissue boundary on bubble cloud dynamics using multiple pulses is more relevant to tissue erosion and needs further investigation.

In our current optical attenuation system, the laser beam only covered a portion of the cloud. However, since the bubbles within the cloud appeared to grow and shrink together,[17] the dynamics of a portion of the bubble cloud may be generalizable to the whole bubble cloud. Future use of high speed photography will be necessary to confirm the assumption. In addition, the current optical system is not sensitive enough to detect small individual bubbles. The current optical system may be improved by widening the laser beam width using optical lenses, increasing laser power, using a photodetector array, and increasing the sensitivity of the photodetector. High-speed imaging is the most direct means to resolve the absolute sizes and spatial distribution of bubbles in the bubble cloud and will be used in a future study.

V. CONCLUSIONS

To investigate the mechanism of histotripsy induced tissue erosion, we used an optical method to monitor the dynamics of a bubble cloud generated by histotripsy pulses at a tissue-water interface. The optical results show that the bubble cloud dynamics depend on pulse parameters and dissolved gas content in water where the erosion occurs. The collapse cycle of the bubble cloud increased with increasing pulse duration, peak rarefactional pressure, and gas concentration. The collapse cycle of the bubble cloud also increased with PRF when the next pulse arrived after the collapse of the previous bubble cloud. When the PRF was too high such that the next pulse arrived before the collapse of the previous bubble cloud, the previous cloud could not collapse effectively. Furthermore, the collapse cycle of the bubble cloud was longer at a tissue-water interface than in free water. These effects of pulse parameters on bubble cloud dynamics may explain our previous *in vitro* results which demonstrated the dependence of tissue erosion on pulse parameters. The cavitating bubble cloud is sensitive to pulse parameters, allowing a huge parameter space for optimization.

ACKNOWLEDGMENTS

The authors thank Dr. Steve Ceccio for generously providing his lab resources to this work. We want to thank Jessica Parsons for her help on calibration. We also would like to thank Dr. Xueding Wang for his support. This research has been funded by grants from the National Institutes of Health R01-HL077629 and Hitachi.

[1]F. J. Fry, G. Kossoff, R. C. Eggleton, and F. Dunn, "Threshold ultrasound dosages for structural changes in the mammalian brain," J. Acoust. Soc. Am. **48**, 1413–1417 (1970).
[2]F. Dunn and F. J. Fry, "Ultrasonic threshold dosages for the mammalian central nervous system," IEEE Trans. Biomed. Eng. **18**, 253–256 (1971).
[3]L. A. Frizzell, C. S. Lee, P. D. Aschenbach, M. J. Borrelli, R. S. Morimoto, and F. Dunn, "Involvement of ultrasonically induced cavitation in hind limb paralysis of the mouse neonate," J. Acoust. Soc. Am. **74**, 1062–1065 (1983).
[4]G. R. ter Haar, S. Daniels, and K. Morton, "Evidence for acoustic cavitation in vivo: Threshold for bubble formation with 0.75-MHz continuous-wave and pulsed beam," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **33**, 162–164 (1986).
[5]J. B. Fowlkes, P. L. Carson, E. H. Chiang, and J. M. Rubin, "Acoustic generation of bubbles in excised canine urinary bladders," J. Acoust. Soc. Am. **89**, 2740–2744 (1991).
[6]K. Hynynen, "Threshold for thermally significant cavitation in dog's thigh muscle in vivo," Ultrasound Med. Biol. **17**, 157–169 (1991).
[7]J. Y. Chapelon, J. Margonari, F. Vernier, F. Gorry, R. Ecochard, and A. Gelet, "In vivo effects of high-intensity ultrasound on prostatic adenocarcinoma Dunning R3327," Cancer Res. **52**, 6353–6357 (1992).
[8]N. B. Smith and K. Hynynen, "The feasibility of using focused ultrasound for transmyocardial revascularization," Ultrasound Med. Biol. **24**, 1045–1054 (1998).
[9]B. C. Tran, J. Seo, T. L. Hall, J. B. Fowlkes, and C. A. Cain, "Microbubble-enhanced cavitation for noninvasive ultrasound surgery," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **50**, 1296–1304 (2003).
[10]J. Debus, P. Peschke, E. W. Hahn, W. J. Lorenz, A. Lorenz, H. Ifflaender, H. J. G. Zabel, G. van Kaick, andM. Pfeiler, "Treatment of the Dunning prostate rat tumor R3327-AT1 with pulsed high energy ultrasound shock waves (PHEUS): Growth delay and histomorphologic changes," J. Urol. (Baltimore) **146**, 1143–1146 (1991).
[11]A. J. Coleman, T. Kodama, M. J. Choi, T. Adams, and J. E. Saunders, "The cavitation threshold of human tissue exposed to 0.2-MHz pulsed ultrasound: Preliminary measurements based on a study of clinical lithotripsy," Ultrasound Med. Biol. **21**, 405–417 (1995).
[12]Z. Xu, A. Ludomirsky, L. Y. Eun, T. L. Hall, B. C. Tran, J. B. Fowlkes, and C. A. Cain, "Controlled ultrasound tissue erosion," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **51**, 726–736 (2004).
[13]J. E. Parsons, C. A. Cain, G. D. Abrams, and J. B. Fowlkes, "Pulsed cavitational ultrasound therapy for controlled tissue homogenization," Ultrasound Med. Biol. **32**, 115–129 (2006).
[14]W. W. Roberts, T. J. Hall, K. Ives, J. J. S. Wolf, J. B. Fowlkes, and C. A. Cain, "Pulsed cavitational ultrasound: A noninvasive technology for controlled tissue ablation (histotripsy) in the rabbit kidney," J. Urol. (Baltimore) **175**, 734–738 (2006).
[15]A. P. Vlahos, J. E. Lock, D. B. McElhinney, and M. E. van der Velde, "Hypoplastic left heart syndrome with intact or highly restrictive atrial septum: Outcome after neonatal transcatheter atrial septostomy," Circulation **109**, 2326–2330 (2004).
[16]Z. Xu, J. B. Fowlkes, and C. A. Cain, "Optical and acoustic monitoring of bubble cloud dynamics at a tissue-fluid interface in ultrasound tissue erosion," J. Acoust. Soc. Am. **121**, 2421–2430 (2007).
[17]Z. Xu, M. Raghavan, T. L. Hall, C.-H. Chang, M.-A. Mycek, J. B. Fowlkes, and C. A. Cain, "High speed imaging of bubble cloud in pulsed cavitational ultrasound therapy—Histotripsy," Proceedings of the IEEE International Ultrasonics Symposium, Vancouver, Canada, 3–6 October 2006).
[18]Y. A. Pishchalnikov, O. A. Sapozhnikov, M. R. Bailey, J. C. J. Williams, R. O. Cleveland, T. Colonius, L. A. Crum, A. P. Evan, and J. A. McAteer, "Cavitation bubble cluster activity in the breakage of kidney stones by lithotripter shockwaves," J. Endourol **17**, 435–446 (2003).
[19]E. A. Zabolotskaya, Y. A. Ilinskii, G. D. Meegan, and M. F. Hamilton, "Bubble interactions in clouds produced during shock wave lithotripsy," IEEE Ultrasonics Symposium, Montreal, Quebec, Canada, 23–27 August 2004 pp. 890–893.
[20]M. Arora, L. Junge, and C. D. Ohl, "Cavitation cluster dynamics in shock-wave lithotripsy. 1. Free field," Ultrasound Med. Biol. **31**, 827–839 (2005).
[21]Y. Matsumoto, "Bubble and bubble cloud dynamics," 15th International Symposium on Nonlinear Acoustics, Gottingen, Germany, 1–4 September 1999, pp. 65–74.
[22]Z. Xu, J. B. Fowlkes, E. D. Rothman, A. M. Levin, and C. A. Cain, "Controlled ultrasound tissue erosion: The role of dynamic interaction between insonation and microbubble activity," J. Acoust. Soc. Am. **117**, 424–435 (2005).
[23]C. C. Church, "A theoretical study of cavitation generated by an extracorporeal shock wave lithotripter," J. Acoust. Soc. Am. **86**, 215–227 (1989).
[24]A. J. Coleman, M. Whitlock, T. Leighton, and J. E. Saunders, "The spatial distribution of cavitation induced acoustic emission, sonoluminescence and cell lysis in the field of a shock wave lithotripter," Phys. Med. Biol. **38**, 1545–1560 (1993).
[25]K. Jochle, J. Debus, W. J. Lorenz, and P. Huber, "A new method of quantitative cavitation assessment in the field of a lithotripter," Ultrasound Med. Biol. **22**, 329–338 (1996).
[26]T. J. Matula, P. R. Hilmo, M. R. Bailey, and L. A. Crum, "In vitro sonoluminescence and sonochemistry studies with an electrohydraulic shock-wave lithotripter," Ultrasound Med. Biol. **28**, 1199–1207 (2002).
[27]P. Huber, J. Debus, P. Peschke, E. W. Hahn, and W. J. Lorenz, "In vivo detection of ultrasonically induced cavitation by a fibre-optic technique," Ultrasound Med. Biol. **20**, 811–825 (1994).
[28]A. J. Coleman, M. J. Choi, and J. E. Saunders, "Detection of acoustic emission from cavitation in tissue during clinical extracorporeal lithotripsy," Ultrasound Med. Biol. **22**, 1079–1087 (1996).
[29]P. Zhong, I. Cioanta, F. H. Cocks, and G. M. Preminger, "Inertial cavitation and associated acoustic emission produced during electrohydraulic shock wave lithotripsy," J. Acoust. Soc. Am. **101**, 2940–2950 (1997).
[30]Z. Xu, J. B. Fowlkes, A. Ludomirsky, and C. A. Cain, "Investigation of intensity threshold for ultrasound tissue erosion," Ultrasound Med. Biol. **31**, 1673–1682 (2005).
[31]K. Kieran, T. L. Hall, J. E. Parsons, J. S. Wolf, J. B. Fowlkes, C. A. Cain, and W. W. Roberts, "Exploring the acoustic parameter space in ultrasound therapy: Defining the threshold for cavitational effects," Sixth International Symposium on Therapeutic Ultrasound, Oxford, UK, 30 August–2 September 2006, p. S01.
[32]V. Ciaravino, H. G. Flynn, and M. W. Miller, "Pulsed enhancement of acoustic cavitation: A postulated model," Ultrasound Med. Biol. **7**, 159–166 (1981).
[33]A. A. Atchley, L. A. Frizzell, R. E. Apfel, C. K. Holland, S. Madanshetty, and R. A. Roy, "Thresholds for cavitation produced in water by pulsed ultrasound," Ultrasonics **26**, 280–285 (1988).
[34]S. Z. Child, C. L. Hartman, L. A. Schery, and E. L. Carstensen, "Lung damage from exposure to pulsed ultrasound," Ultrasound Med. Biol. **16**, 817–825 (1990).
[35]C. K. Holland and R. E. Apfel, "Thresholds for transient cavitation produced by pulsed ultrasound in a controlled nuclei enviroment," J. Acoust. Soc. Am. **88**, 2059–2069 (1990).
[36]P. Huber, J. Debus, K. Jochle, I. Siamiantonakis, J. Jenne, R. Rastert, J. Spoo, W. J. Lorenz, and M. Wannenmacher, "Control of cavitation activity by different shockwave pulsing regimes," Phys. Med. Biol. **44**, 1427–1437 (1999).
[37]P. P. Chang, W. S. Chen, P. D. Mourad, S. L. Poliachik, and L. A. Crum, "Thresholds for inertial cavitation in albunex suspensions under pulsed ultrasound conditions," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **48**, 161–170 (2001).
[38]J. E. Parsons, C. A. Cain, and J. B. Fowlkes, "Cost-effective assembly of a basic fiber-optic hydrophone for measurement of high-amplitude therapeutic ultrasound fields," J. Acoust. Soc. Am. **119**, 1432–1440 (2006).
[39]AIUM, *Acoustic Output Measurement Standard for Diagnostic Ultrasound Equipment, UD2–98* (AIUM/NEMA, 1998).
[40]T. J. Matula, P. R. Hilmo, B. D. Storey, and A. J. Szeri, "Radial response of individual bubbles subjected to shock wave lithotripsy pulses in vitro," Phys. Fluids **14**, 913–921 (2002).
[41]M. J. Choi, A. J. Coleman, and J. E. Saunders, "The influence of fluid properties and pulse amplitude on bubble dynamics in the field of a shock wave lithotripter," Phys. Med. Biol. **38**, 1561–1573 (1993).
[42]P. Huber, K. Jochle, and J. Debuss, "Influence of shock wave pressure amplitude and pulse repetition frequency on the lifespan, size and number of transient cavities in the field of an electromagnetic lithotripter," Phys. Med. Biol. **43**, 3113–3128 (1998).
[43]O. A. Sapozhnikov, V. A. Khokhlova, M. R. Bailey, J. C. Williams Jr., J. A. McAteer, R. O. Cleveland, and L. A. Crum, "Effect of overpressure and pulse repetition frequency on cavitation in shock wave lithotripsy," J. Acoust. Soc. Am. **112**, 1183–1195 (2002).

Impedance of pistons on a two-layer medium in a planar infinite rigid baffle

Scott E. Hassan[a)]
NAVSEA Newport, Newport, Rhode Island 02841-5047

(Received 6 June 2006; revised 29 January 2007; accepted 2 April 2007)

An integral transform technique is used to develop a general solution for the impedance of rigid pistons acting on a two-layer medium. The medium consists of a semi-infinite acoustic fluid on a viscoelastic thick plate in a rigid infinite baffle. The stresses acting on the planar baffle, as a result of piston motion, are determined using theory of linear elasticity and are therefore unrestricted in terms of applicable frequency range. The special case of a circular piston is considered and expressions for the self-and mutual impedances are developed and evaluated numerically. Numerical results are compared with classical piston impedance functions and finite-element model results. At low frequencies ($k_0a<1$), the self-impedances vary significantly from the classical piston impedance functions due to the shear properties of the viscoelastic medium. In the midfrequency range ($1<k_0a<\pi$) the self-impedances vary from the classical piston impedance functions for moderate viscoelastic layer thicknesses ($0.5<h/a<2$). The mutual impedances associated with pistons on a two-layer medium generally exhibit an increased decay, as a function of separation distance, over the classical results. © *2007 Acoustical Society of America.*
[DOI: 10.1121/1.2735104]

PACS number(s): 43.38.Hz, 43.20.Rz, 43.30.Jx [SFW] Pages: 237–246

I. INTRODUCTION

Predicting the self-and mutual impedance associated with pistons, acting on a two-layer medium in a rigid baffle, is of fundamental importance in many areas of engineering acoustics. One common example is sonar transducer arrays located on the nose of undersea vehicles. A moderately thick viscoelastic coating over the array is often required for hydrodynamic purposes and to protect the array elements from damage. The sonar transducer heads act as pistons on a two-layer medium consisting of the viscoelastic solid and fluid. To accurately assess sonar system performance, understanding the influence of a moderately thick viscoelastic layer on the self-and mutual impedances is necessary.

Typically, viscoelastic materials with a close impedance match to the surrounding fluid are chosen for these applications and classical baffled piston impedance functions are used as an approximation. Much of the previous work has addressed various aspects of pistons in rigid infinite baffles acting directly on a semi-infinite fluid. The radiation impedance associated with fluid-loaded circular pistons has been developed extensively using approaches involving surface integrals of Green's functions,[1] integral transforms,[2] or time domain techniques.[3] General calculation methodologies with specific application to a rectangular aperture in a rigid baffle were addressed by Pierce *et al.*[4] The related problem of interaction between circular pistons on a planar rigid baffle was calculated by Klapman[5] and Pritchard.[6]

Previous work related to the response of rigid pistons on viscoelastic or a layered viscoelastic medium are limited to computing self-impedances associated with a traction-free baffle. An integral equation approach was developed by Luco[7,8] and solved numerically for a circular piston on a layered viscoelastic medium. The impedance associated with a circular piston on a viscoelastic half-space was considered by Zhang *et al.*[9] Integral transform and Green's function methods were implemented by Amedin *et al.*[10] to predict the acoustic field in a medium consisting of fluid layer under a fluid half-space.

The related problem of determining the response of fluid-loaded thick plates using elasticity theory has also received considerable attention. Typically, integral transform techniques are applied to reduced wave equations governing the vector and scalar displacement potentials. Ko[11] utilized this approach to evaluate the effectiveness of an air-voided elastomer at reducing structure-borne noise. Pathak and Stepanishen[12] developed an approach to characterize acoustic radiation from fluid-loaded elastic plates. The influence of fluid loading on infinite elastic plates has also been considered by many others including Crighton[13] and Freedman.[14]

In the present paper, integral transform methods are used to develop a general solution for the impedance of pistons acting on a two-layer medium consisting of a semi-infinite acoustic fluid on a viscoelastic solid in a rigid infinite baffle. The theory is presented in Sec. II where the stresses in the solid are determined using the theory of linear elasticity and the pressures in the fluid are assumed to be governed by the linear wave equation. The special case of a circular piston is considered in detail. Results for the self-and mutual impedances of circular pistons for various viscoelastic layer thicknesses are presented in Sec. III along with related finite-element results for the self-impedances.

[a)]Electronic mail: hassanse@npt.nuwc.navy.mil

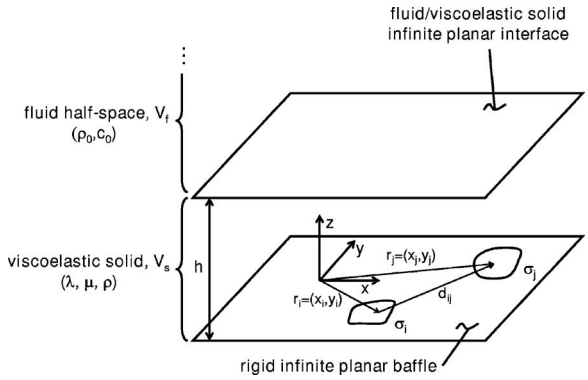


FIG. 1. Two-layer medium in a rigid infinite planar baffle containing general pistons.

II. THEORY

In this section, integral transform techniques are utilized to develop an expression for the self-and mutual impedance of pistons. The final expression is an integral that contains the product of transformed aperture functions and the spectral impedance of a general multilayer medium. The governing differential equations of motion are then developed for the two layers consisting of a viscoelastic solid and fluid. The equations of linear elasticity are used to describe the response of the isotropic viscoelastic medium, and the classical linear acoustic wave equation is used to describe the response of the fluid. These equations are coupled by applying boundary conditions at the fluid/solid interface and rigid baffle resulting in a system of linear equations for the unknown complex displacement potential coefficients. These coefficients are used to evaluate the stresses acting on the rigid baffle and the associated spectral impedance of the two-layer medium.

A. Formulation of impedance expressions

Consider the inhomogeneous two-layer medium depicted in Fig. 1. A semi-infinite acoustic fluid volume, V_f, is above an isotropic viscoelastic solid volume, V_s, of thickness h. The lower surface of the viscoelastic solid is on a rigid infinite planar baffle at $z=0$ and contains two arbitrary surface regions σ_i and σ_j located at $\mathbf{r}_i$ and $\mathbf{r}_j$, respectively, on the rigid baffle. The position vector on the baffle from the ith to jth piston is $\mathbf{d}_{ij}$. The piston described by σ_j is vibrating with a harmonic ($e^{j\omega t}$) uniform velocity. As a result of the piston motion normal to the baffle, a net normal force is exerted on the surfaces σ_i and σ_j. The radiation impedance, $z_{ij}(\omega)$, is defined as the total normal force acting on the ith piston normalized by the complex normal velocity of the jth piston, $V_j(\omega)$. It follows that

$$z_{ij}(\omega) = \left.\frac{F_i(\omega)}{V_j(\omega)}\right|_{v_i(\omega)=0}, \tag{1}$$

where $F_i(\omega)$ is the complex normal component of force obtained by integration of the normal stress component on $z=0$ over the surface σ_i. The two cases of interest for $z_{ij}(\omega)$ include $i=j$ and $i\neq j$ and are referred to as self-and mutual radiation impedance, respectively. The general expression for the impedance can also be written as

$$z_{ij}(\omega) = \lfloor R_{ij}(\omega) + jX_{ij}(\omega)\rfloor, \tag{2}$$

where the R_{ij} and X_{ij} are resistive and reactive components of the complex impedance. In the subsequent development, explicit dependence on $e^{j\omega t}$ is suppressed for convenience.

The solution to the problem of interest is pursued via the use of the following two-dimensional Fourier transform pair:

$$\hat{b}(k_x,k_y) = \int_{-\infty}^{\infty} b(x,y)e^{-j(k_x x+k_y y)}dx\,dy, \tag{3a}$$

$$b(x,y) = \frac{1}{4\pi^2}\int_{-\infty}^{\infty} \hat{b}(k_x,k_y)e^{j(k_x x+k_y y)}dk_x\,dk_y, \tag{3b}$$

where $b(x,y)$ is any field variable of interest and the hat indicates a transformed quantity. Of particular interest for the present study is the case of a radial symmetric kernel associated with a medium exhibiting transverse isotropy. For this case, the Fourier transform pair, given in Eqs. (3a) and (3b), is equivalent to a Hankel transform pair of order 0.

The complex impedance in Eq. (1) can be expressed as an integral of the spectral impedance weighted by an aperture function and a transformed aperture function. It follows that

$$z_{ij} = \frac{1}{4\pi^2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \hat{Z}(k_x,k_y)\hat{S}_i(k_x,k_y)S_j(x,y) \times e^{j(k_x x+k_y y)}dx\,dy\,dk_x dk_y, \tag{4}$$

where $\hat{Z}(k_x,k_y)$ is the spectral impedance relating the transformed normal stress to the transformed normal velocity. This quantity is expressed as

$$\hat{Z}(k_x,k_y) = -\left.\frac{\hat{\tau}_{zz}(k_x,k_y,z)}{\hat{v}(k_x,k_y,z)}\right|_{z=0}, \tag{5}$$

where $\hat{\tau}_{zz}(k_x,k_y,z)$ and $\hat{v}(k_x,k_y,z)$ are the transformed normal stress and velocity, respectively.

The $S_j(x,y)$ is the aperture function for surface σ_j, and $\hat{S}_i(k_x,k_y)$ is the transformed aperture function associated with σ_i. Integration of Eq. (4) over (x,y) results in an expression containing only transformed quantities in the integrand. It follows that

$$z_{ij} = \frac{1}{4\pi^2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \hat{Z}(k_x,k_y)\hat{S}_i(k_x,k_y)\hat{S}_j^*(k_x,k_y)dk_x\,dk_y, \tag{6}$$

where the asterisk denotes complex conjugate and it has been assumed that the aperture function $S_j(x,y)$ is real valued. The expression for impedance in Eq. (6) can be simplified using the following change of variables:

$$k_x = \eta\cos(\theta) \tag{7a}$$

$$k_y = \eta\sin(\theta). \tag{7b}$$

Substituting Eqs. (7a) and (7b) into Eq. (6), and recalling that the viscoelastic solid and fluid composite exhibit transverse isotropy, it follows that Eq. (6) can be expressed as

$$z_{ij} = \frac{1}{4\pi^2}\int_0^{\infty} \eta \hat{Z}(\eta) \int_0^{2\pi} \hat{S}_i(\eta,\theta)\hat{S}_j^*(\eta,\theta) d\eta\, d\theta, \tag{8}$$

where

$$\hat{Z}(\eta) = -\left.\frac{\hat{\tau}_{zz}(\eta,z)}{\hat{v}(\eta,z)}\right|_{z=0} \tag{9}$$

is the spectral impedance of the transversely isotropic two-layer medium. It is noted that expressions similar to Eq. (8) have also been used for the classical fluid-loaded rectangular piston on a planar infinite rigid baffle.[15]

B. Aperture functions

The transformed aperture functions in Eq. (8) act as wave number filters to the spectral impedance associated with the two-layer medium. The filter characteristics are a function of the aperture geometric parameters and the weighting function. For the present study, these transformed aperture functions are associated with two surface regions: (1) the region over which the normal stress on the piston is integrated and (2) the region over which the normal velocity boundary condition is specified. For the classical circular rigid piston on a planar infinite baffle acting on a linear acoustic fluid, a unit weighting over the circular aperture is implied. This results in a discontinuous aperture function and an associated discontinuous velocity boundary condition. The velocity field in the viscoelastic solid, considered in the present study, is continuous and therefore requires a continuous aperture function. Continuous aperture functions have been considered by Tjotta and Tjotta[16] for the study of acoustic fields due a piston with nonuniform velocity distribution. The piecewise linear aperture function used for the present study provides the simplest transition in the small peripheral region, between a uniform normal piston velocity and a rigid baffle. Additionally, for small δ, results from the present study can be compared to classical baffled piston impedance functions.

The aperture function $S_j(x,y)$ is now defined for the special case of a planar circular region of radius a_j with a continuous weighting. This aperture function is expressed as

$$S_j(r) = \begin{cases} 1; & r \le a_j \\ \left(1 + \dfrac{a_j}{\delta}\right) - \dfrac{r}{\delta}; & a_j \le r \le a_j + \delta \\ 0; & \text{otherwise} \end{cases} \tag{10a}$$

where

$$r = \sqrt{(x - x_j)^2 + (y - y_j)^2}, \tag{10b}$$

and (x_j, y_j) is the offset of the aperture geometric center from the origin of the coordinate system. The parameter δ in Eq. (10a) controls the width of the peripheral region of the piston where the weighting function varies linearly from unity to zero. The transformed circular aperture function is found by substituting Eq. (10a) into Eq. (3a) and using Eqs. (7a) and (7b), resulting in

$$\hat{S}_j(\eta,\theta) = \frac{2\pi}{\delta\eta^2}\left\{\frac{1}{\eta}\int_{\eta a_j}^{\eta(a_j+\delta)} J_0(u)du + a_j J_0(\eta a_j) - (a_j + \delta) J_0(\eta(a_j+\delta))\right\} e^{-j(\eta\cos(\theta)x_j + \eta\sin(\theta)y_j)}, \tag{11}$$

where $J_n(\,)$ is the Bessel function of the first kind order n.

The transformed circular aperture function for the classical piston of radius a_i is obtained from Eq. (11) as a limiting case where $\delta \to 0$. It follows that

$$\hat{S}_i(\eta,\theta) = \pi a_i^2 \frac{2J_1(\eta a_i)}{\eta a_i} e^{-j(\eta\cos(\theta)x_i + \eta\sin(\theta)y_i)}, \quad \delta = 0. \tag{12}$$

C. Impedance expressions for circular pistons

The impedance of a circular piston acting on the two-layer medium is obtained by substituting Eq. (12) and the complex conjugate of Eq. (11) into Eq. (8) and integrating over θ, resulting in the following:

$$z_{ij} = 2\pi a_i a_j \int_0^{\infty}\left[\frac{\hat{Z}(\eta)}{\delta\eta^2} J_0(\eta|\mathbf{d}_{ij}|) J_1(\eta a_i) \times \left\{\frac{1}{\eta a_j}\int_{\eta a_j}^{\eta(a_j+\delta)} J_0(u)du + J_0(\eta a_j) - (1 + \delta/a_j) J_0(\eta(a_j+\delta))\right\}\right] d\eta. \tag{13}$$

In the above expression it is noted that a continuous velocity is specified on σ_j, using Eq. (11) and the normal stress integration on σ_i is specified using Eq. (12).

The self-impedance is obtained with $|\mathbf{d}_{ij}| = 0$ and equal piston radii. For this case, Eq. (13) can be further simplified, resulting in

$$z_{ii} = 2\pi a_i^2 \int_0^{\infty} \frac{\hat{Z}(\eta)}{\delta\eta^2} J_1(\eta a_i)\left\{\frac{1}{\eta a_i}\int_{\eta a_{ij}}^{\eta(a_i+\delta)} J_0(u)du + J_0(\eta a_i) - (1 + \delta/a_i) J_0(\eta(a_i+\delta))\right\} d\eta. \tag{14}$$

The classical circular piston impedance expression is obtained from Eq. (13) as a limiting case where $\delta \to 0$. It follows that

$$z_{ij} = 2\pi a_i^2 \int_0^{\infty} \hat{Z}(\eta)\left\{\frac{J_0(\eta|\mathbf{d}_{ij}|) J_1^2(\eta a_i)}{\eta}\right\} d\eta, \tag{15}$$

where it is noted that the self-impedance case corresponds to $|\mathbf{d}_{ij}| = 0$. Analytical solutions to the classical piston self-impedance functions associated with an acoustic fluid, obtained using integral transform methods, can be found in standard texts.[2,15]

D. Equations of motion for the viscoelastic solid

As a result of the transverse isotropic assumption, the integral in Eq. (8) contains the spectral impedance with a

single parameter functional dependence. This enables the use of two-dimensional (plane strain) equations of motion to characterize the displacements of the viscoelastic solid. The irrotational and rotational displacement potentials for the viscoelastic solid satisfy the following reduced wave equations:

$$\nabla^2\phi(\mathbf{x}) + k_d^2\phi(\mathbf{x}) = 0, \tag{16}$$

$$\nabla^2\psi(\mathbf{x}) + k_s^2\psi(\mathbf{x}) = 0, \tag{17}$$

where ∇^2 is the Laplacian operator in the Cartesian coordinate system with $\mathbf{x}=(x,z)$. The complex dilatational and shear wave numbers are, respectively,

$$k_d^2 = \frac{\omega^2}{c_d^2}, \tag{18}$$

$$k_s^2 = \frac{\omega^2}{c_s^2}. \tag{19}$$

The dilatation and shear wave speeds are defined using complex Lamé coefficients, μ and λ, to account for dissipation in the viscoelastic medium. It follows that

$$c_d = \sqrt{\frac{\lambda + 2\mu}{\rho}}, \tag{20}$$

and

$$c_s = \sqrt{\frac{\mu}{\rho}}. \tag{21}$$

Substituting Eqs. (16) and (17) into Eq. (3a) and using Eqs. (7a) and (7b) results in the following set of transformed equations for the displacement potentials:

$$\frac{\partial^2\hat{\phi}(\eta,z)}{\partial z^2} + (k_d^2 - \eta^2)\hat{\phi}(\eta,z) = 0, \tag{22}$$

$$\frac{\partial^2\hat{\psi}(\eta,z)}{\partial z^2} + (k_s^2 - \eta^2)\hat{\psi}(\eta,z) = 0. \tag{23}$$

The solutions to the above equations can be expressed as

$$\hat{\phi}(\eta,z) = \hat{A}_1(\eta)\sin(\gamma_d z) + \hat{B}_1(\eta)\cos(\gamma_d z), \tag{24}$$

$$\hat{\psi}(\eta,z) = \hat{A}_2(\eta)\sin(\gamma_s z) + \hat{B}_2(\eta)\cos(\gamma_s z), \tag{25}$$

where

$$\gamma_d^2 = (k_d^2 - \eta^2), \tag{26}$$

$$\gamma_s^2 = (k_s^2 - \eta^2). \tag{27}$$

The transformed displacements in the viscoelastic medium are found using a standard approach[17] that involves expressing the displacements as a linear combination of a scalar and vector potential. It follows that

$$\hat{u}_x(\eta,z) = j\eta[\hat{A}_1(\eta)\sin(\gamma_d z) + \hat{B}_1(\eta)\cos(\gamma_d z)] - \gamma_s[\hat{A}_2(\eta)\cos(\gamma_s z) - \hat{B}_2(\eta)\sin(\gamma_s z)], \tag{28}$$

$$\hat{u}_z(\eta,z) = \gamma_d[\hat{A}_1(\eta)\cos(\gamma_d z) - \hat{B}_1(\eta)\sin(\gamma_d z)] + j\eta[\hat{A}_2(\eta)\sin(\gamma_s z) + \hat{B}_2(\eta)\cos(\gamma_s z)]. \tag{29}$$

The transformed stresses are now obtained using the standard linear constitutive and strain-displacement relationships[17] for the plane-strain problem of interest. The final result is

$$\hat{\tau}_{xx}(\eta,z) = -(\lambda k_d^2 + 2\mu\eta^2)[\hat{A}_1(\eta)\sin(\gamma_d z) + \hat{B}_1(\eta)\cos(\gamma_d z)] - 2j\mu\eta\gamma_s[\hat{A}_2(\eta)\cos(\gamma_s z) - \hat{B}_2(\eta)\sin(\gamma_s z)], \tag{30}$$

$$\hat{\tau}_{zz}(\eta,z) = -(\lambda k_d^2 + 2\mu\gamma_d^2)[\hat{A}_1(\eta)\sin(\gamma_d z) + \hat{B}_1(\eta)\cos(\gamma_d z)] + 2j\mu\eta\gamma_s[\hat{A}_2(\eta)\cos(\gamma_s z) - \hat{B}_2(\eta)\sin(\gamma_s z)], \tag{31}$$

$$\hat{\tau}_{xz}(\eta,z) = 2j\mu\eta\gamma_d[\hat{A}_1(\eta)\cos(\gamma_d z) - \hat{B}_1(\eta)\sin(\gamma_d z)] - \mu k_s^2[\hat{A}_2(\eta)\sin(\gamma_s z) + \hat{B}_2(\eta)\cos(\gamma_s z)]. \tag{32}$$

E. Equations of motion for the fluid

Following a development similar to that used for the viscoelastic solid, and noting that the fluid is irrotational with a bulk modulus $\lambda_0=\rho_0 c_0^2$, the following reduced wave equation is obtained:

$$\nabla^2\phi_0(\mathbf{x}) + k_0^2\phi_0(\mathbf{x}) = 0, \tag{33}$$

where the acoustic wave number is defined as

$$k_0^2 = \frac{\omega^2}{c_0^2}. \tag{34}$$

Substituting Eq. (33) into Eq. (3a) and using Eqs. (7a) and (7b) results in the transformed acoustic wave equations for the fluid displacement potential,

$$\frac{\partial^2\hat{\phi}_0(\eta,z)}{\partial z^2} + (k_0^2 - \eta^2)\hat{\phi}_0(\eta,z) = 0. \tag{35}$$

The solution to Eq. (35) is obtained by retaining only the outgoing waves (+z direction). It follows that

$$\hat{\phi}_0(\eta,z) = \hat{C}(\eta)e^{-j\sqrt{k_0^2-\eta^2}z}, \tag{36}$$

where the real part of the exponent in Eq. (36) must be negative. The displacements and stresses are obtained in a manner identical to that used for the viscoelastic solid. The displacements can therefore be expressed as

$$\hat{u}_x^f(\eta,z) = j\eta\hat{C}(\eta)e^{-j\sqrt{k_0^2-\eta^2}z}, \tag{37}$$

$$\hat{u}_z^f(\eta,z) = -j\sqrt{k_0^2-\eta^2}\hat{C}(\eta)e^{-j\sqrt{k_0^2-\eta^2}z}, \tag{38}$$

where the superscript f is used to denote a variable in the fluid volume V_f. The stress tensor for the acoustic fluid is diagonal and the transformed stresses and pressure can, respectively, be expressed as

$$\hat{\tau}^f_{zz}(\eta,z) = -\lambda_0 k_0^2 \hat{C}(\eta) e^{-j\sqrt{k_0^2-\eta^2}z}, \tag{39}$$

$$\hat{p}(\eta,z) = -\hat{\tau}^f_{zz}(\eta,z). \tag{40}$$

F. Coupled equations

A coupled system of linear equations for the unknown constants can now be obtained by imposing boundary conditions on the viscoelastic solid at $z=0$ and requiring compatibility of the stresses and displacements in the viscoelastic solid and fluid on the interface at $z=h$.

The transformed boundary conditions on $z=0$ include zero transverse displacement and an imposed velocity distribution. It follows from Eqs. (28) and (29) that these boundary conditions are expressed as

$$\hat{u}_x(\eta,z)|_{z=0} = 0, \tag{41}$$

$$j\omega\hat{u}_z(\eta,z)|_{z=0} = \hat{v}(\eta,z)|_{z=0}, \tag{42}$$

where $\hat{v}(\eta,z)$ is the transformed normal velocity.

The transformed compatibility conditions on $z=h$ can be expressed as

$$\hat{u}^f_z(\eta,z)|_{z=h} = \hat{u}_z(\eta,z)|_{z=h}, \tag{43}$$

$$\hat{\tau}^f_{zz}(\eta,z)|_{z=h} = \hat{\tau}_{zz}(\eta,z)|_{z=h}, \tag{44}$$

$$\hat{\tau}_{xz}(\eta,z)|_{z=h} = 0. \tag{45}$$

The unknown $\hat{C}(\eta)$ in Eq. (36) can be expressed as a function $\hat{A}_1(\eta)$, $\hat{B}_1(\eta)$, $\hat{A}_2(\eta)$, and $\hat{B}_2(\eta)$ by substituting Eqs. (29) and (38) into Eq. (43) and simplifying. Substituting Eqs. (28), (29), (31), (32), and (39) into Eqs. (41), (42), (44), and (45) and simplifying results in the following set of linear equations for the complex unknowns:

$$\begin{bmatrix} t_{11} & 0 & 0 & t_{14} \\ t_{21} & t_{22} & t_{23} & t_{24} \\ 0 & t_{32} & t_{33} & 0 \\ t_{41} & t_{42} & t_{43} & t_{44} \end{bmatrix} \begin{Bmatrix} \hat{A}_1(\eta) \\ \hat{B}_1(\eta) \\ \hat{A}_2(\eta) \\ \hat{B}_2(\eta) \end{Bmatrix} = \begin{Bmatrix} \hat{v}(\eta) \\ 0 \\ 0 \\ 0 \end{Bmatrix}, \tag{46}$$

where the t_{ij} are given in Appendix A.

The spectral impedance in Eq. (9) can now be obtained by substituting Eq. (31) into Eq. (9) and using the solution from Eq. (46).

III. NUMERICAL RESULTS AND DISCUSSION

Numerical results are presented in this section for the self-and mutual impedance of circular pistons. Mechanical, physical, and geometric properties of the fluid and viscoelastic solid are listed in Table I. The solid properties are representative of a typical polyurethane encapsulation material (CONAP EN-7 at 20 C and 50 kHz). The frequency dependence of the viscoelastic solid mechanical properties is not accounted for in the present study. It is common for encapsulation materials to be chosen based on matching the plane dilatational wave characteristic impedance with the fluid. The ratio of characteristic plane-wave impedances for the two layers is expressed as $R=|\rho c_d/\rho_0 c_0|$. For the present study, this ratio is calculated using the properties given in Table I as $R=1.143$.

TABLE I. Material properties.

Property	Value
ρ_0	1000 kg/m^3
c_0	1500 m/s
ρ	1050 kg/m^3
λ	2.44×10^9 N/m^2
μ	$1.80\times10^8(1+j0.150)$ N/m^2
a	0.005 m

The transformed aperture function and the spectral impedances are presented first, followed by self-and mutual impedances. The self-and mutual impedances are investigated over a range of viscoelastic layer thicknesses. Finite-element results, obtained using the commercial software SARA2D,[18] are compared with the associated numerical results obtained using the approach developed in Sec. II.

A. Transformed aperture functions and spectral impedance

The velocity of the piston acting on a viscoelastic solid in a rigid baffle is described by the continuous aperture function. The finite apertures associated with pistons act as a wave number filter to the normal stresses on the piston face. As an illustrative example, consider the continuous aperture described by Eq. (10a). The magnitude of the transformed aperture function, normalized by the zero wave number value, is plotted in Fig. 2 as a function of normalized wave number, ηa. The limiting case of $\delta/a=0$ corresponds to the classical transformed circular piston function of Eq. (12). The roll-off of this aperture function is noted to be proportional to $1/(\eta a)^{3/2}$ for $\eta a \gg 1$. With the exception of $\eta a=0$, the zeros of the aperture function for $\delta/a=0$ occur at the zeros of the Bessel function of the first kind order 1 ($\eta a = 3.832, 7.016, 10.173, \ldots$).

The aperture functions shown in Fig. 2 associated with $\delta/a=0.02, 0.10$, and 0.20 are obtained from Eq. (11). These aperture functions exhibit a roll-off proportional to $1/(\eta a)^{5/2}$

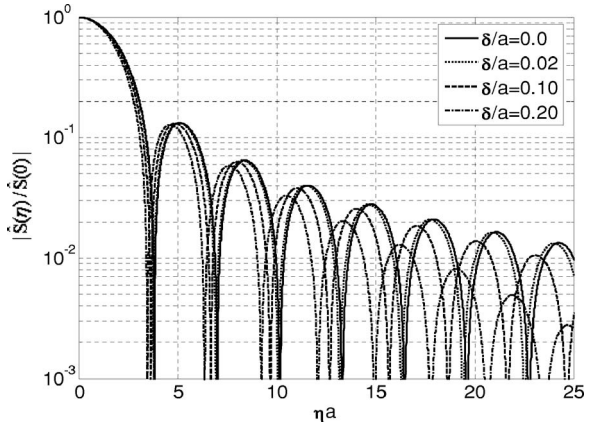


FIG. 2. Transformed circular aperture functions normalized by area.

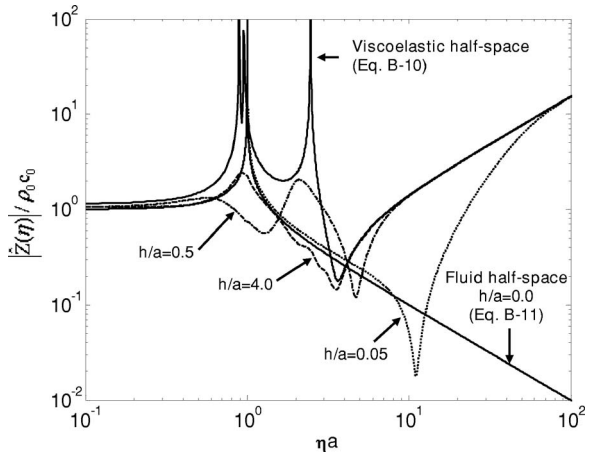


FIG. 3. Magnitude of the normalized spectral impedance, $k_0a=1$.

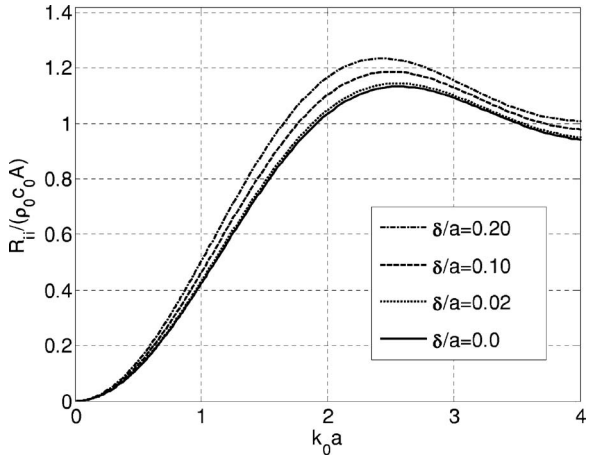


FIG. 4. Influence of the parameter δ on self-radiation resistance of a circular piston acting on a fluid in a rigid infinite baffle with $h/a=0$.

for $\eta a \gg 1$. This spatial filtering contributes to the roll-off in the integration kernel of Eq. (8) as a function of η.

The spectral impedances obtained from Eq. (9), at a normalized frequency of $k_0a=1$, are shown in Fig. 3 as a function of normalized wave number. The results are independent of the wave number normalization factor; however, this factor is retained for continuity with the previous aperture function results and subsequent self-and mutual impedance results. Several normalized thicknesses, including the limiting cases of the fluid half-space ($h/a=0$) and the viscoelastic solid half-space ($h/a=\infty$), are plotted. For the fluid half-space, the peak is readily identified using Eq. (B11) and occurs at the normalized acoustic wave number, $\eta a=k_0a$. The null in the spectral impedance of the viscoelastic half-space at $\eta a=362$ can be found using Eq. (B10) and is associated with the shear wave cutoff frequency. The spectral impedance at intermediate values of thickness exhibits an oscillatory response as a function of wave number that asymptotically approaches the viscoelastic half-space spectral impedance at large ηa. This asymptotic behavior is a result of the viscoelastic solid containing only evanescent waves when $\eta a > k_d a$ and $\eta a > k_s a$. In general, the spectral impedance of the two-layer medium exhibits a local maximum near the acoustic wave number. This results in the normal stress field having a spatial variation with a dominant component at the acoustic wave number. The case of $h/a=0.5$ at $k_0a=1$ is an exception due to the partial cancellation of the normal stress in Eq. (31) in the vicinity of the acoustic wave number. It also is evident that at low wave numbers, the spectral impedances approach a limiting value of the fluid half-space for small thickness (e.g., $h/a=0.05$). Additionally, the spectral impedance for all $h/a>0$ at large ηa is proportional to ηa, whereas the spectral impedance for the fluid half-space ($h/a=0$) is proportional to $1/\eta a$. This asymptotic behavior is an important consideration for the convergence of the integral in Eq. (13). In general, for $h/a>0$, integration to higher wave numbers is required for convergence as δ/a is reduced.

The behavior of the spectral impedance at zero wave number for fluid and viscoelastic solid half-spaces can be obtained from Eqs. (B11) and (B10), respectively. As indicated in Fig. 3, the normalized spectral impedance magnitude, at $\eta a=0$, for the fluid and viscoelastic half-spaces are unity and $R=1.143$, respectively. Further information pertaining to waves in semi-infinite and layered elastic media can be found in standard texts.[17,19,20]

B. Circular piston self-impedance results

Real and imaginary components of the normalized impedance for a circular piston acting directly on a fluid ($h/a=0$) in a rigid infinite baffle are shown in Figs. 4 and 5, respectively, as a function of normalized frequency. These results are obtained from Eq. (14) with the spectral impedance computed from Eq. (B11). The results for $\delta/a=0$, obtained using the present approach, are in agreement with classical circular piston impedance functions.[1] The radiation resistance increases monotonically, for all frequencies, over the range of δ/a investigated. The reactance also exhibits a monotonic dependence on δ/a; however, the increase or decrease from the $\delta/a=0$ results are frequency dependent. Furthermore, it is noted that the resistance and reactance for $\delta/a=0.020$ are within 2.01% and 1.97%, respectively, of the $\delta/a=0$ results.

Real and imaginary components of impedance for a circular piston with $\delta/a=0.02$ and various viscoelastic layer

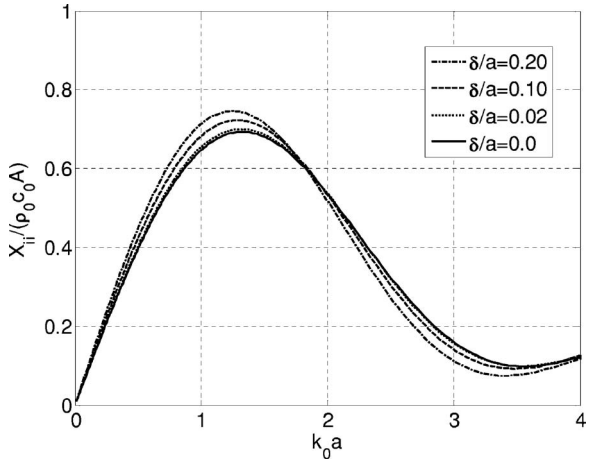


FIG. 5. Influence of the parameter δ on self-radiation reactance of a circular piston acting on a fluid in a rigid infinite baffle with $h/a=0$.

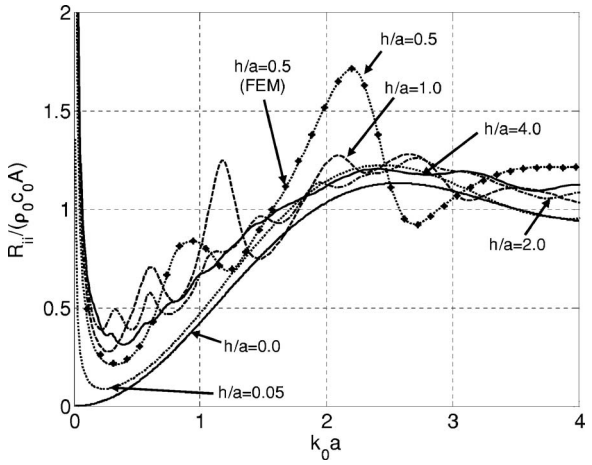


FIG. 6. Self-radiation resistance of a circular piston ($\delta/a=0.02$) in a two-layer medium on an infinite rigid baffle. Numerical (lines) and finite-element method (symbols).

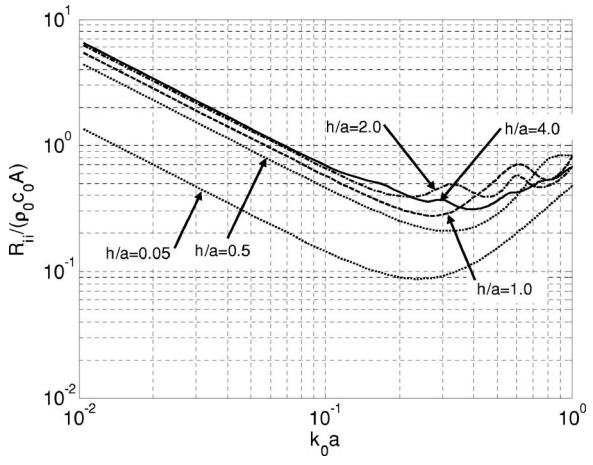


FIG. 8. Log-log plot of self-radiation resistance for a circular piston ($\delta/a=0.02$) in a two-layer medium on an infinite rigid baffle.

thicknesses are shown in Figs. 6 and 7, respectively. These results, obtained using Eq. (14), are plotted along with the associated results for the piston acting directly on a fluid half-space ($h/a=0$). Finite-element results, obtained using SARA2D,[18] are also shown in Figs. 6 and 7 for the case of $h/a=0.5$. The finite-element model utilizes isoparametric axisymmetric continuum elements with quadratic interpolation functions. Acoustic and infinite acoustic elements are utilized to represent the fluid near-field and semi-infinite fluid domain, respectively. The domain in the longitudinal plane is truncated at a radius of $r=20a$. The boundary conditions on the baffle and piston are imposed using kinematic constraints. A high degree of mesh refinement at all boundaries in the vicinity of the piston is required to accurately represent the displacement field associated with the evanescent waves. Finite-element results and those obtained from the approach developed in Sec. II are noted to be in excellent agreement.

The self-radiation resistance and reactance results for $h/a>0$ can deviate substantially from the classical piston impedance functions. The low-frequency behavior of the resistance and reactance (magnitude) are shown in Figs. 8 and 9, respectively, as log-log plots. At low frequency, the piston impedance behaves as a dissipative spring due to shear deformation in the solid layer and exhibits the classical $1/f$ frequency dependence. Furthermore, as shown in Fig. 7, the reactance is negative in the low-frequency region for $h/a>0$ due to the finite real component of the shear modulus (μ) associated with the viscoelastic solid. This is in contrast to the case of a piston acting directly on the fluid half-space, where the reactance decreases as frequency is reduced but always remains positive over all frequencies. In general, for $h/a>0$, the radiation resistances are greater than the $h/a=0$ case. At lower frequencies this is due to the viscoelastic losses associated with shear deformation, and at higher frequencies this is due to the characteristic plane-wave impedance of the viscoelastic layer being greater than the fluid. The radiation reactance in Fig. 7, for $h/a>0$, is generally lower than the $h/a=0$ case. For the fluid half-space, the reactance is positive (mass) for all frequencies. The viscoelastic layer introduces a frequency-dependent negative (stiffness) reactive component in addition to the mass reactance typically associated with pistons acting on a fluid. The fundamental spring-mass resonance of the two-layer medium

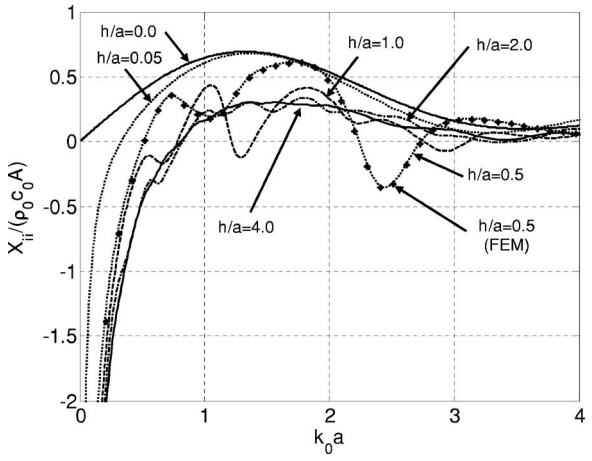


FIG. 7. Self-radiation reactance of a circular piston ($\delta/a=0.02$) in a two-layer medium on an infinite rigid baffle. Numerical (lines) and finite-element method (symbols).

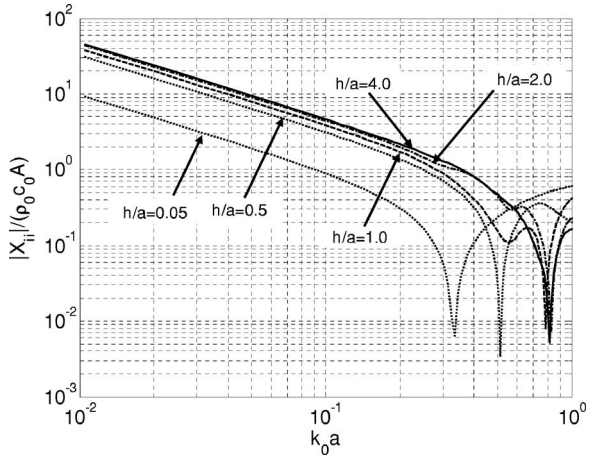


FIG. 9. Log-log plot of the self-radiation reactance for a circular piston ($\delta/a=0.02$) in a two-layer medium on an infinite rigid baffle.

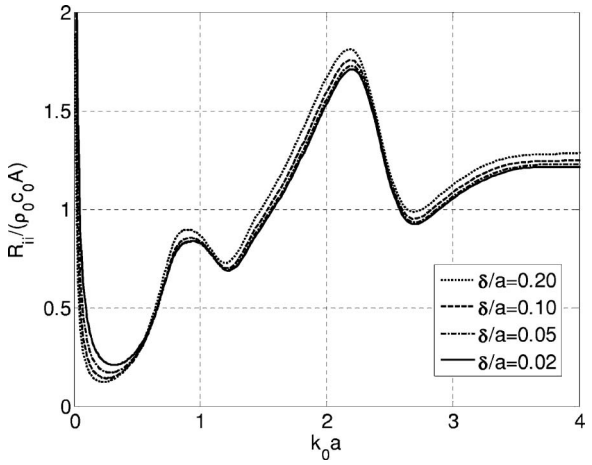


FIG. 10. Influence of the parameter δ on self-radiation resistance of a circular piston acting on two-layer medium on an infinite rigid baffle with $h/a=0.50$.

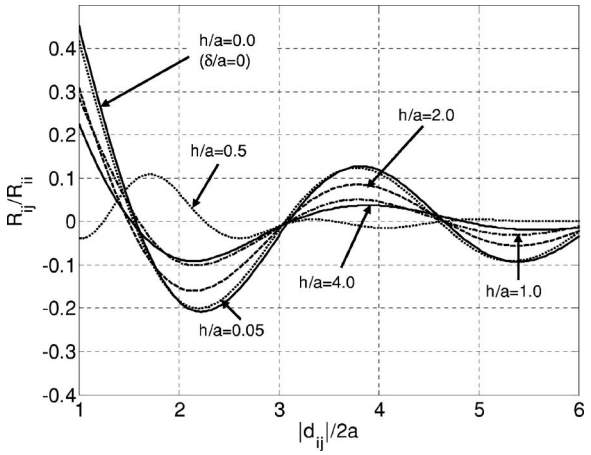


FIG. 12. Real component of the mutual impedance between circular pistons, $k_0a=1$, $\delta/a=0.020$.

occurs at the lowest zero crossing of the reactance curves in Fig. 7. Additional resonances occurring at higher frequencies are the result of the coupled shear and dilatational waves interacting with the boundaries of the viscoelastic layer. As the layer thickness becomes large relative to the shear wavelength (e.g., $h/a=4.0$ and $k_0a>1$), the dissipation associated with propagating shear waves results in an effectively semi-infinite viscoelastic medium.

The influence of the parameter δ on the self-radiation resistance and reactance of a circular piston acting on a two-layer medium with $h/a=0.50$ is shown in Figs. 10 and 11, respectively. The influence on resistance, shown in Fig. 10, is greatest in the low-frequency region where losses associated with the viscoelastic material are the dominant contributor to the real component of the piston impedance. It is clearly evident from Fig. 11 that the reactive impedance is only significantly influenced by the parameter δ in the frequency range of $ka<2$.

C. Circular piston mutual impedance results

The resistive and reactive mutual impedance components of pistons, computed using Eq. (13) at a normalized frequencies of $k_0a=1$, are shown in Figs. 12 and 13 as a function of normalized separation distance. These results are computed for pistons with equal diameter and $\delta/a=0.02$. The associated self-resistance and reactance are used to normalize the mutual impedance results.

The fluid half-space case corresponds to $h/a=0$ and $\delta/a=0$, with the pistons acting directly on a fluid in an infinite baffle. For this special case, the presented results are noted to be in excellent agreement with those of Pritchard.[6] Additionally, it is evident from the $h/a=0.05$ results that the classical limit is approached as the layer thickness as reduced. The results for $h/a=0.5$ appear atypical and suggest a normal stress field with a higher dominant wave number. This higher dominant wave number is clearly evident in Fig. 3 for the $h/a=0.5$ at the $k_0a=1$ case. As previously indicated, there is a partial cancellation of the normal stress in the vicinity of the acoustic wave number for this case, resulting in a higher dominant wave number.

These results can be considered in the context of sonar transducers and arrays. Typically, a narrow-band sonar system will have elements spaced at half-wavelength, based on the highest frequency of operation, to eliminate the presence of grating lobes in the array response pattern. At a normal-

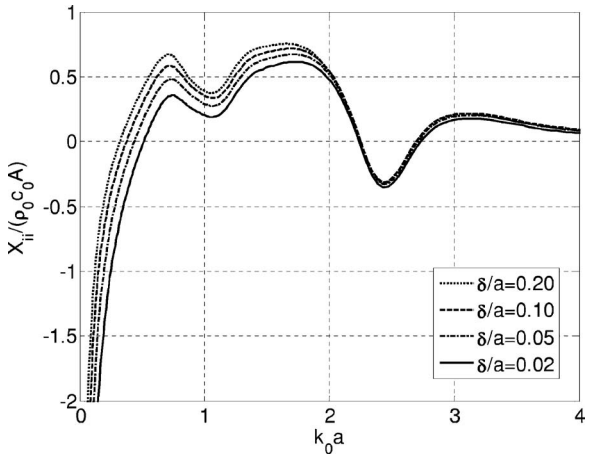


FIG. 11. Influence of the parameter δ on self-radiation reactance of a circular piston acting on two-layer medium on an infinite rigid baffle with $h/a=0.50$.

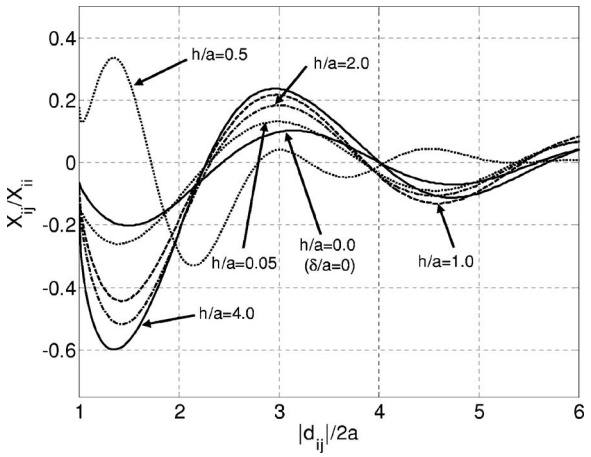


FIG. 13. Imaginary component of the mutual impedance between circular pistons, $k_0a=1$, $\delta/a=0.020$.

ized frequency of $k_0a=1$, half-wavelength spacing corresponds to $|\mathbf{d}_{ij}|/2a=m\pi/2(m=1,2,3,\ldots)$. For spacing less than half of an acoustic wavelength, the mutual resistance is a significant fraction of the self-resistance. In general, the mutual resistances for $h/a>0$ tend to decay more rapidly with separation distance than the classical baffled piston $(h/a=0)$.

IV. SUMMARY

A general approach to solve for the self-and mutual impedances associated with pistons on a two-layer semi-infinite medium, consisting of a fluid on a viscoelastic plate in a rigid infinite baffle, has been developed. The approach utilizes the linear theory of elasticity and integral transform techniques to determine the stresses in the viscoelastic plate as a result of an imposed piston velocity. The approach is validated, by comparison with finite-element model results, for the self-impedances of circular pistons. The limiting case of zero viscoelastic plate thickness is in agreement with classical results for pistons on a planar rigid infinite baffle. The presented results have clearly shown that a viscoelastic layer can significantly influence the resulting self-and mutual impedances. The basic approach can be readily extended to cases involving noncircular piston shapes and multilayered media.

ACKNOWLEDGMENTS

This work was supported by funds from the Office of Naval Research, Code 333.

APPENDIX A

Elements of the coefficient matrix in Eq. (46):

$$t_{11}=j\omega\gamma_d,$$

$$t_{12}=0,$$

$$t_{13}=0,$$

$$t_{14}=-\omega\eta,$$

$$t_{21}=-(\lambda k_d^2+2\mu\gamma_d^2)\sin(\gamma_d h)-\frac{\lambda_0 k_0^2\gamma_d}{j\sqrt{k_0^2-\eta^2}}\cos(\gamma_d h),$$

$$t_{22}=(\lambda k_d^2+2\mu\gamma_d^2)\cos(\gamma_d h)+\frac{\lambda_0 k_0^2\gamma_d}{j\sqrt{k_0^2-\eta^2}}\sin(\gamma_d h),$$

$$t_{23}=2\mu j\eta\gamma_s\cos(\gamma_s h)-\frac{\lambda_0 k_0^2\eta}{\sqrt{k_0^2-\eta^2}}\sin(\gamma_s h),$$

$$t_{24}=2\mu j\eta\gamma_s\sin(\gamma_s h)-\frac{\lambda_0 k_0^2\eta}{\sqrt{k_0^2-\eta^2}}\cos(\gamma_s h),$$

$$t_{31}=0,$$

$$t_{32}=j\eta,$$

$$t_{33}=-\gamma_s,$$

$$t_{34}=0,$$

$$t_{41}=2\mu j\eta\gamma_d\cos(\gamma_d h),$$

$$t_{42}=2\mu j\eta\gamma_d\sin(\gamma_d h),$$

$$t_{43}=\mu(2\gamma_s^2-k_s^2)\sin(\gamma_s h),$$

$$t_{44}=\mu(2\gamma_s^2-k_s^2)\cos(\gamma_s h),$$

APPENDIX B

The spectral impedance of a viscoelastic half-space and fluid half-space are derived using the transformed equations of motion for the displacement potentials. The starting point is Eqs. (22) and (23), rewritten below,

$$\frac{\partial^2\hat{\phi}(\eta,z)}{\partial z^2}+(k_d^2-\eta^2)\hat{\phi}(\eta,z)=0, \tag{B1}$$

$$\frac{\partial^2\hat{\psi}(\eta,z)}{\partial z^2}+(k_s^2-\eta^2)\hat{\psi}(\eta,z)=0, \tag{B2}$$

For the problem of interest, the solutions to the above equations can be expressed for a semi-infinite medium by retaining the outgoing wave solution. This leads to the following expressions for displacement potential:

$$\hat{\phi}(\eta,z)=\hat{G}(\eta)e^{-j\gamma_d z}, \tag{B3}$$

$$\hat{\psi}(\eta,z)=\hat{H}(\eta)e^{-j\gamma_s z}, \tag{B4}$$

where the wave numbers γ_d and γ_s are given in Eqs. (26) and (27), respectively. The associated displacements are obtained from Eqs. (B3) and (B4), using a standard approach,[17] with the following final result:

$$\hat{u}_x(\eta,z)=j\eta\hat{G}(\eta)e^{-j\gamma_d z}+j\gamma_s\hat{H}(\eta)e^{-j\gamma_s z}, \tag{B5}$$

$$\hat{u}_z(\eta,z)=-j\gamma_d\hat{G}(\eta)e^{-j\gamma_d z}+j\eta\hat{H}(\eta)e^{-j\gamma_s z}. \tag{B6}$$

The unknown $\hat{G}(\eta)$ and $\hat{H}(\eta)$ in the above equations are obtained by substituting Eqs. (B5) and (B6) into Eqs. (41) and (42) and solving. The final result is expressed as

$$\hat{G}(\eta)=\frac{\hat{v}(\eta,0)}{\omega}\left(\frac{\eta_s}{\gamma_d\gamma_s+\eta^2}\right), \tag{B7}$$

$$\hat{H}(\eta)=\frac{\hat{v}(\eta,0)}{\omega}\left(\frac{\eta}{\gamma_d\gamma_s+\eta^2}\right). \tag{B8}$$

The transformed normal stress on the baffle $(z=0)$ is obtained by substituting Eqs. (B5) and (B6) into standard linear strain-displacement relationships.[17] The final result is

$$\hat{\tau}_{zz}(\eta,0) = -\frac{\hat{v}(\eta,0)}{\omega}\left(\frac{(\lambda+2\mu)k_d^2}{\gamma_d+\eta^2/\gamma_s}\right). \tag{B9}$$

It follows from Eq. (9) that the spectral impedance of a viscoelastic half-space subjected to zero transverse motion on $z=0$ can be expressed as

$$\hat{Z}^s(\eta) = -\frac{(\lambda+2\mu)}{\omega}\left(\frac{k_d^2}{\gamma_d+\eta^2/\gamma_s}\right). \tag{B10}$$

The spectral impedance of a fluid half-space follows directly from Eq. (B10) by noting that for an inviscid fluid $\mu=0$ and the associated $\gamma_s \to \infty$. After substituting the fluid Lamé parameters for the viscoelastic solid parameters, it follows that the spectral impedance for a fluid half-space can be expressed as

$$\hat{Z}^f(\eta) = \frac{\lambda_0}{\omega}\left(\frac{k_0^2}{\sqrt{k_0^2-\eta^2}}\right). \tag{B11}$$

[1] L. Kinsler and A. Frey, *Fundamentals of Acoustics* (John Wiley and Sons, New York, 1964).
[2] M. Junger and D. Feit, *Sound, Structures, and Their Interaction* (AIP, New York, 1993).
[3] P. Stepanishen, "The time dependent force and radiation impedance on a piston in a rigid infinite planar baffle," J. Acoust. Soc. Am. **49**, 841–849 (1971).
[4] A. Pierce, R. Cleveland, and M. Zampolli, "Radiation impedance matrices for rectangular interfaces within rigid baffles: Calculation methodology and applications," J. Acoust. Soc. Am. **111**, 672–684 (2002).
[5] S. Klapman, "Interaction impedance of a system of circular pistons," J. Acoust. Soc. Am. **11**, 289–295 (1940).
[6] R. Pritchard, "Mutual acoustic impedance between radiators in an infinite baffle," J. Acoust. Soc. Am. **32**, 730–737 (1960).
[7] J. Luco, "Impedance functions for a rigid foundation on a layered medium," Nucl. Eng. Des. **31**, 204–217 (1974).
[8] J. Luco, "Vibrations of a rigid disc on a layered viscoelastic medium," Nucl. Eng. Des. **36**, 325–340 (1976).
[9] X. Zhang, T. J. Royston, H. A. Mansy, and R. H. Sandler, "Radiation impedance of a finite circular piston on a viscoelastic half-space with application to medical diagnosis," J. Acoust. Soc. Am. **109**, 795–802 (2001).
[10] C. Amedin, A. Berry, and Y. Champoux, "Sound field of a baffled piston source covered by a porous medium layer," J. Acoust. Soc. Am. **98**, 1757–1766 (1995).
[11] S. Ko, "Reduction of structure-borne noise using an air-voided elastomer," J. Acoust. Soc. Am. **101**, 3306–3312 (1997).
[12] A. Pathak and P. Stepanishen, "Acoustic harmonic radiation from fluid-loaded infinite elastic plates using elasticity theory," J. Acoust. Soc. Am. **94**, 1700–1710 (1993).
[13] D. Crighton, "The free and forced waves on a fluid-loaded elastic plate," J. Sound Vib. **63**, 225–235 (1979).
[14] A. Freedman, "Effects of fluid-loading on Lamb mode spectra," J. Acoust. Soc. Am. **98**, 3488–3496 (1996).
[15] P. Morse and K. Ingard, *Theoretical Acoustics* (McGraw-Hill, New York, 1968).
[16] J. Naze Tjotta and S. Tjotta, "Nearfield and farfield of pulsed acoustic radiators," J. Acoust. Soc. Am. **71**, 824–834 (1982).
[17] J. Achenbach, *Wave Propagation in Elastic Solids* (North-Holland, New York, 1973).
[18] H. Allik, R. Dees, S. Moore, and D. Pan, SARA-2D User's Manual, Version 99–3, BBN Technologies Inc, 1995.
[19] K. Graf, *Wave Motion in Elastic Solids* (Dover, New York 1991).
[20] W. Ewing, W. Jardetsky, and F. Press, *Elastic Waves in Layered Media* (McGraw-Hill, New York, 1957).

Acoustic beam scattering and excitation of sphere resonance: Bessel beam example

Philip L. Marston[a)]
Department of Physics and Astronomy, Washington State University, Pullman, Washington 99164-2814

(Received 17 January 2007; revised 9 April 2007; accepted 9 April 2007)

The exact partial wave series for the scattering by a sphere centered on an ideal Bessel beam was recently given by Marston ["Scattering of a Bessel beam by a sphere," J. Acoust. Soc. Am. **121**, 753–758 (2007)]. That series is applied here to solid elastic spheres in water and to an empty spherical shell in water. The examples are selected to illustrate the effect of varying the beam's conical angle so as to modify the coupling to specific resonances in the response of each type of sphere considered. The backscattering may be reduced or increased depending on properties of the resonance and of the specular contribution. Changing the conical angle is equivalent to changing the beamwidth. Some applications of the Van de Hulst localization principle to the interpretation of the partial wave series and to the interpretation of the scattering dependence on the beam's conical angle are discussed. Some potential applications to the analysis of the scattering by spheres of more general axisymmetric beams are noted. © *2007 Acoustical Society of America.*
[DOI: 10.1121/1.2735810]

PACS number(s): 43.40.Fz, 43.20.Fn, 43.30.Gv, 43.80.Qf [TDM] Pages: 247–252

I. INTRODUCTION

There is a long history of experimental and theoretical investigations of the acoustic excitation of the resonances of elastic spheres and spherical shells in water.[1–6] In most studies, however, the incident wave was taken to be a plane wave or, in the case of experiments, the variations in the amplitude of the incident wave were insignificant over the transverse dimension of the target. When the acoustic illumination is in the form of a localized *acoustic beam*, it is anticipated that the excitation of resonances will depend on the beam parameters. While there are applications of acoustic beams to sonar[7–9] and to medical imaging,[10] the effect of beam parameters on the response of targets is relatively unexplored. The present investigation illustrates modifications to resonance excitation in the case of a Bessel beam[11–14] centered on a spherical target of radius a. The exact partial wave series for this scattering in an ideal fluid having a sound speed c was recently derived by Marston.[13] In addition to the dimensionless frequency ka where $k=\omega/c$, the width of the beam relative to the diameter of the sphere is characterized by one additional parameter which is taken to be the cone angle β of the Bessel beam. That angle is illustrated in Fig. 1 (from Marston[14]). The complex acoustic pressure in an ideal Bessel beam is[11–14]

$$p_B(R,z) = p_0 \exp(i\kappa z) J_0(\mu R), \tag{1a}$$

$$\mu = k \sin(\beta), \tag{1b}$$

where p_0 is the axial amplitude, z is the axial coordinate, $R = \sqrt{(x^2+y^2)}$ is the transverse coordinate, J_0 is a Bessel function, and $\kappa = \sqrt{(k^2-\mu^2)}$. The physical acoustic pressure is given by the real part of $p_B(R,z)\exp(-i\omega t)$. As explained in Sec. VII of Marston,[13] the ratio of the beam half width of the central spot to the sphere radius increases with decreasing β. The plane-wave limit is recovered by taking $\beta=0$. The purpose of this paper is to show the pronounced modification of the excitation of three types of sphere resonances for spheres in water with appropriate selection of the parameter β. The types of resonances considered are: (i) those associated with subsonic waves on a solid polymer sphere,[6] (ii) those associated with supersonic Rayleigh-like waves on a solid sphere having a large shear modulus,[2,3] and (iii) those associated with low frequency deformation of an empty stainless steel shell.[4,5] These targets were selected because the resonances are associated with guided waves having differing phase velocities and because of the ease of comparison with prior studies involving plane-wave illumination. Resonances are referred to as subsonic or supersonic depending on whether the guided wave phase velocity is less than or greater than the speed of sound in the surrounding water. In addition to providing physical insight into the scattering of acoustic beams, this investigation may provide a useful test of finite-element codes for the evaluation of scattering.[15]

II. REVIEW OF THE PARTIAL WAVE SERIES FOR THE SCATTERING

For an isotropic sphere centered at the origin and illuminated by a Bessel beam in an ideal inviscid fluid, the exact solution for the far-field complex scattered pressure expressed using a dimensionless form function F is [Eqs. 6 and 9 of Marston].[13]

$$p_s(r,\theta) = (a/2r) p_0 F e^{ikr}, \tag{2}$$

a)Electronic mail: marston@wsu.edu

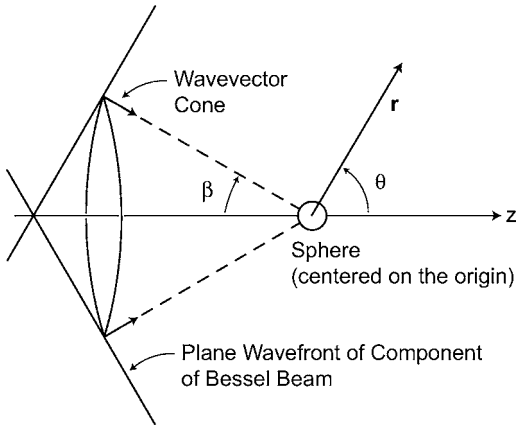


FIG. 1. Geometry of the scattering calculation. The sphere is centered on the z axis of an ideal Bessel beam described by Eq. (1). As discussed in Refs. 11–14, the beam may be represented by a superposition of plane waves having a cone angle β. The scattering angle relative to the beam axis is denoted by θ.

$$F(ka,\cos\theta,\cos\beta) = (-\mathrm{i}/\mathrm{k}a)\sum_{n=0}^{\infty}(2n+1)(s_n-1) \times P_n(\cos\theta)P_n(\cos\beta), \quad (3)$$

where the scattering angle relative to the z axis is denoted by θ, and the P_n are Legendre polynomials. The s_n are the known scattering functions from the corresponding plane wave scattering,[13] Eq. (3) with the factor $P_n(\cos\beta)$ omitted, The nth partial wave coefficient is commonly designated by $a_n=(s_n-1)/2$. In the plane wave limit $\beta=0$ and $P_n(\cos\beta)=1$ for all n so that Eq. (3) gives the standard series for plane wave scattering. In the case of an ideal sphere causing no dissipation of acoustic energy, the complex functions s_n are unimodular: $|s_n|=1$. For the numerical examples considered here, it is a useful approximation to take $|s_n|=1$ so that $s_n=-D_n^*/D_n$, where, D_n is the denominator of the partial wave coefficient and an asterisk denotes complex conjugation. Expressions for D_n are reviewed in Appendices A and B for the solid sphere[2] and empty shell[16] cases, respectively. These involve spherical Bessel, Neumann, and Hankel functions, j_n, n_n, and $h_n^{(1)}$, and the material properties including the densities of the solid and of water, ρ_E and ρ, and the longitudinal and transverse wave velocities in the solid, c_L and c_T. In the computations given here, the series in Eq. (3) was be terminated for n in excess of 1.6 ka, a convergence condition discussed by Ihlenburg.[17] The excess was typically more than four terms.

Appendix B of Marston[13] gives a partial wave expansion of the incident Bessel beam. That expansion is not essential to the derivation of Eq. (3) since Eq. (3) follows directly from superposition.[13] That expansion agrees numerically with Eq. (1).[13]

III. INTERPRETATION USING THE LOCALIZATION PRINCIPLE

The modification to the partial wave series, Eq. (3), resulting from Bessel beam illumination is the factor $P_n(\cos\beta)$. When β is small the factor $P_n(\cos\beta)$ has a simple interpretation using the Van de Hulst localization principle which associates the nth partial wave with scattering contributions associated with a localized impact parameter.[18] (The impact parameter is the offset of an incident ray from the center of the sphere.) Williams and Marston[19] describe applications of this principle to acoustical scattering and note the following approximation (given by Szego[20] and by Nussenzveig[21]):

$$P_n(\cos\beta) = (\beta/\sin\beta)^{1/2}J_0[(n+(1/2))\beta] + O(n^{-3/2}), \quad (4)$$

where $n\gg 1$, $\beta\ll 1$ rad, and O denotes the order of the leading correction. As noted in Eq. (1), $\beta=\arcsin(\mu/k)$. When β is small and n is large it follows that

$$P_n(\cos\beta) \approx J_0(\mu R) \quad (5)$$

provided that n and R are related by

$$R \approx [n+(1/2)]/k. \quad (6)$$

The significance of inserting Eq. (5) into Eq. (3) is that $P_n(\cos\beta)$ is replaced by the local relative Bessel beam amplitude in the plane $z=0$ from Eq. (1). One interpretation is that the incident beam may be represented as a bundle of parallel rays having an amplitude proportional to $J_0(\mu R)$. Equation (6) is the standard localization principle relationship between n and the impact parameter R given by Van de Hulst.[18] The relationships (5) and (6) are noted for heuristic reasons and not as a replacement of the derivation of Eq. (3) given in Marston.[13]

IV. SUBSONIC RESONANCES OF SOLID PMMA SPHERES

Bessel beam modifications to the coupling to resonances of solid acrylic spheres are examined in this section. Hefner and Marston[6] (referred to as HM) observed such resonances in the backscattering by polymethylmethacrylate (PMMA) spheres suspended in water by a nylon fishing line and illuminated by waves similar to plane waves. The computations given here use the material properties they considered: $c_L=2690$ m/s, $c_T=1340$ m/s, and $\rho_E=1.19$ g/cm^3 for PMMA, and $c=1479$ m/s and $\rho=1.00$ g/cm^3 for water. The solid curve shown in Fig. 2 is the backscattering ($\theta=\pi$) computed from Eq. (3) for plane wave illumination ($\beta=0$). The curve agrees with results in HM neglecting absorption. Because ka is small, the curve is also similar to measurements and to computations that include PMMA absorption (HM Fig. 7). Hickling has computed a similar curve with properties differing slightly from the ones used here.[22] HM explain that the sequence of peaks close to (or exceeding) $|F|=5$ are associated with subsonic waves guided by the sphere and give a quantitative ray model. Numbering from left-to-right in Fig. 2, the sequence of peaks are primarily associated with partial waves having $n=2$, 3, 4, and 5. The peaks are nearly equally spaced because the phase velocity of the guided wave (designated here as c_R) depends only weakly on ka in this range. See Fig. 4 of HM.

The dashed curves in Fig. 2 show modifications to the backscattering resulting from Bessel beam illumination with different values of β. To facilitate the discussion which follows it is convenient to define β_n as the lowest root of

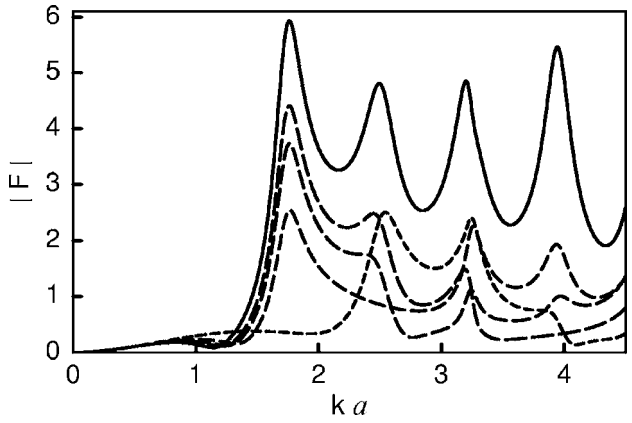


FIG. 2. Backscattering form function modulus $|F|$ computed as a function of ka for an elastic PMMA sphere in water having different values of the beam conical angle β. The solid curve is the plane-wave result discussed by Hefner and Marston (Ref. 6). As described in Sec. IV the dashed curves have $\beta \approx \beta_n$ for $n=2$, 3, 4, and 5 beginning with the curve having the shortest dashes ($\beta \approx \beta_2$) and ending with the curve having the longest dashes ($\beta \approx \beta_5$). The resonance associated with the nth partial wave is suppressed when $\beta \approx \beta_n$. The peak in the solid curve near $ka=3.26$ is associated primarily with the partial wave with $n=4$. When $\beta \approx \beta_4$ there is some residual structure in $|F|$ near $ka=3.26$ because of the dependence on ka of the partial wave with $n=1$.

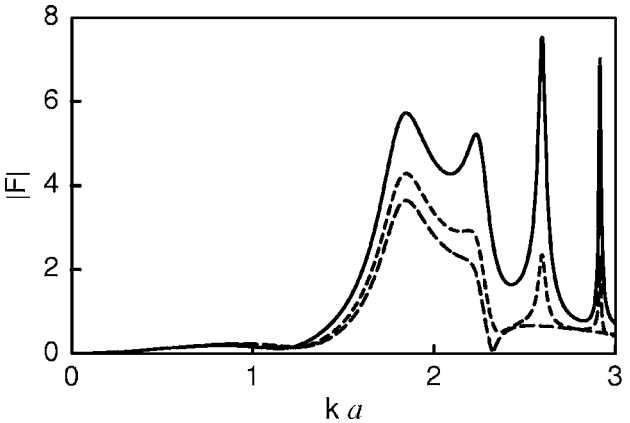


FIG. 3. Backscattering form function modulus $|F|$ computed as a function of ka for an empty stainless steel spherical shell having an inner to outer radius $b/a=0.95$. The solid curve is the plane wave ($\beta=0$) case and the curves having long and short dashes correspond to $\beta=\beta_4$ and $\beta=\beta_5$ as discussed in Sec. V.

$P_n(\cos\beta)=0$. The 6-digit approximations to β_n for $n=2$, 3, 4, and 5 are $\beta_n=54.7346°$, 39.2315°, 30.5556°, and 25.0173°. Application of one of these approximations will be indicated as $\beta \approx \beta_n$ with $n=2$, 3, 4, and 5. Inspection of those curves shows that the resonance with the nth partial wave is suppressed when $\beta \approx \beta_n$. This is most easily seen from the curve having the shortest dashes which has $\beta \approx \beta_2$. This behavior is in agreement with the form of Eq. (3) because of the dependence on $P_n(\cos\beta)$. While Fig. 2 illustrates the effect of suppressing a resonance on the backscattering, there is also a corresponding suppression of the contribution of the nth partial wave to the bistatic scattering. An example of this suppression is shown in Sec. VI.

The terminology used to describe the relevant waves guided by surfaces of the elastic spheres considered in Secs. IV and VI may be summarized as follows. Following Überall[23] a wave is described as a (generalized) Rayleigh wave when appreciable energy resides in the solid and the usual Rayleigh wave velocity is recovered in the limit of very large ka and negligible fluid loading. With this convention a Stoneley wave is a wave for which the strain energy of an immersed solid is relatively localized. This convention has also been used in seismology.[24] It is noted here for ease of comparison with discussions of relevant wave properties by Hefner and Marston[6] and by Williams and Marston.[3,19] For justifiable reasons, however, some authors refer to subsonic resonances of polymer spheres or cylinders in water as being associated with a Scholte-Stoneley wave.[25] What is important for the present discussion is the existence of weakly damped guided waves and not the specific terminology used to describe them. Ansell[26] discusses some of the relevant wave properties in the flat-surface limit.

V. SUBSONIC RESONANCES OF A STEEL SHELL

Consider now the case of an empty hollow stainless steel spherical shell.[4,5] The example considered here is similar to a shell that has been used in underwater acoustics experiments.[27] The ratio of the inner radius to outer radius is taken to be $b/a=0.95$ and the material and water properties used were: $c_L=5790$ m/s, $c_T=3100$ m/s, and $\rho_E=7.90$ g/cm^3 for steel, and $c=1482$ m/s and $\rho=1.00$ g/cm^3 for water. The solid curve in Fig. 3 is the backscattering ($\theta=\pi$) computed from Eq. (3) for plane wave illumination ($\beta=0$). The general features are similar to those described by various authors for metallic shells having similar values of b/a. Some authors designate the wave associated with the prominent resonances as the a_{0-} Lamb wave because of similarities with the lowest antisymmetric (flexural) branch of generalized Lamb waves on a fluid loaded plate.[4,28,29] As is the case in Fig. 2, numbering from left-to-right in Fig. 3, the sequence of prominent peaks are primarily associated with partial waves having $n=2$, 3, 4, and 5. Unlike Fig. 2, however, the peaks are unequally spaced since the phase velocity of the a_{0-} wave is strongly dependent on ka for this range of ka since the shell is sufficiently thin that $(a-b) \ll a$. For the scattering from some ideal thin shells evaluated for loading by ideal (inviscid) water the radiation damping of some of the highly subsonic resonances is very small.[4,5,28,29] For some resonances a localized computational step size as small as $\Delta ka=10^{-10}$ has been used to completely resolve the *width* of the narrow features.[28] Such extremely narrow features are rarely (if ever) observed since real steel shells in water will have imperfections or dissipation. For the purposes of the present discussion, the computational increment of $\Delta ka=10^{-3}$ used here is adequate for indicating the relevant responses.

The dashed curves in Fig. 3 show modifications to the backscattering resulting from Bessel beam illumination with $\beta=\beta_n$ where, as in Sec. IV, $P_n(\cos\beta_n)=0$. Only the cases with $n=5$, short dashes, and $n=4$, long dashes are shown to simplify Fig. 3. (For this example the relevant roots β_n were evaluated to better than 1 part in 10^{10}.) Inspection of Fig. 3 shows that when $\beta=\beta_5$ the narrow $n=5$ resonance is completely suppressed. Figure 3 also shows that when $\beta=\beta_4$ the slightly broader $n=4$ resonance is completely suppressed and there is almost a null in the backscattering for $ka=2.323$.

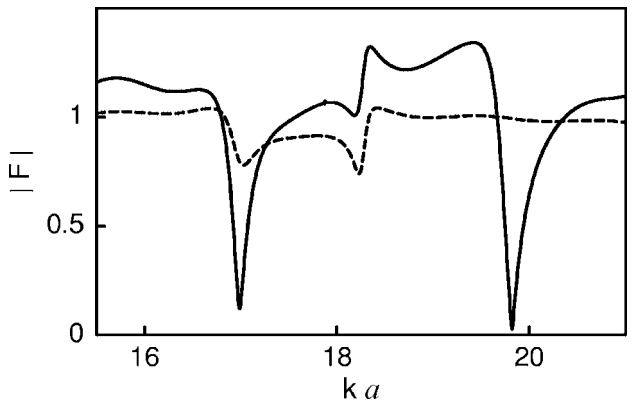


FIG. 4. Backscattering form function modulus $|F|$ computed as a function of ka for an elastic tungsten carbide sphere in water. The solid curve is the plane wave case ($\beta=0$) discussed using resonance scattering theory by Gaunaurd and Überall (Ref. 2) and quantitative ray theory by Williams and Marston (Ref. 3). The dashed curve is with $\beta \approx \beta_6$ so as to suppress the partial wave with $n=6$. The pronounced dip in the solid curve at $ka=19.82$ is filled in when $\beta \approx \beta_6$.

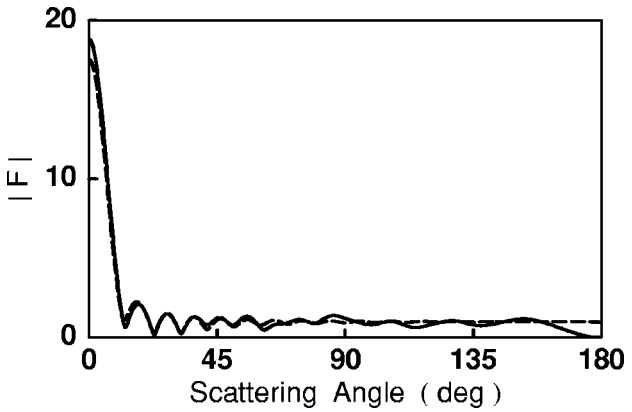


FIG. 5. The solid curve is the angular distribution of the scattering amplitude $|F|$ from Eq. (3) for plane wave illumination ($\beta=0$) for the tungsten carbide sphere when $ka=19.82$. That ka corresponds to the pronounced dip in the solid curve in Fig. 4. The dashed curve is an analogous plot for a rigid sphere. The region with $\theta>60°$ in the rigid case has $|F| \approx 1$ as a consequence of specular reflection from the sphere.

While it is not shown here (so as to simplify Fig. 3), it was verified that the $n=2$ and 3 resonances were suppressed by taking $\beta=\beta_2$ and β_3. As with the example in Sec. IV, suppression of one of the large subsonic resonances also produces an appreciable modification of the bistatic scattering.

VI. SUPERSONIC RESONANCES OF A TUNGSTEN CARBIDE SPHERE

The solid sphere example considered in Sec. IV was atypical in that the material (PMMA) is such that the shear wave velocity was less than the speed of sound in water. A more widely investigated case is that of hard materials such as metals, ceramics, glass, or rocks for which the shear wave speed significantly exceeds the speed of sound in water. A widely investigated example is the case of tungsten carbide, which is a hard dense material.[2,3,19,30,31] Unlike the PMMA case, the guided wave velocity significantly exceeds the speed of sound in water and the generalization of the Rayleigh wave guided by the water-loaded sphere is classified as a supersonic leaky wave.[3,19,23] Unlike the case of aluminum,[22] however, tungsten carbide is sufficiently dense that the radiation damping is weak so that the resonances remain narrow over a *wide range* of ka. The material and water properties used here are: $c_L=6860$ m/s, $c_T=4185$ m/s, and $\rho_E=13.8$ g/cm^3 for tungsten carbide, and $c=1476$ m/s and $\rho=1.00$ g/cm^3 for water. Williams and Marston[4] (WM) found from ray theory that the resonances could cause either peaks or depressions in the backscattering depending on whether the specular reflection was in phase (for a peak) or out-of-phase (for a depression) with the guided-wave contribution. The solid curve shown in Fig. 4 is the backscattering ($\theta=\pi$) computed from Eq. (3) for plane wave illumination ($\beta=0$). This curve agrees with one computed by WM.[4] In the range of ka shown the dips near ka of 16.99 and 19.82 correspond to Rayleigh wave resonances having respective n of 5 and 6. The smaller feature near ka of 18 is associated with a higher-order internal wave commonly referred to as a whispering gallery wave.[4]

Figure 5 shows the angular scattering pattern for plane wave illumination ($\beta=0$) computed at $ka=19.82$. While the forward region has some similarity with the rigid sphere plane-wave result, the dashed curve, comparison of the region with $\theta>60°$ shows significant qualitative differences associated with the excitation of the Rayleigh wave resonance. Most noticeably, the backscattering in the angular region near 180° is depressed in Fig. 5 as a consequence of the aforementioned destructive interference between the specular and the Rayleigh wave contributions.

A computational method widely used to gain insight into the elastic response of simple targets is the method of[1,2] "background subtraction." Generalized to the case of a Bessel beam, the complex *background subtracted form function* becomes

$$F^{\mathrm{sub}}(ka,\cos\theta,\cos\beta) = F(ka,\cos\theta,\cos\beta) - F^{\mathrm{b}}(ka,\cos\theta,\cos\beta) \quad (7)$$

where the background F^b is given by Eq. (3) with $s_n(ka)$ replaced by a corresponding background function $s_n^b(ka)$ for the intended background target. The background target is selected in such a way that the F^{sub} displays the *elastic response* of interest. In the present situation a fixed *rigid sphere* is a useful background. That choice gives[1,2] $s_n^b(ka) = -h_n^{(2)}(ka)'/h_n^{(1)}(ka)'$ where primes denote differentiation with respect to the indicated argument. The curve having long dashes in Fig. 6 shows $|F^{\mathrm{sub}}|$ with plane wave illumination ($\beta=0$) evaluated for a tungsten carbide sphere at $ka=19.82$, corresponding to the conditions of the dip in the solid curve in Fig. 4. This curve has six nodes confirming the identification of the mode as associated with s_6. The curve deviates from the form of $|P_6|$ because of weak elastic contributions not included in the background.

Consider now the consequences of illumination with a Bessel beam with $\beta=21.1769°$, the six-digit approximation to β_6, the first root of $P_6(\cos\beta)=0$. The dashed curve in Fig. 4 gives $|F|$ for backscattering. The dip at $ka=19.82$ associated with the resonance behavior of s_6 is completely suppressed in agreement with the suppression of the term having

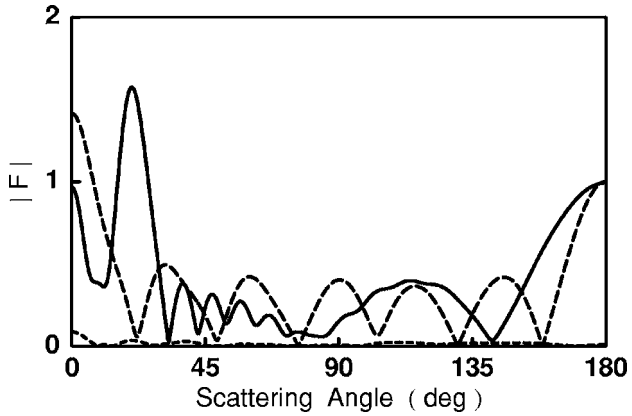


FIG. 6. The curve with the long dashes is the angular distribution of the background subtracted form function given by $|F^{\text{sub}}|$ from Eq. (8) with plane wave illumination ($\beta=0$). This is shown for ka of 19.82 for the sphere considered in Fig. 4. The other curves are also with ka of 19.82 but now (as for the dashed curve in Fig. 4) with $\beta\approx\beta_6$. The solid curve is $|F|$ from Eq. (3) and the curve having short dashes is $|F^{\text{sub}}|$ from Eq. (7).

$n=6$ in Eq. (3). The solid curve in Fig. 6 shows the angular dependence of $|F|$ for $ka=19.82$ and $\beta=21.1769°$. Since the resonance is suppressed the general appearance is similar to perfectly reflecting sphere examples (Marston,[13] Figs. 3 and 5–7) and the interpretation of those features parallels the discussion in Marston. The highly suppressed curve with short dashes in Fig. 6 gives $|F^{\text{sub}}|$ for the previously noted ka and β. The small magnitude of this curve confirms that the solid curve is nearly that for a rigid sphere.

VII. DISCUSSION

The β value for the suppression of a guided-wave resonance may be estimated using the localization principle discussed in Sec. III provided the known ka of the resonance is not small and a good estimate of the phase velocity of the guided wave is available. In this case take $R=\sqrt{(x^2+y^2)}$ to be the impact parameter associated with *excitation* of a given guided wave. Ray theory gives[3] $R=ac/c_{\text{gw}}$ where c_{gw} denotes the phase velocity of the guided wave along the outer surface of the sphere. (The ray theory applies to subsonic[6,28] as well as supersonic[3,16,19] guided waves. See, e.g., Fig. 3 of HM[6] and Fig. 1 of WM.[3]) With Bessel beam illumination, excitation of a given guided wave is suppressed when the exciting ray lies on a *pressure null* of the beam. The β of the incident Bessel beam is selected so that at the ka of interest, μ is such that $\mu R=2.405$, the first root of $J_0(\mu R)=0$. Since Eq. (1) gives $\beta=\arcsin(\mu/k)$, the required β becomes

$$\beta\approx\arcsin(2.405c_{\text{gw}}/kac). \tag{8}$$

Consider the examples from Secs. IV and VI where for simplicity the guided waves are both classified as a generalized Rayleigh wave and c_{gw} is denoted by c_R. For the resonance at $ka=19.82$ of the tungsten carbide sphere in Fig. 4, c_R/c is approximately[19] $ka/[n+(1/2)]\approx 19.82/6.5=3.034$ and Eq. (8) gives $\beta=21.62°$. This is close to the value 21.1769° used in the evaluation of the dashed line in Fig. 6. A related approach gives[4] $c_R/c\approx \text{ka}/\sqrt{[n(n+1)]}\approx 3.043$ and $\beta=21.67°$ from Eq. (8). Similarly c_R/c for the PMMA sphere considered in Fig. (2) is estimated from Fig. 4 of HM[6] to be 0.72 near the resonance at $ka=3.95$: Eq. (8) gives $\beta\approx 26.0°$ which is close to the value of 25.0173° used in the evaluation of the curve having long dashes in Fig. 2. In the localization-principle representation, the phase change of a partial wave contribution in Eq. (3) associated with a null of $P_n(\cos\beta)$ becomes associated with a zero crossing of $J_0(\mu R)$.

While numerical examples are outside the scope of this discussion, a plausible application of this scattering formulation is the local representation of axisymmetric beams as a superposition of Bessel beams in the region near $z=0$. Taking $R=\sqrt{(x^2+y^2)}$, with that superposition, the complex incident wave becomes

$$p_{\text{inc}}(R,z)=\int_{\beta=0}^{\pi/2} B(\beta)J_0(kR\sin\beta)\exp(ikz\cos\beta)d\beta, \tag{9}$$

where $B(\beta)$ is the complex *conical angular spectrum* of the generalized beam. The far-field scattering from Eqs. (2) and (3) becomes

$$p_s(r,\theta)=(a/2r)e^{ikr}\int_{\beta=0}^{\pi/2} B(\beta)F(ka,\cos\theta,\cos\beta)d\beta. \tag{10}$$

Let δ denote a Dirac δ function. Setting $B(\beta)=p_0\delta(\beta-\beta_{\text{BB}})$ in Eqs. (9) and (10) recovers the results in Eqs. (2) and (3) with β for the pure Bessel beam replaced by β_{BB}. It is anticipated from the form of Eq. (9) that for a general axisymmetric beam, because of the range of conical angles present in the incident wave, the suppression of specific partial waves will not be complete as in the Bessel beam cases illustrated here.

ACKNOWLEDGMENT

This research was supported by the Office of Naval Research.

APPENDIX A: DETERMINANT FOR A SOLID SPHERE

For the solid sphere D_n is the determinant of the 3-by-3 matrix having the following elements.[2] Letting $X=ka$, $Y=(c/c_{\text{L}})X$, $Z=(c/c_T)X$, and $N=n(n+1)$, and using primes to denote differentiation, the elements are: $d_{11}=(\rho/\rho_E)Z^2h_n^{(1)}(X)$, $d_{12}=(2N-Z^2)j_n(Y)-4Yj_n(Y)'$, $d_{13}=2N[Zj_n(Z)'-j_n(Z)]$, $d_{21}=-Xh_n^{(1)}(X)'$, $d_{22}=Yj_n(Y)'$, $d_{23}=Nj_n(Z)$, $d_{31}=0$, $d_{32}=2[j_n(Y)-Yj_n(Y)']$, $d_{33}=2Zj_n(Z)'+[Z^2-2N+2]j_n(Z)$.

APPENDIX B: DETERMINANT FOR AN EMPTY SHELL

Let a and b denote the outer and inner radius of the shell and use the notation from Appendix A augmented by $U=(c/c_L)kb$ and $W=(c/c_T)kb$. For an evacuated shell D_n is the determinant of the 5-by-5 matrix with elements as in Appendix A and the remaining elements given by:[16] $d_{14}=(2N-Z^2)n_n(Y)-4Yn_n(Y)'$, $d_{15}=2N[Zn_n(Z)'-n_n(Z)]$, $d_{24}=Yn_n(Y)'$, $d_{25}=Nn_n(Z)$, $d_{34}=2[n_n(Y)-Yn_n(Y)']$, $d_{35}=2Zn_n(Z)'+[Z^2-2N+2]n_n(Z)$, $d_{41}=0$, $d_{42}=(2N-W^2)j_n(U)-4Uj_n(U)'$, $d_{43}=2N[Wj_n(W)'-j_n(W)]$, $d_{44}=(2N-W^2)n_n(U)-4Un_n(U)'$, $d_{45}=2N[Wn_n(W)'-n_n(W)]$, $d_{51}=0$, $d_{52}=j_n(U)-Uj_n(U)'$, $d_{53}=2Wj_n(W)'+[W^2-2N+2]j_n(W)$, $d_{54}=2[n_n(U)-Un_n(U)']$, $d_{55}=2Wn_n(W)'+[W^2-2N+2]n_n(W)$.

[1]L. Flax, L. R. Dragonette, and H. Überall, "Theory of elastic resonance excitation by sound scattering," J. Acoust. Soc. Am. **63**, 723–731 (1978).
[2]G. C. Gaunaurd and H. Überall, "RST analysis of monostatic and bistatic acoustic echoes from an elastic sphere," J. Acoust. Soc. Am. **73**, 1–12 (1983).
[3]K. L. Williams and P. L. Marston, "Synthesis of backscattering from an elastic sphere using the Sommerfeld–Watson transformation and giving a Fabry–Perot analysis of resonances," J. Acoust. Soc. Am. **79**, 1702–1708 (1986).
[4]G. S. Sammelmann, D. H. Trivett, and R. H. Hackman, "The acoustic scattering by a submerged shell. I. The bifurcation of the dispersion curve for the antisymmetric Lamb wave," J. Acoust. Soc. Am. **85**, 114–124 (1989).
[5]M. Talmant, H. Überall, R. D. Miller, M. F. Werby, and J. W. Dickey, "Lamb waves and fluid-borne waves on water-loaded, air-filled thin spherical shells," J. Acoust. Soc. Am. **86**, 278–289 (1989).
[6]B. T. Hefner and P. L. Marston, "Backscattering enhancements associated with subsonic Rayleigh waves on polymer spheres in water: Observation and modeling for acrylic spheres," J. Acoust. Soc. Am. **107**, 1930–1936 (2000).
[7]G. C. Gaunaurd, "Monostatic and bistatic cross sections of a large (capped) sphere partially insonified at a circular spot," J. Acoust. Soc. Am. **61**, 1121–1132 (1977).
[8]G. C. Gaunaurd and H. Überall, "Acoustics of finite beams," J. Acoust. Soc. Am. **63**, 5–16 (1978).
[9]K. G. Foote, D. Chu, T. R. Hammar, K. C. Baldwin, L. A. Mayer, L. C. Hufnagle, Jr., and J. M. Jech, "Protocols for calibrating multibeam sonar," J. Acoust. Soc. Am. **117**, 2013–2027 (2005).
[10]P. N. T. Wells, "Ultrasound imaging," Phys. Med. Biol. **51**, R83–R98 (2006).
[11]J. Durnin, "Exact solutions for nondiffracting beams. I. The scalar theory," J. Opt. Soc. Am. A **4**, 651–654 (1987).
[12]J. Durnin, J. J. Miceli, Jr., and J. H. Eberly, "Diffraction-free beams," Phys. Rev. Lett. **58**, 1499–1501 (1987).
[13]P. L. Marston, "Scattering of a Bessel beam by a sphere," J. Acoust. Soc. Am. **121**, 753–758 (2007).
[14]P. L. Marston, "Axial radiation force of a Bessel beam on a sphere and direction reversal of the force," J. Acoust. Soc. Am. **120**, 3518–3524 (2006).
[15]L. L. Thompson, "A review of finite-element methods for time-harmonic acoustics," J. Acoust. Soc. Am. **119**, 1315–1330 (2006).
[16]S. G. Kargl and P. L. Marston, "Observations and modeling of the backscattering of short tone bursts from a spherical shell: Lamb wave echoes, glory, and axial reverberations," J. Acoust. Soc. Am. **85**, 1014–1028 (1989).
[17]F. Ihlenburg, *Finite Element Analysis of Acoustic Scattering* (Springer, New York, 1998), pp. 34–35.
[18]H. C. Van de Hulst, *Light Scattering by Small Particles* (Dover, New York, 1981), p. 208.
[19]K. L. Williams and P. L. Marston, "Axially focused (glory) scattering due to surface waves generated on spheres: Model and experimental confirmation using tungsten carbide sphere," J. Acoust. Soc. Am. **78**, 722–728 (1985).
[20]G. Szego, *Orthogonal Polynomials* (American Mathematical Society, New York, 1939).
[21]H. M. Nussenzveig, "High-frequency scattering by a transparent sphere. 2. Theory of the rainbow and the glory," J. Math. Phys. **10**, 125–176 (1969).
[22]R. Hickling, "Analysis of echoes from a solid elastic sphere in water," J. Acoust. Soc. Am. **34**, 1582–1592 (1962).
[23]H. Überall, "Surface waves in acoustics," Phys. Acoust. **10**, 1–60 (1973).
[24]D. Gubbins, *Seismology and Plate Tectonics* (Cambridge University Press, Cambridge, 1992), pp. 73–77.
[25]F. Chati, F. Léon, and G. Maze, "Acoustic scattering by a metallic tube with a concentric solid polymer cylinder coupled by a thin water layer. Influence of the thickness of the water layer on the two Scholte–Stoneley waves," J. Acoust. Soc. Am. **118**, 2820–2828 (2005).
[26]J. H. Ansell, "The roots of the Stoneley wave equation for solid-liquid interfaces," Pure Appl. Geophys. **94**, 172–188 (1972).
[27]H. J. Simpson, B. H. Houston, and R. Lin, "Laboratory measurements of sound scattering from a buried sphere above and below the critical angle," J. Acoust. Soc. Am. **113**, 39–42 (2003).
[28]P. L. Marston and N. H. Sun, "Resonance and interference scattering near the coincidence frequency of a thin spherical shell: An approximate ray synthesis," J. Acoust. Soc. Am. **92**, 3315–3319 (1992).
[29]G. Kaduchak, C. S. Kwiatkowski, and P. L. Marston, "Measurement and interpretation of the impulse response for backscattering by a thin spherical shell using a broad-bandwidth source that is nearly acoustically transparent," J. Acoust. Soc. Am. **97**, 2699–2708 (1995).
[30]L. R. Dragonette, S. K. Numrich, and L. J. Frank, "Calibration technique for acoustic scattering measurements," J. Acoust. Soc. Am. **69**, 1186–1189 (1981).
[31]K. G. Foote and D. N. MacLennan, "Comparison of copper and tungsten carbide calibration spheres," J. Acoust. Soc. Am. **75**, 612–616 (1984).

Investigation of microphones as near-ground sensors for seismic detection of buried landmines

Gregg D. Larson and James S. Martin
Woodruff School of Mechanical Engineering, Georgia Institute of Technology, Atlanta, Georgia 30332

Waymond R. Scott, Jr.
School of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, Georgia 30332

(Received 25 May 2006; revised 12 April 2007; accepted 29 April 2007)

Commercially available microphones were investigated as near-ground sensors to measure the acoustic pressure and the vertical pressure gradient of evanescent air-acoustic waves associated with audio-frequency seismic waves. Measurements in close proximity to the surface and the use of waveguides were found to improve the microphone signal's quality, the comparison of its seismic sensitivity to its sensitivity to propagating sound (ambient acoustic noise and nonseismic reverberation). Landmine images formed using microphone data collected in a laboratory experimental model clearly locate buried inert landmines but exhibit more clutter than images of the same objects formed with seismic displacement data collected using other techniques.
 [DOI: 10.1121/1.2743155]

PACS number(s): 43.40.Ph, 43.40.At, 43.35.Pt, 43.38.Rh [RAS] Pages: 253–258

I. INTRODUCTION

Experimental seismic landmine detection techniques use noncontact[1–5] and ground-contacting[6] sensors to measure surface-normal displacements due to the propagation of seismic waves and their interactions with shallow-buried objects. In previous research, landmines and minelike objects were found to be characterized by resonant scattering of incident seismic waves in the low audio-frequency range (50 to 2000 Hz).[7–9] In this study, microphones were investigated as an attractive low-cost alternative to noncontact sensors such as radar, ultrasound, and laser-Doppler vibrometer systems[1–5] for seismic landmine detection.

Cook and Wormser[10] first reported the use of microphones for landmine detection in 1973. Their system used a reciprocal speaker as a microphone to sense seismic waves produced by a ground-contacting source.

For seismic landmine detection, the use of a microphone in lieu of a vibrometer dictates the use of a ground-coupled seismic source, rather than an airborne source. This is because propagating acoustic-field components directly generated by the source can easily overwhelm the measurement of the acoustic field generated by the ground's surface motion. Thus, sources should be chosen that maximize the ratio of energy in the source-generated seismic wave field to the energy in the source-generated propagating acoustic field.

The predominant surface waves generated by ground-coupled sources for seismic landmine detection systems may be referred to as Rayleigh waves. These are the lowest order *P-SV* (coupled pressure and vertical shear wave) mode that is guided by the interface between a vertically heterogeneous medium and the overlying air. In soil, over the frequency range of interest, these waves typically propagate at speeds between 70 and 160 m/s.[11,12] For the measurements reported here, Rayleigh wave speeds were measured at approximately 80 m/s. Because they are subsonic, soil-borne Rayleigh waves produce evanescent acoustic fields in the air above the soil. Below 2 kHz, these signals have been measured with microphones up to 10 cm above the air-soil interface.[13,14]

The evanescent field decays in amplitude with increasing distance from the ground's surface. Since the higher frequency (shorter wavelength) waves decay more quickly, it is desirable to measure the acoustic pressure as close to the surface as possible in order to preserve the full bandwidth of the seismic wave field in the measured data. The obvious limit to this is when the microphone physically contacts the ground. Here, the microphone can still function as a near-ground rather than a ground-contacting sensor if only the case of the microphone touches the ground and the sensing element is separated from the surface by air. Recent work with ground-contacting vibrometers has demonstrated that ground contact is not inherently problematic for a landmine detection system, provided that the contact force is sufficiently small.[6] In the case of the microphones used in these experiments, which weigh only 1.5 N as freestanding sensors, ground contact is not problematic.

II. EXPERIMENTAL SETUP

To investigate the capabilities of microphones for seismic detection of buried landmines, measurements were conducted in the experimental model shown in Fig. 1. Seismic surface waves were generated with an electrodynamic shaker coupled to the ground with a 1.27 cm wide (normal to a wavefront) by 21.6 cm long (along a wavefront) aluminum foot to preferentially excite Rayleigh waves.[15] Two Behringer ECM8000 microphones, used for all of the measurements, are shown in Fig. 1 on the positioning system that was used to scan the microphones through the region of interest; an inset photograph shows the microphone with and without the flexible (translucent latex rubber) and rigid (red plastic) waveguides used in the experiments. Damp, compacted sand was used in the experimental model, because its properties are typical of soils and it allows for easy burial

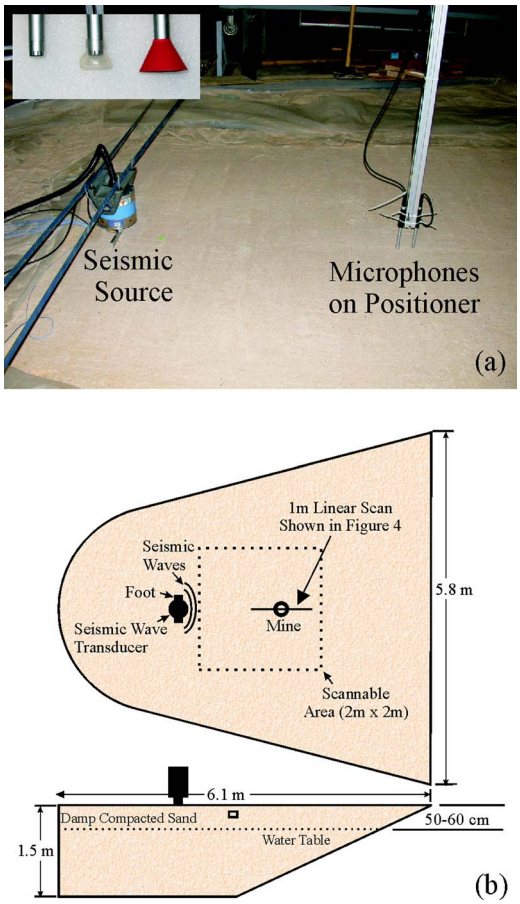


FIG. 1. Experimental model in (a) photograph with seismic source and microphone sensors (inset photograph show waveguides and microphones with 5-cm center-to-center spacing between microphones) and (b) drawing of scan region.

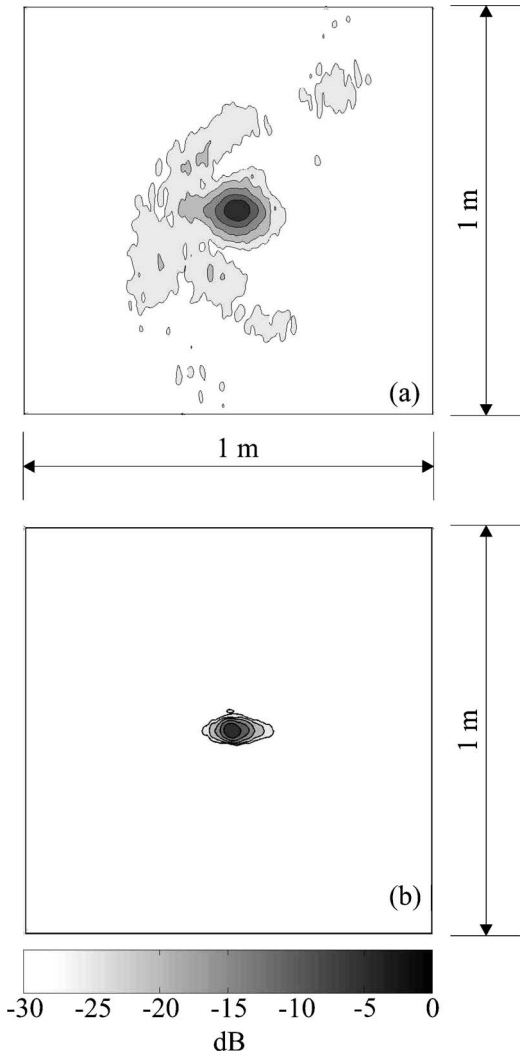


FIG. 2. Images of a TS-50 antipersonnel landmine (1 cm deep) made from data measured by (a) a microphone as a near-ground sensor 1 cm above the surface and (b) a noncontact radar vibrometer. Both images are plotted on a 30-dB scale relative to the maximum value in the image and show a 1-m^2 region in the experimental model.

and compaction to maintain lateral homogeneity and repeatability of measurements.[12,16–20] For all of the experiments, incident seismic signals were excited by chirps spanning a band from 100 Hz to 2 kHz in 4.096 s. These measurements were used to compute the response of the system to a differentiated Gaussian pulse with a 450 Hz center frequency.

III. LANDMINE DETECTIONS

Images made from pressure data measured with a microphone and surface displacement data measured with a radar-based vibrometer over a buried landmine are shown in Fig. 2 for a 1 m^2 region. Measurements were made at 1 cm intervals in both directions for the radar sensor and at 0.78 cm intervals for the microphone sensor. However, the microphone has a larger effective sensing area than the radar sensor. Both images show clear indications of the landmine's location, although there is less clutter and higher resolution in the image formed from the radar data. Similarly, microphone measurements over a variety of other landmine burials have yielded images with more clutter and less resolution than those from the radar data.[13] These images were formed by filtering out the forward-propagating waves, determining the reverse-propagating energy at each location and the reflected energy from each location, and then combining the two energy levels as a weighted image.[21]

In previous work, the measured acoustic signals were back-projected to the surface of the sand using near-field acoustical holography.[13,14] This, at best, minimally improved the resolution and clutter rejection.

The increased clutter in the microphone image is due to the microphones' comparatively high intrinsic noise, sensitivity to ambient acoustic noise in the laboratory, and sensitivity to the reverberant air-acoustic waves produced by the electrodynamic shaker. Both microphone and direct surface motion measurements are also susceptible to ambient seismic noise, which can be problematic in the experimental model.[22] The crescent-shaped artifact between the landmine and the source is caused by the multiple wave types that are present in the model. The time window used on the data includes some faster propagating components of the seismic wave field (prograde surface waves[6,23–26] and bulk waves). Because only the reverse-propagating components are used to form the image, these faster waves result in image clutter between the source and the buried objects. The microphone is proportionally more sensitive to faster waves, which decay at a slower rate away from the surface.

IV. SENSITIVITY IMPROVEMENTS

To improve the microphone's sensitivity to Rayleigh waves relative to propagating acoustic waves, two waveguides were tested. Measurements were made with the microphone in five configurations: (1) the microphone on the surface; (2) the microphone approximately 0.5 cm above the surface; (3) the microphone with the rigid waveguide on the surface; (4) the microphone with the flexible waveguide on the surface; and (5) the microphone with the flexible waveguide above the surface. The microphone has a 4.8 mm diameter opening for the electret sensing element surrounded by a 1.27 cm rigid case. The rigid waveguide was made of hard plastic and tapered linearly from the 1.27 cm diameter of the microphone case to a 3.8 cm diameter opening over a 2.54 cm length. The tip of the microphone extended approximately 1.27 cm into the rigid waveguide. The flexible waveguide, made of latex rubber, fit snugly over the end of the microphone case and expanded to a 2.54 cm diameter at the open end over a length of 1.27 cm. The tip of the microphone extended approximately 0.3 cm into the flexible waveguide. The microphone with the flexible waveguide was coupled to the ground by pushing the sensor into contact with the surface. This ensured that the waveguide was in uniform contact with the ground, but caused some deformation of its shape. The microphone with the rigid waveguide was simply placed upon the surface. Other waveguides including long pipes and flexible tubes were tested as methods to increase sensor standoff distance. These were not found to improve the surface motion sensitivity as well as the waveguides previously described. For comparison purposes, a ground-contacting accelerometer[6] was used to measure the surface motions. After burial of a VS-1.6 antitank (AT) landmine 5 cm deep in the middle of the scan region and careful recompaction of the experimental model, each sensor was scanned in rapid succession to minimize changes in the soil properties (particularly due to drying of the near-surface layers) during the experiment.

In Figs. 3(a)–3(e), 25 ms of data are plotted for each sensor configuration at a location 121 cm away from the source and 50 cm in front of the mine. All of the microphone data were normalized to the maximum microphone signal; the ground-contacting accelerometer signal was normalized by its own peak value. Three distinct waves can be seen in the data presented in Figs. 3(a)–3(e). The first arrival, the direct in-air acoustic wave, is only apparent with the microphone. Local excitation of the ground by this signal is negligible in comparison to sand-borne waves. The direct acoustic signal has fairly constant amplitudes and arrival times for all of the microphone configurations. The second and third arrivals, the prograde surface wave and the Rayleigh surface wave, respectively, are apparent in all of the data sets. Arrival times of the two different waves for the different sensor configurations change somewhat due to slight variations in the sensing area and measurement locations.

The goal of adding a waveguide to the microphone was to increase the amplitude of the Rayleigh surface wave relative to the response due to the in-air acoustic waves. This is most readily apparent through comparison of the measured Rayleigh wave and the direct in-air acoustic signal. The direct in-air acoustic signal is not a significant practical concern because it can be removed by time gating; however, its strength is directly related to the sensitivity of the microphone to reverberant in-air acoustic waves, which cannot be removed by time gating because of their longer path lengths and their possible coincident arrival with the surface waves. In Fig. 3(f), the root-mean-square (rms) amplitudes of the measurement from three time windows (2.5 to 6.5 ms for the direct in-air acoustic wave, 6.5 to 10 ms for the prograde surface wave, and 10 to 25 ms for the Rayleigh surface wave) have been plotted for each sensor configuration. The rigid waveguide improved the relative sensitivity of the microphone to the seismic surface waves more than the other configurations, but the flexible waveguide also demonstrated increased sensitivity to the surface waves. The flexible and rigid waveguide measurements showed nearly the same rms level for the prograde wave, but the rigid waveguide measurement showed about twice the rms level for the Rayleigh wave as that measured with the flexible waveguide. The height dependence of the evanescent acoustic waves in air generated by the prograde and Rayleigh surface waves can be observed by comparison of the measured response of the ground-contacting and in-air microphones. The decay of the evanescent wave is more noticeable for the Rayleigh wave signal because of its shorter wavelength due to its slower propagation speed.

Waterfall plots of microphone data from this experiment are shown in Fig. 4 for (a) the microphone in air; (b) the microphone on the ground; and (c) the microphone with the rigid waveguide on the ground. Data are plotted as a function of time on the horizontal axis and as a function of measurement location (from 50 cm in front of to 50 cm past the center of the burial location at 1-cm intervals) along the vertical axis. The seismic source was located 71 cm in front of the first measurement location. For comparison purposes, the data presented have been normalized such that the maximum amplitude of the measurement at the first location is the same for all sensor configurations. The location of the VS-1.6 AT landmine is indicated by the shaded region in the center of the plots. The prograde and Rayleigh surface waves can be seen to propagate through the scan region, exciting a resonance of the landmine-soil system, and scattering surface waves. The direct in-air acoustic wave generated by the seismic source is apparent in all of the measurements. The results for the rigid waveguide clearly show its stronger sensitivity to the Rayleigh wave relative to the prograde and in-air waves. The front and back edges of the landmine reflect the incident waves. The resonance effects can be seen to persist over the landmine after the passage of the Rayleigh wave. Acoustic and seismic reverberation and noise contribute to the measured pressure following the passage of the Rayleigh wave. Some of the observed effects in the microphone measurements may be due to seismically induced motion of the microphone's case, but this was not investigated.

Vertical gradients are comparatively large in the evanescent field and virtually nonexistent in the direct air-acoustic field where the acoustic velocity is nearly parallel to the surface of the sand. This suggests that buried landmines could

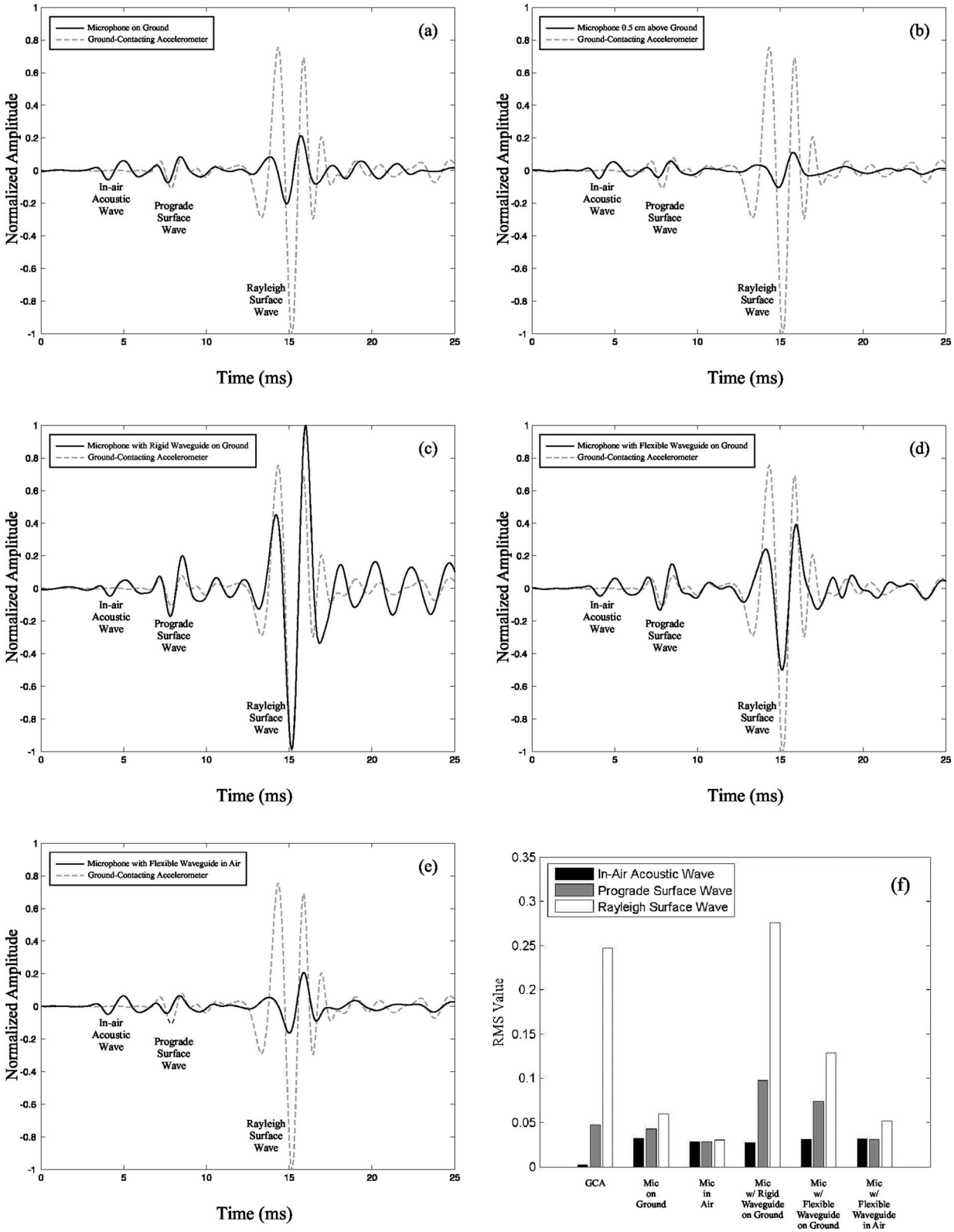


FIG. 3. Experimental measurements of normalized pressure as a function of time (a)–(e) and root-mean-square (rms) amplitude (f) with (a) microphone on the ground; (b) microphone in air approximately 0.5 cm above the ground; (c) microphone with rigid waveguide on ground; (d) microphone with flexible waveguide on ground; and (e) microphone with flexible waveguide in air. The normalized surface acceleration measured by a ground-contacting accelerometer is shown as a reference in all plots.

be detected via a pressure-gradient measurement more effectively than with a simple pressure measurement. By subtracting the measured in-air response of the microphone [shown in Fig. 4(a)] from the response of the microphone with the rigid waveguide on the ground [shown in Fig. 4(c)], the approximate vertical pressure gradient can be synthesized from the data and plotted as a waterfall in Fig. 4(d). This technique reduced the effects from both the direct air-acoustic

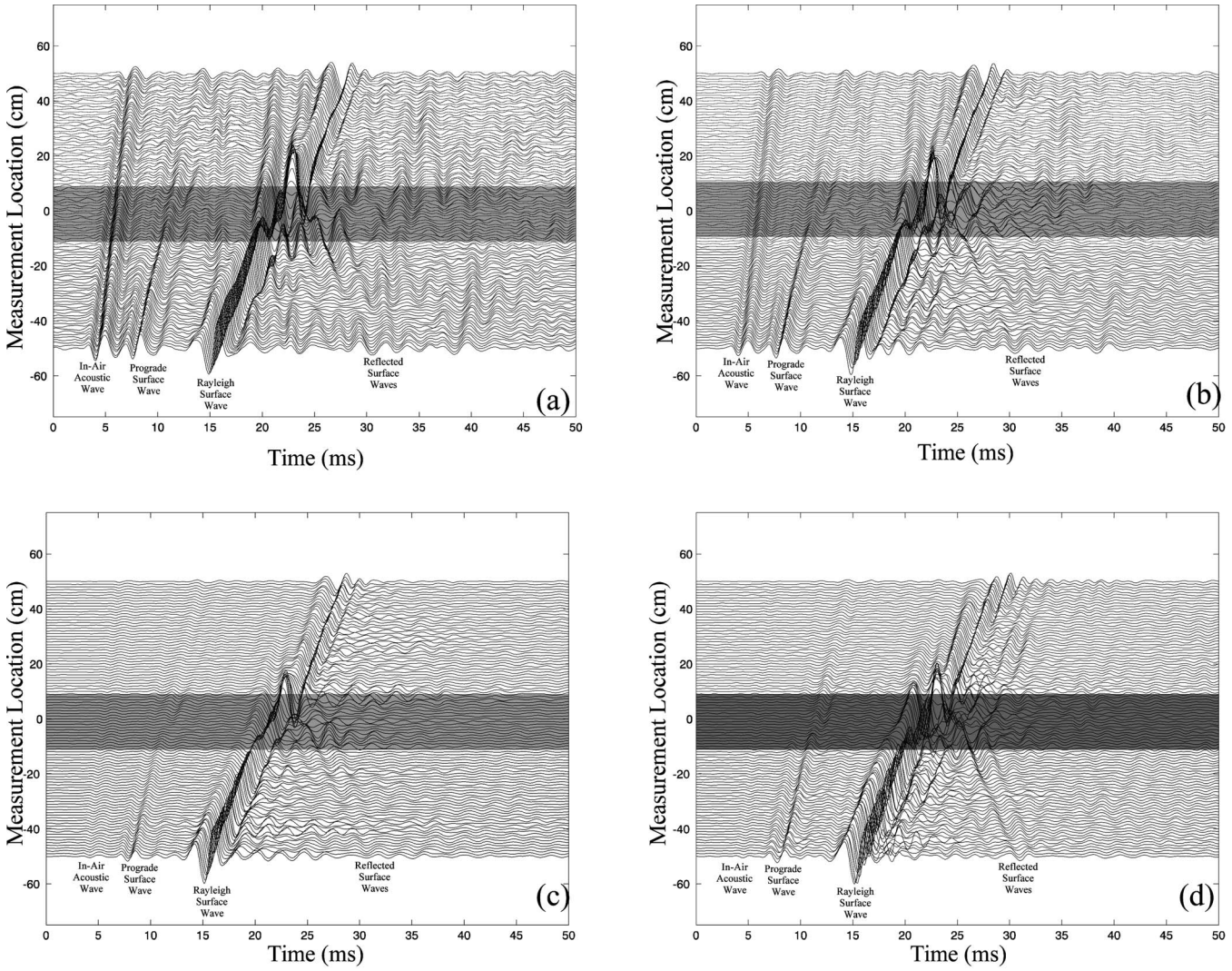


FIG. 4. Waterfall plots of surface motion measured by (a) the microphone 0.5 cm above the surface in the air; (b) the microphone on the ground; (c) the microphone with the rigid waveguide on the ground; and (d) the pressure difference between (c) the microphone with the rigid waveguide on the ground and (a) the microphone 0.5 cm above the surface in the air. A VS-1.6 AT landmine was buried 5 cm deep in the middle of the 1-m-long scan (location indicated by shaded region).

wave and the prograde surface wave relative to the Rayleigh wave. The effects of the prograde wave are diminished in the approximated gradient measurements because the prograde wave propagates faster than the Rayleigh wave in the experimental model.[11] The subsonic prograde wave (at approximately 180 m/s) produces an evanescent acoustic field in the air that decays more slowly with distance from the surface than the signal associated with the Rayleigh wave. Prograde waves have been measured as fast as 460 m/s at field test sites.[11] These would generate propagating air-acoustic signals with predominantly horizontal velocities that could be discriminated against by a measurement of the vertical pressure gradient.

One limitation of approximating the vertical gradient from this data is that the measurements were not conducted simultaneously for each location. Thus, fluctuations in the ambient noise between the different measurements can elevate the apparent noise floor in comparison to what would have been measured with a gradient microphone. However, these measurements were made with the same microphone, which eliminates the mutual scattering and interaction effects that would be encountered with a gradient microphone and eliminates the problem of matching microphones' sensitivities. Subtraction of simultaneously recorded measurements from two different microphones with similar locations above the surface resulted in more apparent noise than the reported synthetic measurements.

V. CONCLUSIONS

As near-ground sensors for seismic landmine detection, commercially available microphones can provide sufficient seismic sensitivity to detect and localize buried antipersonnel and antitank landmines. Images formed from microphone measurements clearly indicate the location and extent of buried landmines, albeit with more clutter than images generated with other vibration-sensing methods. Minimal distance between the microphone and the soil is desirable for optimal signal quality. The main drawback of increased standoff is decreased sensitivity to surface motion relative to the effects

of nonseismic noise and reverberation. The sensitivity of microphones to the Rayleigh surface wave, relative to in-air acoustic waves generated by the seismic source, was improved through the use of ground-coupling waveguides. A rigid waveguide in direct contact with the ground provided the most improvement among the configurations that were investigated. Signal quality was improved in measurements of the vertical pressure gradient above the surface, which also reduced the relative effects of prograde seismic waves.

ACKNOWLEDGMENTS

This work was supported by the Office of Naval Research under Contract Number N00014-04-1-0613 and by the U. S. Army RDECOM CERDEC Night Vision and Electronic Sensors Directorate under Contract Number DAAB07-03-D-C213-0006.

[1]W. R. Scott, Jr., J. Martin, and G. Larson, "Experimental model for a seismic landmine detection system," IEEE Trans. Geosci. Remote Sens. **39**(6), 1155–1164 (2001).
[2]J. S. Martin, D. J. Fenneman, F. Codron, P. H. Rogers, W. R. Scott Jr., G. D. Larson, and G. S. McCall, II, "Ultrasonic displacement sensor for the seismic detection of buried land mines," Proc. SPIE **4742**(2), 606–616 (2002).
[3]A. G. Petculescu and J. M. Sabatier, "Air-coupled ultrasonic sensing of grass-covered vibrating surfaces; qualitative comparisons with laser Doppler vibrometry," J. Acoust. Soc. Am. **115**(4), 1557–1564 (2004).
[4]N. Xiang and J. M. Sabatier, "An experimental study on antipersonnel landmine detection using acoustic-to-seismic coupling," J. Acoust. Soc. Am. **113**(3), 1333–1341 (2003).
[5]D. Donskoy, A. Ekimov, N. Sedunov, and M. Tsionskiy, "Nonlinear seismo-acoustic land mine detection and discrimination," J. Acoust. Soc. Am. **111**(6), 2705–2714 (2002).
[6]J. S. Martin, G. D. Larson, and W. R. Scott Jr., "An investigation of surface-contacting sensors for the seismic detection of buried landmines," J. Acoust. Soc. Am. **120**(5), 2676–2685 (2006).
[7]C. T. Schröder and W. R. Scott Jr., "Resonance behavior of buried landmines," Proc. SPIE **4394**(1), 553–562 (2001).
[8]C. T. Schröder, "On the Interaction of Elastic Waves with Buried Land Mines: An Investigation Using the Finite-Difference Time-Domain Method" (Ph.D. dissertation, School of Elec. and Comp. Engineering, Georgia Institute of Technology, Atlanta, GA, 2001).
[9]A. Zagrai, D. Donskoy, and A. Ekimov, "Structural vibrations of buried land mines," J. Acoust. Soc. Am. **118**(6), 3619–3628 (2005).
[10]J. C. Cook and J. J. Wormser, "Semi-remote acoustic, electric and thermal sensing of small buried nonmetallic objects," IEEE Trans. Geosci. Electron., **GE-11**(3), 135–152 (1973).
[11]G. D. Larson, M. Alam, J. S. Martin, W. R. Scott, Jr., J. H. McClellan, G. S. McCall, II, P. D. Norville, and B. Declety, "Surface-wave-based inversions of shallow seismic structure," Proc. SPIE **5089**(2), 1231–1242 (2003).
[12]M. Westebbe, J. F. Bohme, H. Krummel, and M. B. Matthews, "Model fitting and testing in near surface seismics using maximum likelihood in frequency domain," in Conference Record 32nd Asilomar Conf. Signals, Systems, and Computers, Pacific Grove, CA, **2**, 1311–1315 (1998).
[13]G. D. Larson, J. S. Martin, and W. R. Scott, Jr, "Detection of buried landmines using seismic waves and microphones," Proc. SPIE **5794**(1), 655–664 (2005).
[14]G. D. Larson, J. S. Martin, W. R. Scott, Jr, and C. Jia, "Air acoustic sensing of seismic waves," J. Acoust. Soc. Am. **107**(5), 2896–2897 (2000).
[15]P. Ferrari, "Influence of the shape of an exciting foot on the propagation of elastic waves in the ground" (MSME thesis, Woodruff School of Mechanical Engineering, Georgia Institute of Technology, Atlanta, GA, 2003).
[16]E. Smith, P. S. Wilson, F. W. Bacon, J. F. Manning, J. A. Behrens, and T. G. Muir, "Measurement and localization of interface wave reflections from a buried target," J. Acoust. Soc. Am. **103**(5), 2333–2343 (1998).
[17]D. A. Sachs, "Feasibility of Acoustic Landmine Detection: Final Technical Report," BBN Technical Report No. 7677, Submitted to U.S. Army Belvoir Research Development and Engineering Center, 19 May 1992.
[18]R. Bachrach and A. Nur, "High-resolution shallow-seismic experiments in sand. I. Water table, fluid flow, and saturation," Geophysics **63**(4), 1225–1233 (1998).
[19]R. Bachrach, J. Dvorkin, and A. Nur, "High-resolution shallow-seismic experiments in sand. II. Velocities in shallow unconsolidated sand," Geophysics **63**(4), 1234–1240 (1998).
[20]J. M. Sabatier, H. E. Bass, L. N. Bolen, and K. Attenborough, "Acoustically induced seismic waves," J. Acoust. Soc. Am. **80**(2), 646–649 (1986).
[21]A. Behboodian, W. R. Scott, Jr., and J. H. McClellan, "Signal processing of elastic surface waves for localizing buried land mines," in *Proceedings of the 33rd Assilomar Conference on Signals, Systems, and Computers*, Pacific Grove, CA, **2** pt. 2, pp. 827–830 (1999).
[22]J. S. Martin, G. D. Larson, W. R. Scott, Jr., and G. S. McCall, II, "Evaluation of seismic noise for landmine detection system development," Proc. SPIE **5089**(1), 653–664 (2003).
[23]C. T. Schröder and W. R. Scott, Jr, "On the complex conjugate roots of the Rayleigh equation: The leaky surface wave," J. Acoust. Soc. Am. **110**, 2867–2877 (2001).
[24]F. Gilbert and S. J. Laster, "Excitation and propagation of pulses on an interface," Bull. Seismol. Soc. Am. **52**, 299–319 (1962).
[25]J. A. Hudson, *The Excitation and Propagation of Elastic Waves* (Cambridge University Press, New York, 1980), pp. 171–172.
[26]L. Tsang, "Time-harmonic solution of the elastic head wave problem incorporating the influence of Rayleigh poles," J. Acoust. Soc. Am. **63**, 1302–1308 (1978).

Numerical and experimental validation of a hybrid finite element-statistical energy analysis method

Vincent Cotoni[a] and Phil Shorter
ESI US R&D, 12555 High Bluff Drive, Suite 250, San Diego, California

Robin Langley
Department of Engineering, University of Cambridge, Trumpington Street, Cambridge, CB2 1PZ, United Kingdom

(Received 22 August 2006; revised 12 April 2007; accepted 21 April 2007)

The finite element (FE) and statistical energy analysis (SEA) methods have, respectively, high and low frequency limitations and there is therefore a broad class of "mid-frequency" vibro-acoustic problems that are not suited to either FE or SEA. A hybrid method combining FE and SEA was recently presented for predicting the steady-state response of vibro-acoustic systems with uncertain properties. The subsystems with long wavelength behavior are modeled deterministically with FE, while the subsystems with short wavelength behavior are modeled statistically with SEA. The method yields the ensemble average response of the system where the uncertainty is confined in the SEA subsystems. This paper briefly summarizes the theory behind the method and presents a number of detailed numerical and experimental validation examples for structure-borne noise transmission. © *2007 Acoustical Society of America.* [DOI: 10.1121/1.2739420]

PACS number(s): 43.40.Qi, 43.40.Dx [LPF] Pages: 259–270

I. INTRODUCTION

The dynamic analysis of a complex vibro-acoustic system is difficult for two reasons: first the system may require many degrees of freedom to describe the response, and second the response may be sensitive to small imperfections in the system (arising, for example, from a given manufacturing process). Both of these problems become increasingly severe as the excitation frequency increases, due to the decreasing wavelength of the deformation. This situation is commonly encountered in the automotive industry, for example, where finite element (FE) models with several million degrees of freedom are often required to model the response of a vehicle to a few hundred hertz, while successive vehicles from a production line can show significant differences in their measured responses.[1,2]

When modeling the response of a vibro-acoustic system at higher frequencies it is therefore desirable to adopt an analysis method that provides statistical information and requires relatively few degrees of freedom. Statistical energy analysis (SEA)[3] provides such a method as a system is modeled as an assembly of relatively few subsystems, and the method gives the "mean" vibrational energy level in each subsystem (averaged across an ensemble of nominally identical systems). The approach is also capable of predicting the ensemble variance of the energy levels,[3,4] giving a measure of the statistical spread in the response. One of the assumptions in SEA is that the dynamic properties of the subsystems are very uncertain. If a subsystem is large compared with a deformation wavelength, then this assumption is typically valid and the "ensemble-average" response given by an SEA model provides a useful way to characterize the response of a system. However, at mid and low frequencies, this assumption may not be valid for all subsystems and the resulting ensemble average response becomes less useful.

This mix of dynamic behavior in which some subsystems are large compared with a wavelength (and are therefore well modeled by SEA), while others are short compared with a wavelength (and are not well modeled by SEA) is commonly encountered at mid frequencies. For example, consider the structure shown in Fig. 1, in which a number of thin panels are bolted to a stiff framework. Above 50 Hz, the wavelength of free flexural waves in the panels is much shorter than the panel dimensions; the bending modes of the panels are then typically very sensitive to perturbations. Conversely, the framework is very stiff and consequently exhibits few modes, which are insensitive to perturbation over a wide frequency range. In order to investigate the sensitivity of the response of the built-up structure, the panels were perturbed by adding a number of small masses at random locations, and a numerical Monte Carlo analysis was performed for different realizations of the mass positions. The structure was excited by a point force on the framework and the modulus squared velocity at a point on a panel was computed for each realization, as shown in Fig. 1. The results for 200 realizations are plotted in gray and one particular realization is plotted with a dotted line. The bold curve is the ensemble-averaged response. The response varies significantly from one realization to another, and the prediction of the response of one given configuration does not provide meaningful information regarding the behavior of the random ensemble. It is of more interest to consider the statistical properties of the response, such as the ensemble mean. The mean response exhibits distinct peaks that can be traced to the robust dynamic properties of the framework. Typically, a

[a]Author to whom correspondence should be addressed. Electronic mail: vincent.cotoni@esi-group-na.com

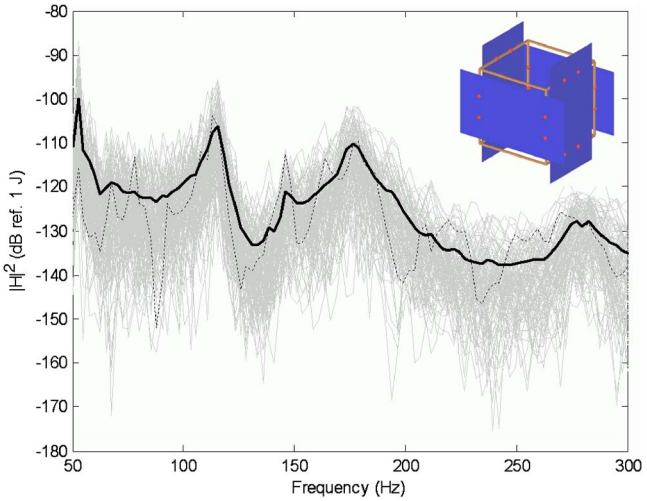


FIG. 1. (Color online) Typical modulus squared displacement response of an ensemble of slightly perturbed framework-panel structures at mid frequencies. Gray: a set of 200 Monte Carlo realizations; dotted: one particular Monte Carlo realization; thick black: ensemble average.

SEA model of the framework and panel would fail to describe this trend (since a SEA model implicitly assumes that the dynamic properties of the framework are very uncertain), while a FE model would be computationally expensive to apply to even a single realization of a more realistic structure. This example highlights the need for a hybrid modeling approach for describing the response of a system in the mid-frequency range.

The literature on mid-frequency analysis is extensive, and several authors have discussed attempts to combine deterministic and statistical analysis methods into a single model. The objective of a hybrid FE-SEA (or more generally deterministic-statistical) method is to combine the low frequency performance of the FE method with the high frequency performance of SEA to produce a robust technique that can be applied across a wide frequency range. However, the coupling of FE and SEA in a single model is difficult because the methods differ in two ways: (i) FE is based on dynamic equilibrium while SEA is based on the conservation of energy flow, and (ii) FE is a deterministic method while SEA is inherently statistical.

Belyaev and Palmov[5] included locally averaged statistical behavior into a global deterministic model of a system. This type of approach was developed further by Soize[6] in "fuzzy structure" theory. Langley and Bremner[7] developed a hybrid FE-SEA method based on these ideas combined with a wavelength partitioning scheme. Jayachandran and Bonilha[8] developed a hybrid technique where the structural vibration response obtained from SEA is used to excite a modal model of an interior acoustic space. Grice and Pinnington[9], Ji, Mace, and Pennington,[10] and Hong, Wang, and Vlahopoulos[11] also each developed hybrid methods in a series of papers (only the most recent references are given here) in which the short wavelength components are described statistically by effective impedances applied to the long wavelength components. The approaches differ in the way the effective impedance is computed and in the way the response of the short wavelength components is recovered. Most of those approaches assume that a statistical subsystem acts as an energy sink and therefore do not explicitly account for the flow of energy back into a deterministic subsystem. While such an assumption is often adequate for heavily damped subsystems (and/or weakly coupled subsystems), the assumption is not applicable for more general lightly damped reverberant systems.

Recently Shorter and Langley[12] developed a general method for coupling FE and SEA based on wave concepts rather than the modal-type of approach employed in Ref. 7. At the heart of the method is a reciprocity result[13,14] regarding the forces exerted at the boundaries of a SEA subsystem. The method yields the average response over an ensemble of random systems where the randomness is confined in the components described with SEA (recent work has also focused on the prediction of the variance of the response in the hybrid method, allowing prediction of the spread of the response around the mean value[15]).

The main aim of the present paper is to assess the performance of the hybrid FE-SEA method for predicting the ensemble average response of a number of example structures. The hybrid FE-SEA method is summarized in the following section, and numerical and experimental validation studies are reported in Sec. III.

II. HYBRID FE-SEA METHOD

A. Concept of "direct field" dynamic stiffness matrix

As discussed in the previous section, in the mid-frequency range some components of a complex structure (for example, thin panels) typically display short wavelength response and are sensitive to the effects of random uncertainties, while others (for example, stiff beams) show little variation in their dynamic properties and are essentially deterministic. In the hybrid method proposed by Shorter and Langley,[12] the deterministic components are modeled using the finite element method, while the random components are modeled as SEA subsystems. A key feature of the method is the concept of a direct field dynamic stiffness matrix associated with each SEA subsystem

Consider, for example, a thin plate that is excited at the boundaries. The excitation generates waves that propagate through the plate and are reflected repeatedly at the boundaries; the total dynamic stiffness matrix of the plate, phrased in terms of the edge degrees of freedom, has contributions from all of these reflections. Suppose now that the response is viewed in two parts: (i) the contribution from the initial generated waves, prior to any boundary reflections—this can be called the direct field; (ii) the contribution from waves produced on the first and all subsequent reflections—this can be called the "reverberant field." The direct field dynamic stiffness matrix can be defined as that resulting from the presence of the direct field waves. The presence of uncertainties in the system is accounted for in the reflections of the waves on boundaries and interior scattering devices, with the result that the direct field dynamic stiffness matrix is deterministic across the ensemble. Such a matrix can in principle

be computed from a boundary element analysis. However, in many instances, analytical approaches can be used to find the matrix.

Consider a thin plate bolted to a structural component at a point. The direct field dynamic stiffness of the plate as seen from the bolt can be obtained for any value of the bolt diameter by using recently developed analytical formulas.[16] These formulas are obtained by describing the displacement and stress fields at the bolt-plate interface in terms of a series of outgoing cylindrical waves (composed of Bessel functions), and by solving for the wave amplitudes using displacement compatibility and force equilibrium. In physical coordinates, this dynamic stiffness matrix is a 6×6 matrix relating the three displacements and three rotations at the connecting point to the three forces and three moments. If only the flexural wave field is considered in the description of the panel, the dynamic stiffness matrix reduces to a 3×3 matrix. If N bolts connect the plate with the structural component, then in principle the analysis should account for the coherence of the waves emanating from each bolt. The $6N\times6N$ direct field dynamic stiffness matrix would then be widely populated. However, if the wavelength of a free wave in the plate is short compared to the spacing between the bolts, then the phase of a wave arriving at a bolt is extremely sensitive to perturbation, and the resulting coherence can be neglected. In this case, the direct field dynamic stiffness can be derived from the analysis of each connection in isolation.

Consider now a thin plate connected to structural components, such as beams or frames, along the plate boundaries. If the plate is to be described by SEA and the structural components with FE, then the degrees of freedom of the boundary can be described in terms of the physical nodal degrees of freedom of the FE mesh. If the boundaries are straight, the direct field dynamic stiffness matrix associated with those degrees of freedom can be found by considering each straight boundary in turn, and taking it to form a segment of the edge of a semi-infinite plate. Motion of the boundary will generate waves into the semi-infinite plate, and for a given boundary motion the generated waves can be found by Fourier transform techniques, as detailed in the Appendix . The entries of the dynamic stiffness matrix are then found from the prescribed motion shapes and the boundary forces associated with the generated waves: i.e., strictly the dynamic stiffness matrix of a segment of the edge of a semi-infinite plate, when the motion of the segment is described by FE nodal degrees of freedom. This is the required direct field dynamic stiffness matrix for the plate edge, and by neglecting the coherence between the waves emanating from each plate edge, the process can be repeated for each of the edges to give the total direct field dynamic stiffness matrix of the plate subsystem. Another approach would be to model the vibration of the plate as a combination of incoming cylindrical waves (each carrying energy away from the boundary towards the center of the plate) and then use point collocation to assemble the direct field dynamic stiffness matrix for the boundary degrees of freedom.[17]

Given the concept of the direct field dynamic stiffness matrix, the hybrid FE-SEA equations can now be described.

B. Hybrid FE-SEA equations

The starting point for the hybrid method is to identify those parts of the system response that will be described by SEA subsystems. The remaining part of the system (which can be considered to be the "deterministic" part) is then modeled by using the FE method. The degrees of freedom associated with the SEA subsystems will be omitted from the FE model of the system, at all points other than the subsystems' boundaries. The relevant direct field dynamic stiffness matrix is then added to the FE model at the subsystems' boundaries, and this augmented FE model is used in the subsequent analysis. If the degrees of freedom of the FE part are labeled $\mathbf{q}$, then the governing equations of motion (for harmonic vibration of frequency ω say) will have the form

$$\mathbf{D}_{\mathrm{tot}}\mathbf{q} = \mathbf{f} + \sum_k \mathbf{f}_{\mathrm{rev}}^{(k)}, \tag{1}$$

$$\mathbf{D}_{\mathrm{tot}} = \mathbf{D}_d + \sum_k \mathbf{D}_{\mathrm{dir}}^{(k)}. \tag{2}$$

The summation is over the number of SEA subsystems in the model, and $\mathbf{D}_{\mathrm{dir}}^{(k)}$ represents the direct field dynamic stiffness matrix associated with subsystem k at frequency ω. Furthermore, $\mathbf{D}_d$ is the dynamic stiffness matrix given by the finite element model of the deterministic part of the system (i.e., generally written in terms of a mass, stiffness and damping matrices, and the frequency), $\mathbf{f}$ is the set of external forces applied to this part of the system, and $\mathbf{f}_{\mathrm{rev}}^{(k)}$ represents the force arising from the reverberant field in subsystem k (which is not accounted for in $\mathbf{D}_{\mathrm{dir}}^{(k)}$ only including the direct field effect). The matrix $\mathbf{D}_{\mathrm{tot}}$ is the dynamic stiffness matrix of the FE model (excluding the SEA subsystem degrees of freedom), when augmented by the direct field dynamic stiffness matrix of each SEA subsystem. It should be noted that Eqs. (1) and (2) are exact—all that has been done is to split the forces arising from the SEA subsystems into a direct field part, which is accounted for by $\mathbf{D}_{\mathrm{dir}}^{(k)}$, and a reverberant part which is carried to the right hand side of Eq. (1). The following result[13,14] regarding the cross-spectral matrix of the reverberant blocked force is central to the development of the hybrid method:

$$\mathbf{S}_{ff,\mathrm{rev}}^{(k)} \equiv \mathrm{E}[\mathbf{f}_{\mathrm{rev}}^{(k)}\mathbf{f}_{\mathrm{rev}}^{(k)*\mathrm{T}}] = \left(\frac{4E_k}{\omega\pi n_k}\right)\mathrm{Im}\{\mathbf{D}_{\mathrm{dir}}^{(k)}\}. \tag{3}$$

Here E[] denotes the ensemble average, where the ensemble describes a state of maximum uncertainty in the SEA subsystem with a given modal density, damping and direct field dynamic stiffness at the junction.[13] E_k and n_k are, respectively, the ensemble average energy and the modal density of the kth subsystem. Equation (3) implies that the cross-spectral matrix of the force exerted by the reverberant field on its surrounding boundaries is proportional to the resistive part of the direct field dynamic stiffness matrix of the boundary, and to the energy of the reverberant field. This forms a diffuse field reciprocity statement, and was shown to be valid when the (ensemble average) response of the subsystem constitutes a diffuse random wave field.[14] Equation (3) makes a connection between the energetics of the subsystem and the

elastic forces at the boundary, and as such it forms the key to coupling SEA to FE.

From Eq. (1), the response $\mathbf{q}$ can be expanded in the form

$$\mathbf{q} = \mathbf{q}_d + \sum_k \mathbf{q}^{(k)}, \quad \mathbf{q}_d = \mathbf{D}_{\text{tot}}^{-1}\mathbf{f}, \quad \mathbf{q}^{(k)} = \mathbf{D}_{\text{tot}}^{-1}\mathbf{f}_{\text{rev}}^{(k)}. \tag{4}$$

Now the ensemble and time averaged power input to the direct field of subsystem j can be written as

$$P_{\text{in},j} = (\omega/2)\text{Im}\{\text{E}[\mathbf{q}^{*\text{T}}\mathbf{D}_{\text{dir}}^{(j)}\mathbf{q}]\} = (\omega/2)\sum_{rs} \text{Im}\{D_{\text{dir},\text{rs}}^{(j)}\}S_{qq,rs}, \tag{5}$$

where it has been noted that the direct field dynamic stiffness matrix is symmetric and is fixed across the ensemble. Due to the statistical nature of the reverberant response of the subsystems, the various contributions $\mathbf{q}^{(k)}$ that appear in Eq. (4) are taken to be uncorrelated and of zero mean, and Eqs. (3)–(5) then yield

$$P_{\text{in},j} = P_{\text{in},j}^{\text{ext}} + \sum_k \omega\eta_{jk}n_j(E_k/n_k), \tag{6}$$

where

$$P_{\text{in},j}^{\text{ext}} = (\omega/2)\sum_{rs} \text{Im}\{D_{\text{dir},rs}^{(j)}\}(\mathbf{D}_{\text{tot}}^{-1}\mathbf{S}_{ff}\mathbf{D}_{\text{tot}}^{-1*\text{T}})_{rs}, \tag{7}$$

$$\omega\eta_{jk}n_j = (2/\pi)\sum_{rs} \text{Im}\{D_{\text{dir},rs}^{(j)}\}(\mathbf{D}_{\text{tot}}^{-1}\,\text{Im}\{\mathbf{D}_{\text{dir}}^{(k)}\}\mathbf{D}_{\text{tot}}^{-1*\text{T}})_{rs}. \tag{8}$$

In Eq. (7), $\mathbf{S}_{ff}$ denotes the cross-spectral matrix of the external forces applied to the FE part. Given that the dynamic stiffness matrices are symmetric, it is readily shown from Eq. (8) that reciprocity holds, in the sense that $\eta_{jk}n_j = \eta_{kj}n_k$. As will be shown in what follows, the terms η_{jk} are equivalent to the coupling loss factors that appear in SEA.

The ensemble averaged power output from the reverberant field in subsystem j can be written as the sum of the power dissipated through damping and the power output at the boundary

$$\begin{aligned} P_{\text{out},j} &= \omega\eta_j E_j + (\omega/2)\text{Im}\{\text{E}[\mathbf{q}^{(j)*\text{T}}\mathbf{f}_{\text{rev}}^{(j)}]\} \\ &= \omega\eta_j E_j + \sum_k \omega\eta_{kj}n_k(E_j/n_j) + \omega\eta_{d,j}E_j, \end{aligned} \tag{9}$$

where Eqs. (1)–(4) were used to express the expectation in terms of the subsystem energy and the various dynamic stiffness matrices. The last term in Eq. (9) is the power dissipated in the deterministic part of the system, given by

$$\omega\eta_{d,j} = \left(\frac{2}{\pi n_j}\right)\sum_{rs} \text{Im}\{D_{d,rs}\}(\mathbf{D}_{\text{tot}}^{-1}\,\text{Im}\{\mathbf{D}_{\text{dir}}^{(j)}\}\mathbf{D}_{\text{tot}}^{-1*\text{T}})_{rs}. \tag{10}$$

Equations (6) and (9) then lead to the following energy balance equation for subsystem j:

$$\omega(\eta_j + \eta_{d,j})E_j + \sum_k \omega\eta_{jk}n_j(E_j/n_j - E_k/n_k) = P_{\text{in},j}^{\text{ext}}. \tag{11}$$

Furthermore, the cross-spectral matrix of the response $\mathbf{q}$ can be derived from Eqs. (3) and (4), which yields

$$\mathbf{S}_{qq} = \mathbf{D}_{\text{tot}}^{-1}\left[\mathbf{S}_{ff} + \sum_k \left(\frac{4E_k}{\omega\pi n_k}\right)\text{Im}\{\mathbf{D}_{\text{dir}}^{(k)}\}\right]\mathbf{D}_{\text{tot}}^{-1*\text{T}}. \tag{12}$$

Equations (11) and (12) form the two main equations of the hybrid method. These equations couple FE and SEA methodologies: Eq. (11) has precisely the form of SEA, but the coupling loss factors η_{jk} and loss factors $\eta_{d,j}$ are calculated by using the FE model (augmented by the direct field dynamic stiffness matrices) via Eqs. (8) and (10); furthermore, Eq. (12) has the form of a standard random FE analysis, but additional forces arise from the reverberant energies in the SEA subsystems. If no SEA subsystems are included then the method becomes purely FE; on the other hand, if only the junctions between the SEA subsystems are modeled by FE, then the method becomes purely SEA, with a novel method of computing the coupling loss factors through complex and possibly dissipative junctions.

C. Steps in the Hybrid FE-SEA method

The first step when building a hybrid model of a system is to partition the system into deterministic and statistical portions. This partitioning can be based on the expected wavelengths within the various components of a system. The local dynamic response of a subsystem that is large compared to a wavelength is typically sensitive to perturbations, and an SEA description should be used; conversely, subsystems that are small compared with wavelength should be described with FE. Although the value defining small *versus* large wavelength depends on the actual level of uncertainty or perturbation (i.e., the dynamic properties of any components could eventually be made random by sufficiently increasing the level of the perturbations), the practical limit value of a few wavelengths within the component is typically used. It should be noted that the choice of partitioning implicitly defines the statistical ensemble in a hybrid analysis.

Once the partition is defined, the hybrid method proceeds as follows:

(1) A finite element model of the deterministic part of the system is constructed. All degrees of freedom associated with SEA subsystems are omitted from this model, other than those that lie on the subsystem boundaries.
(2) A direct field dynamic stiffness matrix is constructed for each SEA subsystem in terms of the relevant boundary degrees of freedom. These matrices are then coupled to the FE model to yield the total dynamic stiffness matrix $\mathbf{D}_{\text{tot}}$.
(3) The various terms that appear in the SEA equation, Eq. (11), are calculated from Eqs. (7), (8), and (10).
(4) The SEA equations are solved to yield the subsystem energies E_j.
(5) Given the SEA subsystem energies, Eq. (12) is used to yield the response of the deterministic part of the system.

A key feature of the method is that the FE mesh considered in step (1) does not need to capture the short wavelength response of the subsystems, and thus relatively few degrees of freedom are required compared to a conventional finite element model. This leads to very significant reductions in

the computer time and memory needed to solve the problem. Furthermore, the method immediately yields ensemble average response quantities, without any need for a Monte Carlo simulation of the ensemble.

III. EXAMPLE APPLICATIONS

Three validation studies are presented in this section. The first two are concerned with FE and SEA subsystems which are connected at discrete points, while the third involves extended connections along line junctions. In all cases, benchmark Monte Carlo results for the response of an ensemble of randomized structures have been obtained for comparison with the ensemble-averaged response yielded by the hybrid method. In the second case, the benchmark Monte Carlo results were obtained experimentally; in the other cases, the benchmark results were obtained from extensive detailed FE calculations.

The systems were randomized by attaching point masses at random locations on various components. Ten masses per component were considered, with a constant total added mass of 15% or 20% of the mass of the bare component depending on the particular example. When there is sufficient uncertainty in the dynamic properties of a system, the overall response statistics become insensitive to the way in which a system is randomized.[4,18] The use of point masses in the following examples is therefore representative of more general ensembles.

The predictions of the modulus squared velocity response at a number of locations on the structures are compared. For the hybrid results, if the point was located on the FE part, the modulus square velocity associated with the ith degree of freedom was obtained from Eq. (12) with $\mathrm{E}[|v_i|^2] = \omega^2\{S_{qq}\}_{ii}$. If the point belonged to an SEA subsystem, the response was recovered using the formula[3] $\mathrm{E}[|v_{i\in \mathrm{sub}k}|^2] = 2E_k/M_k$, where M_k is the mass of the subsystem and the energy is computed from Eq. (11).

A. Framework-panel structure: A numerical benchmark

1. Description of the structure

The hybrid method has been applied to the example structure shown in Fig. 1. The structure consists of a beam framework with four thin panels which are each bolted to the framework at four points. All components are made of aluminium with elastic properties $E=71$ GPa, $\nu=0.33$, and $\rho = 2700\ \mathrm{kg/m^3}$. The beams are all 0.7 m long, with a square hollow section of external side width 25.4 mm and wall thickness 3.2 mm. They are connected at right angles and form the edges of a cube. The panel dimensions are 0.6 m $\times$ 1.1 m, with thickness 1 mm. They are connected to the framework at four points with plain cylindrical bolts of 5 mm radius, with an offset of 17 mm to the neutral axis of the beams. Two successive bolts on a common beam are 0.2 m apart. The damping loss factor is 0.05% for the framework, and 2% for the panels.

Two excitation locations were considered: a point force was applied in the transverse direction on panel 1, or on one of the lower horizontal beams of the framework in the vertical direction at 0.25 m from the corner, as indicated by the arrow in the left hand picture of Fig. 2. Regarding the loading of the randomized panel, "rain-on-the-roof" excitation commonly used in SEA would yield the same ensemble mean response as the point force (but the variance of the response would be smaller).[4] The dynamic response was investigated at a point in each panel in the out-of-plane direction, and at two points on the framework. The observation point 1 on the framework is the same as the loading point and the same vertical direction is considered. The second point is on the other side of the framework at 0.25 m from the corner, and the horizontal direction is considered as shown on the left hand side of Fig. 2.

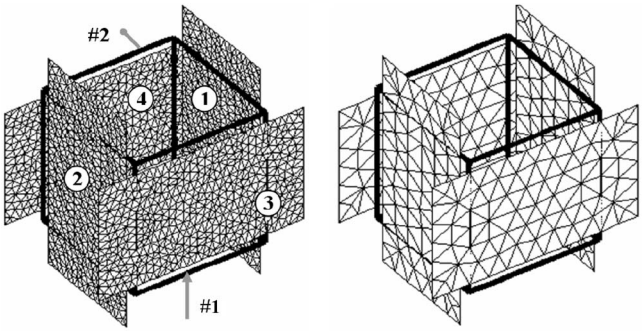


FIG. 2. Two finite element meshes of a framework-panel structure. Left: mesh for reference Monte Carlo simulation (10 481 degrees of freedom); right: mesh for hybrid prediction where the surface elements are membrane only (1184 degrees of freedom).

The reference results were obtained using the detailed FE model of the structure shown on the left hand side of Fig. 2. The FE software package VA ONE[19] was used to create the model composed of 10 481 degrees of freedom for analysis up to 300 Hz. The framework and bolts were described with 96 beam elements and the panels with 3354 triangular shell elements. The structure was randomized by adding a set of ten masses at random locations on each panel, each mass having 2% of the mass of the bare panel. By moving the masses to random locations, 200 realizations were generated for the Monte Carlo analysis.

2. Hybrid FE-SEA model

The partition of the system into deterministic and statistical subsystems can be based on the free propagating wavelengths across the frequency range of interest. The shear and extensional wavelengths in the panels are, respectively, 18 and 16 m at 300 Hz, suggesting that the in-plane motion associated with these waves should be described with FE. The bending wavelength in the panels is 0.31 m at 100 Hz, and 0.18 m at 300 Hz. This corresponds to between 2 and 4 wavelengths within the shortest dimension of the panels, which implies that the out-of-plane motion should be described with SEA. Below 50 Hz, this SEA description is not expected to be appropriate as the bending wavelength is larger than the panel dimensions, and strictly a FE description should be used. The first four propagating waves in the beams are a torsional wave, an extensional wave, and two bending waves (additional higher order wave types cut on

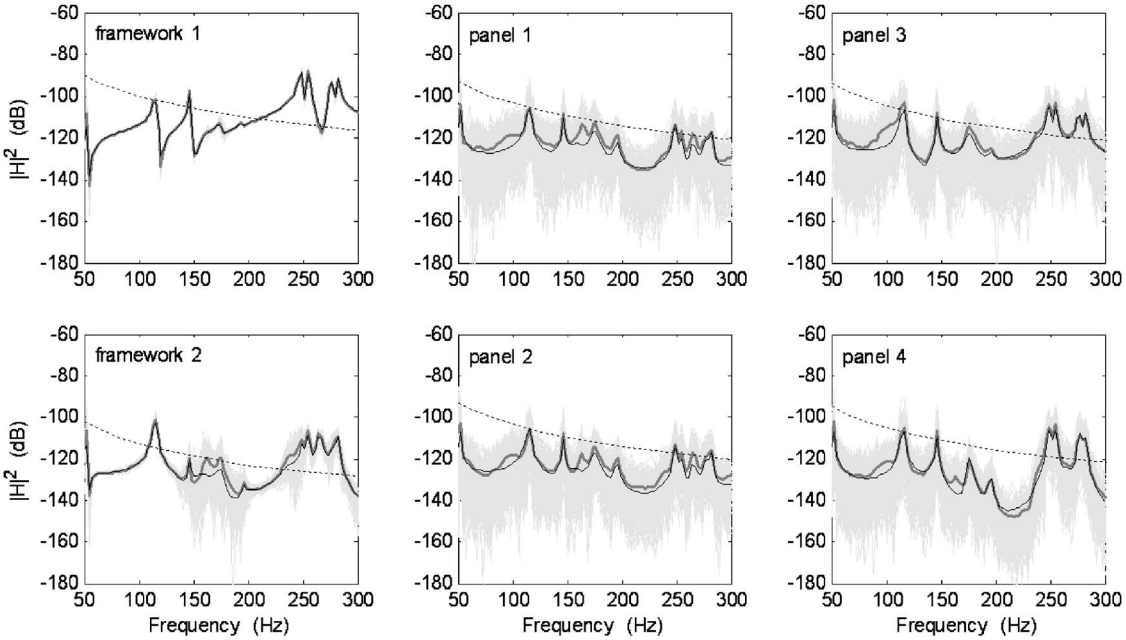


FIG. 3. Modulus squared velocity response of the framework-panel structure to a point force excitation applied on the framework. Light gray: each Monte Carlo realization; bold gray: Monte Carlo ensemble average. Solid black: hybrid FE-SEA; dotted: pure SEA.

above 9650 Hz). The corresponding wavelengths at 300 Hz are all larger than 1 m so that a FE description should be adopted for the beams.

There are approximately 250 bending modes in the four panels, no in-plane modes and approximately 30 modes in the framework. This emphasizes the efficiency gains that can be achieved by using a statistical description of the panel bending motion, rather than a detailed FE description.

The FE model employed in the hybrid method is shown on the right hand side of Fig. 2, and comprises 1184 degrees of freedom. The framework and bolts were described with the same elements and mesh as the reference Monte Carlo FE model. Conversely, the panels were described with a much coarser mesh composed of 520 triangular membrane elements, as only the in-plane motion is of concern. The coupling between the FE and SEA subsystems was described by a series of 16 point junctions. The various point junctions were assumed to be incoherent and the local direct field dynamic stiffness matrices were computed separately.

3. Comparison with Monte Carlo results

The velocity squared frequency response at two points of the framework and at one point of each panel is shown in Figs. 3 and 4 for excitation respectively on the framework, and on panel 1. The response for each Monte Carlo realization is plotted with a light gray line, the ensemble-averaged reference response with a bold gray line, and the hybrid FE-SEA prediction with a black line. In addition to the reference and hybrid results, the results from a pure SEA model (created using the software VA ONE[19]) are shown with dotted lines. In the SEA model, all the components are described with SEA: each of the 12 beams of the framework is described with four wave fields corresponding to the first four propagating waves (one extensional, two flexural, one torsional); each of the four panels is described with three wave fields (one extensional, one flexural, one shear). The complete SEA model is made of 60 SEA subsystems, all coupled through point junctions.

The hybrid prediction provides a good estimate of the ensemble average response, and significantly improves the SEA prediction. In particular, the hybrid approach is able to capture the "modal" behavior of the ensemble mean response arising from the robust dynamics of the framework. The overestimation of the response of the panels in the pure SEA model is due to the strong coupling provided by the long-wavelength subsystems (the framework and the in-plane motion of the plates). This can also be viewed as being due to a lack of a sufficient amount of uncertainty in the modal properties for these subsystems to be represented by SEA subsystems. Some differences between the reference results and the hybrid method are to be expected due to the limited number of trials used in the Monte Carlo simulation, and the lack of randomness introduced by the added masses (especially at lower frequencies).

It is noted in passing that the variance of the response depends on the load case, the observation location and the frequency; recent work has been directed at including variance prediction within the hybrid method.[15]

B. Framework-panel structure: An experimental benchmark

1. Structure and experimental setup

The hybrid method has been used to predict the response of the test structure shown in Fig. 5. The upper part of the structure is similar to the previous numerical example and consists of a cubic beam framework with four thin panels each bolted to the framework at four points. A number of viscoelastic patches are glued onto the thin panels in order to increase the damping, as shown in Fig. 5(b). Four additional circular hollow beams connect the cubic framework to a stiff

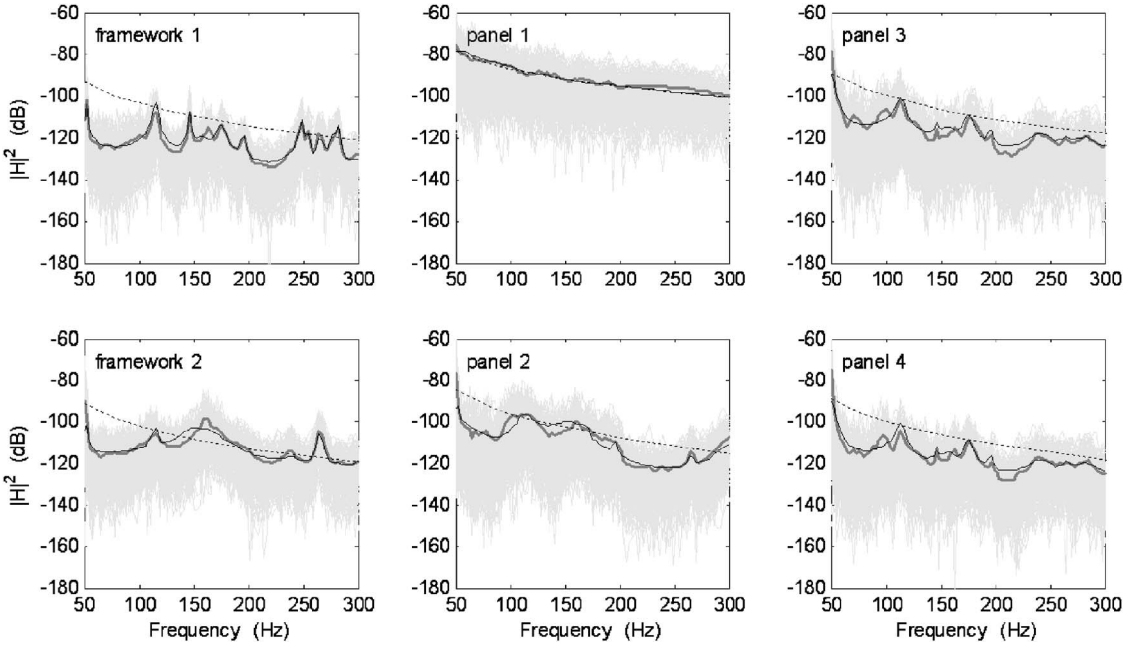


FIG. 4. Modulus squared velocity response of the framework-panel structure to a point force excitation applied on panel 1. Light gray: each Monte Carlo realization; bold gray: Monte Carlo ensemble average. Solid black: hybrid FE-SEA; dotted: pure SEA.

panel via welded junctions (see Fig. 5(a) where the thin panels are removed). The bottom panel is 6.35 mm thick and of dimensions 0.4 m × 0.4 m. The circular hollow beams are 1.3 m long and the section external diameter is 19 mm with wall thickness 1.6 mm. All components are made of aluminium.

The structure was randomized by attaching ten masses to each of the four thin panels at random locations (some masses can be seen in Fig. 5(b)), representing 18% of the mass of the bare panel. Twenty realizations of the structure were tested. The structure was suspended from a corner of the framework with an elastic string so that a close approximation to free boundary conditions was obtained. The framework was excited by an inertial shaker at the location referenced Acc 1 in Fig. 5(a), in the direction of the main axis of the structure (almost vertical in the picture). The shaker is shown in Fig. 5(b), and the precise location was 0.314 m from the closest framework corner. The acceleration response was measured on the framework at the two points Acc 1 and 2 indicated on Fig. 5(a), in the direction of the main axis. The second accelerometer was at 76.7 mm from the center of the bottom plate. The acceleration response was also measured at one point on each of the four panels in the out-of-plane direction. The points were chosen to be remote from the connection points and from the edges to avoid near-field and correlation effects. A white noise signal was applied to the shaker, and acceleration and input force data were measured up to 1000 Hz. The modulus square velocity responses per unit force were computed from the measured acceleration-force cross-spectra S_{af}, and the force auto-spectrum S_{ff}, by using the relation $|v(\omega)|^2 = |S_{af}(\omega)/S_{ff}(\omega)|^2/\omega^2$. The responses were then averaged over all realizations.

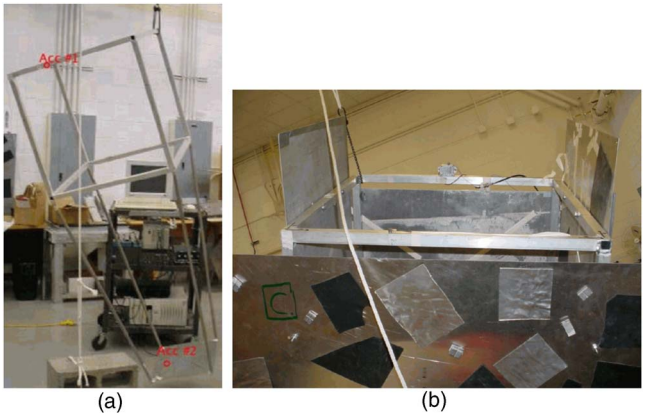


FIG. 5. (Color online) Test structure suspended from a corner. Left: framework and bottom thick plate alone; right: top part of the framework, with the four thin panels each bolted at four points. The inertial shaker used for excitation is on the more remote beam.

2. Hybrid FE-SEA model

The same approach as in the previous example was employed to build the hybrid model of the test structure: For a specified frequency range, the free wavelength of each component was compared to the characteristic dimension of the component in order to choose either an FE or an SEA description. The hybrid model of the structure was developed for analysis up to 1000 Hz. The bending wavelength of the thin panels becomes shorter than the panel dimension at about 30 Hz, and thus an SEA model was adopted. Conversely, the framework, the top plate and the in-plane motion of the panels were modeled by FE. The coupling between the FE model and the four SEA subsystems was achieved by a series of 16 point junctions using the direct field dynamic stiffness matrix from each connection in isolation.

The FE part of the hybrid model excluding the in-plane motion of the thin panels was first validated against experimental data obtained with the structure prior to the attach-

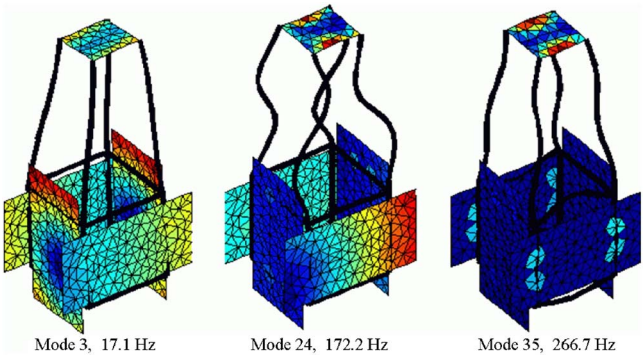


FIG. 6. (Color online) Selected mode shapes of the FE part of the hybrid model of the test structure. Only the membrane behavior of the four large panels is considered (bending is removed).

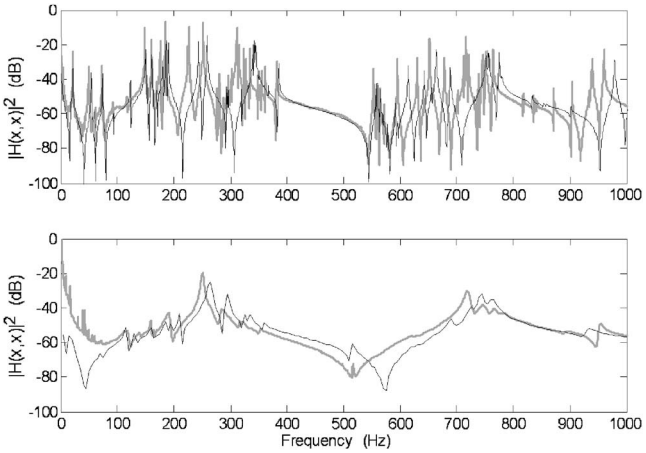


FIG. 7. Experimental (gray) and simulated (black) drive point mobility response of the test structure. Upper diagram: framework without panels (prediction by FE); lower diagram: framework with panels (prediction by hybrid FE-SEA).

ment of the thin panels (see Fig. 5(a)). The model was meshed with triangular shell elements for the top plate and beam elements for the framework, with a total of 876 degrees of freedom for analysis up to 1000 Hz. Starting with standard material properties for the aluminum ($E=71$ GPa, $\nu=0.33$, $\rho=2700$ kg/m^3), the model was updated by adding mass at the welded junctions, then adjusting the aluminum density to fit the mass of the test structure, and finally adjusting the Young modulus to fit the first few natural frequencies. A damping loss factor of 0.5% was found to give satisfactory predictions over the whole frequency range. The predicted and measured dynamic response of the structure at the drive point (point Acc 1 in Fig. 5(a)) are respectively shown with the black and gray lines on the top diagram of Fig. 7. The agreement is acceptable for the purposes of the study. It may seem that the experimental curve has more modes than the predicted one. The symmetry of the simulated structure might explain this: although lumped masses have been added in the model to account for the sensors and actuators, the perturbation may be too weak to break the symmetry and make all the modes appear in the predicted response. For example, between 300 and 400 Hz where five modes can be seen on the predicted response, closer examination shows that there actually are nine resonances. It is expected that with this level of accuracy for the base FE part of the model, the hybrid model will be accurate up to approximately 800 Hz.

The FE component of the complete hybrid model includes the in-plane motion of the panels. The resulting model comprises 68 shell elements (bottom plate), 1068 membrane elements (for the in-plane motion of the thin panels), and 204 beam elements (framework and bolts), giving a total of 924 nodes and 2659 degrees of freedom. The corresponding mesh is shown in Fig. 6. The bolts were described by using small beam elements with attention paid to the offset from the framework (which plays a significant role in the transmission of power to the flexural field of the panels).

A modal analysis was performed on the FE component of the hybrid model and 120 modes were extracted with frequencies below 1000 Hz. This is to be compared with more than 1000 modes that would arise if the bending modes of the panels were included. Some of the mode shapes are illustrated in Fig. 6. In most instances, the thin panels behave almost like rigid bodies (recall that only the in-plane motion is considered). Indeed, from a free wavelength analysis, it is found that the first membrane mode of the panels is expected to occur at around 2000 Hz. At lower frequencies the panels act to provide additional mass and static stiffness to the framework. Global modes of the type shown in Fig. 6 are expected to govern the ensemble-averaged response of the structure.

The bending motion of each panel was described with an SEA subsystem. The effect of the randomly added masses was distributed over the subsystem by increasing the mass density by a factor of 1.18. The damping of the panels was measured using the power injection method[3] with an instrumented hammer impact on a panel in isolation. The power input to the panel was estimated from the acceleration and force measurements at the impact point, and the energy was estimated from four accelerometers placed at random locations on the panel. The value of the damping loss factor calculated from a power balance at each frequency was found to be approximately constant across the frequency range of interest and equal to 1.1%.

3. Comparison with Monte Carlo experiment

The response of the structure due to a point force acting at the point Acc 1 in Fig. 5(a) was computed for each of the six observation points described above. The measured and computed drive point velocities are shown in the lower diagram of Fig. 7. For comparison, the measured and computed responses of the structure without the thin panels (see previous section) are shown in the upper diagram with the same vertical scale. The hybrid model reproduces some of the changes that occur when adding the thin panels to the framework. As already discussed, the membrane effect of the panel is mainly a reactive effect and this explains the shift of the natural frequencies. Conversely, the bending point impedance of the panels is resistive and corresponds mainly to a damping effect that changes the width of the resonance peaks (as seen in Fig. 7).

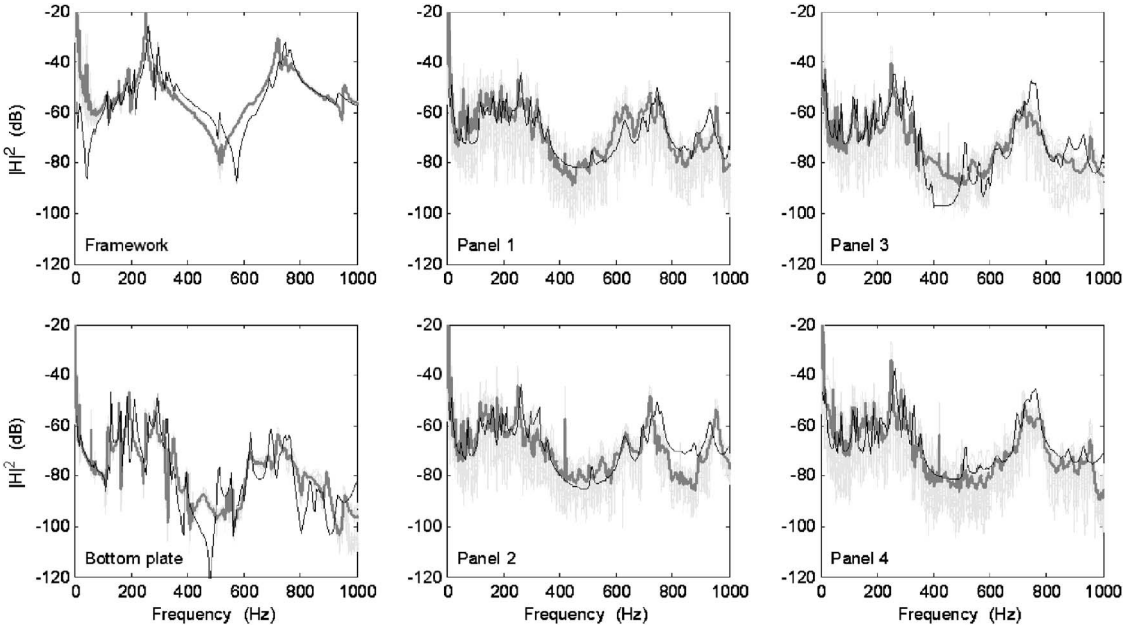


FIG. 8. Modulus squared velocity response at six locations of the test structure to a point force excitation on the framework. Light gray; 20 experimental realizations; bold gray: averaged experimental response; black: hybrid FE-SEA prediction.

The velocity squared response at each of the six observation points on the structure is shown in Fig. 8. Three sets of results are shown in each diagram: the experimental results for 20 realizations are shown in light gray; the ensemble-averaged experimental response is plotted with a bold gray line; and the hybrid prediction is plotted with a black line. The test data exhibit the mid-frequency characteristics, with a significant sensitivity of the response to perturbations, and at the same time some distinct features (peaks and drops) in the ensemble mean response. Those features are reasonably well predicted by the hybrid method. For instance, the response maximum at 260 Hz can be traced to global mode 35 (266 Hz) in Fig. 6; this mode has a large displacement at the excitation location in the direction of the excitation, and consequently features strongly in the hybrid prediction. It also appears in the experimental averaged data although it is not an actual natural mode of vibration of any single realization of the ensemble. The prediction, however, is not perfect. As previously noted, the FE model of the framework is valid only up to 800 Hz, and this might explain the approximate results yielded by the hybrid method above that frequency. This emphasizes the need for an accurate description of the deterministic part of the hybrid model: as the system randomness is assumed to arise in the short-wavelength components, only those components are suitable to a coarse description as provided by SEA. Any errors in the FE model will clearly propagate through to the response prediction.

The first two test cases above demonstrated how statistics can be included in a traditional FE model using the hybrid method. It is noticed that no explicit information was specified regarding the nature of the uncertainty of the SEA subsystems. It has been shown that the statistics underlying the SEA diffuse reverberant field concept corresponds to a state of maximum uncertainty,[13] and as such may not be identical to the actual uncertainty in a given structure. However, in most practical cases of engineering interest, very little is known about the actual uncertainty, and a state of maximum uncertainty provides a pragmatic approach that can be readily included in a model. In addition, recent studies have shown that the dynamic properties of an uncertain system at high frequencies exhibit the same universal statistics regardless of the exact nature of the uncertainty provided that the system is random enough.[4,18] It is then expected that the hybrid method yields an appropriate prediction of the ensemble average response of a system provided the subsystems modeled with SEA exhibit a sufficient amount of randomness across the actual ensemble.

C. Two panels connected by an extended line junction

Consider the structure shown in Fig. 9; two identical

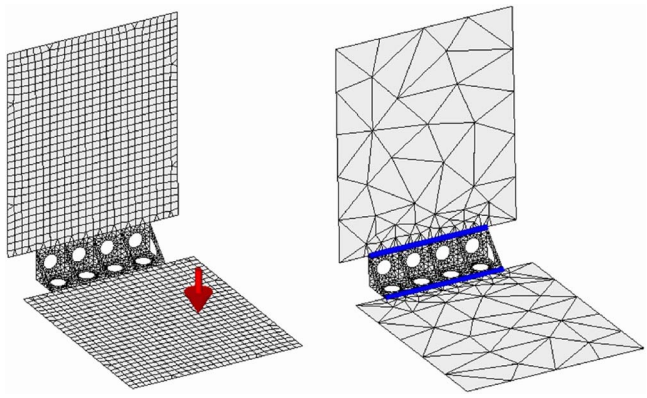


FIG. 9. (Color online) Two finite element meshes of a bracket-panels structure. Left: detailed mesh for Monte Carlo simulation (11 922 degrees of freedom); Right: mesh for hybrid prediction where the surface elements are membrane only (3502 degrees of freedom). The bold lines are the hybrid line junctions between the finite element freedoms and the panels bending wave fields described by SEA.

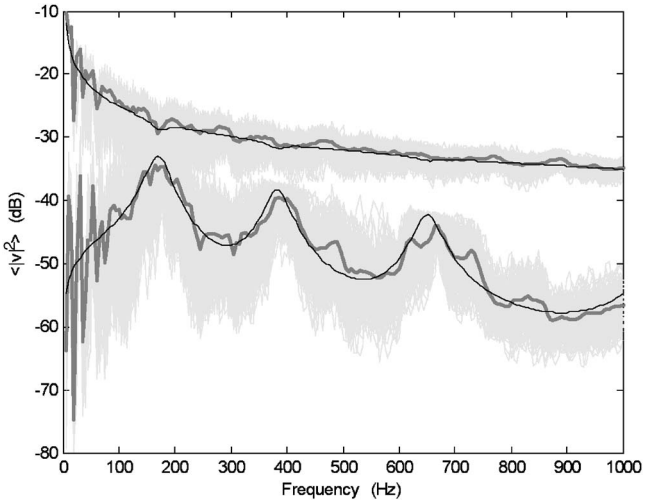


FIG. 10. Mean squared velocity response in the source and receiving plates. Light gray: reference Monte Carlo results for 200 realizations; bold gray: Monte Carlo ensemble average; black: hybrid FE-SEA.

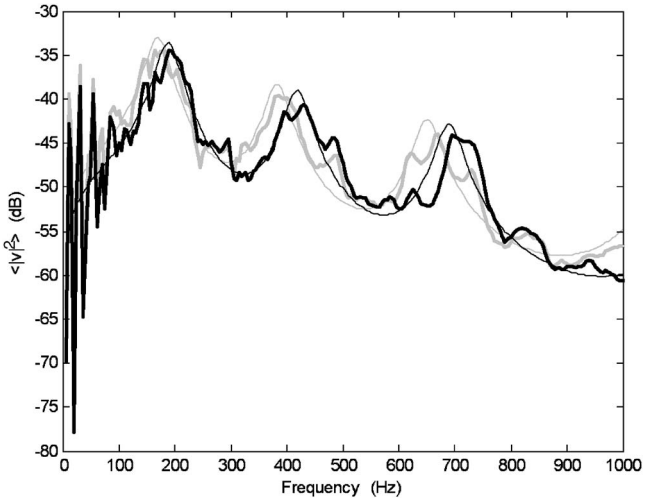


FIG. 11. Mean squared velocity response in the receiving plate for pinned (gray) and clamped (black) nodes at the bracket ends. Bold: Monte Carlo ensemble average; thin: hybrid FE-SEA.

thin aluminum plates of dimension 0.6 m $\times$ 0.7 m and thickness 1 mm are coupled by a stiff bracket. Interest lies in predicting the transmission of structure-borne noise between the two plates when one plate is excited by spatially distributed (rain-on-the-roof) excitation. The bracket is 0.4 m long and is made of 4 mm thick steel. The three corner points at each end of the bracket are pinned. The structure has a uniform damping loss factor of 1%.

A reference FE model with triangular and quadrangular shell elements was created with 551 nodes in the bracket and 773 nodes per panel; 370 modes were extracted with frequencies below 1200 Hz. In order to reduce computational expense when performing a Monte Carlo simulation, a component mode synthesis model was created using three super elements, and the space-averaged squared velocity in the plates due to rain-on-the-roof excitation was then calculated using the energy flow postprocessing equations presented by Mace and Shorter.[20] A Monte Carlo simulation was performed by adding ten masses to each plate at random locations, each mass being 0.5% of the total mass of the structure. The velocity response computed using the FE Monte Carlo simulation is plotted in Fig. 10. The transmission across the bracket is clearly sensitive to perturbation, while the ensemble average transmission exhibits a number of distinct peaks that are related to the local dynamic behavior of the bracket (the relative variance of the response in the receiving plate is significantly reduced around these peaks).

A hybrid model was created in which the stiff bracket was modeled in detail with FE using the same shell elements, and the (long wavelength) membrane behavior of the plates was modeled using a coarse mesh of triangular membrane elements (74 nodes per plate). The (short wavelength) flexural wave fields of the plates were modeled using two SEA subsystems. Two hybrid line junctions were used to couple the FE and SEA subsystems together. The hybrid line junction formulation described in the Appendix provides a method for calculating the direct field dynamic stiffness of an SEA subsystem. A modal analysis was performed on the FE subsystem and six modes were retained below 1200 Hz (to be compared with the 370 modes of the full FE model). Two of those modes at 547 and 554 Hz only involve the in-plane motion of the panels, and they are not expected to affect the response. The other modes at, respectively, 196, 428, 701, and 1200 Hz involve torsion of the bracket with simply supported ends. A point force was applied to one of the SEA subsystems and the hybrid model was used to calculate the response. The solution is computationally efficient and took approximately 1/100th of the time it took to compute 20 Monte Carlo trials. The ensemble average response predicted with the hybrid FE-SEA approach is in good agreement with the FE Monte Carlo simulations. In particular, the peak in the transmission can be explained by the local modal dynamics of the brackets (modified by the reactive and resistive impedance of the direct fields of the panels). The reactive part can be seen to be mainly a mass effect since the peaks in the transmission are lower in frequency than the natural frequencies of the bracket in isolation.

The computational efficiency of the hybrid approach enables robust design changes (i.e., design changes that are insensitive to uncertainties) to be quickly identified. As an example, the following section investigates modifications to the boundary conditions of the bracket and the value of damping. The resulting hybrid predictions are compared with additional FE Monte Carlo computations.

In the initial configuration, three nodes at each end of the bracket were pinned. This boundary condition was changed to clamped nodes and the hybrid model was re-run using the updated modes for the FE portion of the structure. The ensemble averaged mean squared velocity response of the transmitted plate obtained by the hybrid and Monte Carlo simulations is shown in Fig. 11. The initial configuration results from Fig. 10 are plotted in gray. The Monte Carlo results show that the primary effect of changing the bracket boundary conditions is to increase the frequency of the peaks in the transmission (arising from the stiffening of the structure). This change is correctly predicted by the hybrid model.

Keeping the initial boundary conditions, the damping loss factor of the bracket was changed from 1 to 10%. The

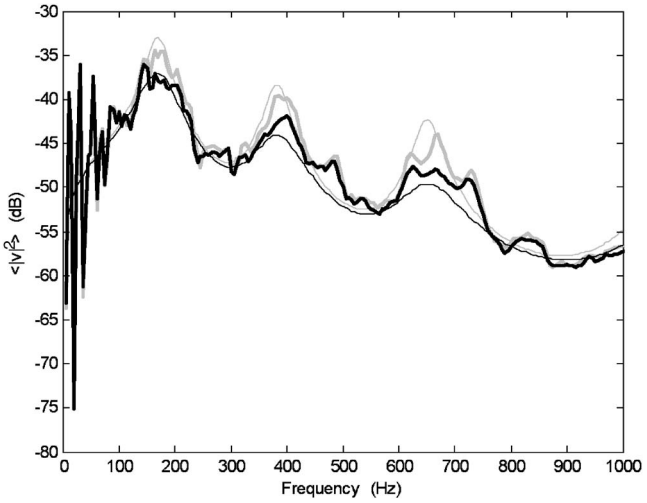


FIG. 12. Mean squared velocity response in the receiving plate for two values of damping loss factor of the bracket: 1% (gray) and 10% (black). Bold: Monte Carlo ensemble average; thin: hybrid FE-SEA.

ensemble averaged mean squared velocity response of the transmitted plate is shown in Fig. 12, where good agreement is seen between the Monte Carlo and hybrid results.

This last test case is typical of a problem traditionally handled by SEA, as interest lies in predicting the energy transmission between two subsystems with high modal density. The example shows how the hybrid method can be used to enhance a pure SEA model by providing a rigorous and automatic calculation of the SEA coupling loss factors. The method enables an arbitrary amount of deterministic detail to be added to a junction (for example, detailed modifications to the boundary conditions of the junction), and allows the junction to be dissipative (which is often problematic in traditional SEA). Similarly, by describing the vicinity of some external excitations with FE while keeping the rest of the model with SEA, the hybrid method can be used to improve the estimation of the power input into an SEA model. This application is not demonstrated here.

IV. CONCLUSIONS

A hybrid analysis method has been discussed which combines the finite element method with statistical energy analysis. The combination of these two methods in the same model provides an efficient way to model the vibro-acoustic behavior of systems with mid-frequency behavior. In particular, (i) components with few modes (or long wavelength behavior) are described deterministically with FE, as their dynamic behavior is robust to uncertainty; (ii) components with many modes (or short wavelength behavior) are described statistically with SEA, as their dynamic behavior is very sensitive to uncertainty. This partitioning matches the mix of dynamic behavior encountered in the mid-frequency range and also makes the method computationally efficient (since components that would require a large number of deterministic descriptive degrees of freedom are described statistically with SEA).

No information about the nature of the uncertainty in the SEA subsystems needs to be provided as those components are assumed to have a state of maximum uncertainty (for a given modal overlap, damping, and junctions impedance). The hybrid method predicts the mean response across this ensemble and provides a pragmatic nonparametric approach to deal with the lack of information regarding the system uncertainty.

The method has been shown to yield good results for the ensemble averaged response of a number of example systems. The first two validation cases demonstrated how statistics are easily introduced in a standard FE analysis, by modeling modally dense components using SEA subsystems. The third validation case demonstrated how deterministic details can be introduced in a standard SEA model, by modeling with FE the local details of a complex junction between a number of SEA subsystems. The numerical and experimental validations presented in this paper are focused on structure-borne noise problems, so that only point and line junctions have been demonstrated. The hybrid FE-SEA equations are, however, general and apply to any vibro-acoustic system including the coupling of structures and acoustic fluids though area junctions.

ACKNOWLEDGMENTS

The validation part of the work was funded by the Air Force Laboratory, Space Vehicles Directorate, Kirtland AFB, NM, under the SBIR Phase II Contract No. F29601-02-C-0109 managed by Dr. Steven Lane. The testing was performed by Dr. Mike Kidner and Dr. Cory Papenfuss at the Vibration and Acoustic Laboratory of the Virginia Polytechnic Institute.

APPENDIX: LINE IMPEDANCE ALONG THE EDGE OF A SEMI-INFINITE PLATE

Consider a hybrid model where a structural component described with finite elements (FEs) is connected to a plate described with SEA, along a straight junction of finite length L. The direct field dynamic stiffness matrix associated with the boundary degrees of freedom of the plate is found from the analysis of a segment of length L on the edge of the semi-infinite plate.

The displacement field along the segment is described with a set of shape functions: if FE nodal degrees of freedom are used, the shape functions are the elementary shape functions associated with each node in the segment; if a modal decomposition is employed, the shape functions are the mode shapes along the segment. In any case, the displacement at any point is expressed as the sum of the shape function contributions. Assuming thin plate theory, three translations and the rotation of the midplane fully describe the motion of the edge of the plate. Therefore, at any point x along the segment the four displacement components can be written: $[u(x),v(x),w(x),\theta(x)]^{\mathrm{T}}=\Sigma_j q_j \boldsymbol{\varphi}_j(x)$ where $\boldsymbol{\varphi}_j(x)=[u_j(x),v_j(x),w_j(x),\theta_j(x)]^{\mathrm{T}}$ is a shape function with four components.

The ji entry of the direct field dynamic stiffness looking into the plate is written

$$\{D_{\mathrm{dir}}\}_{ji} = \int_{-L/2}^{L/2} \boldsymbol{\varphi}_j^{*T}(x)\mathbf{f}_i(x)dx, \tag{A1}$$

where $\mathbf{f}_i(x)$ is the distribution of the four forces (dual of the four displacements) due to the prescribed displacement field $\boldsymbol{\varphi}_i$ along the segment.

In what follows, the displacement field outside of the junction is assumed to be zero (clamped boundary condition) so that the wave number representation of a given shape function is easily obtained with spatial Fourier transform:

$$\boldsymbol{\Phi}_i(k) = \int_{-\infty}^{\infty} \boldsymbol{\varphi}_i(x)\exp(-ikx)dx. \tag{A2}$$

The integral can be reduced to the interval $[-L/2, L/2]$ due to the clamped boundary condition assumed outside the junction.

The semi-infinite plate wave dynamic stiffness matrix $\mathbf{D}^{\infty}$ relates the four displacement wave number transforms to the four forces wave number transforms $\mathbf{F}_i$ by

$$\mathbf{F}_i(k) = \mathbf{D}^{\infty}(k)\boldsymbol{\Phi}_i(k). \tag{A3}$$

The interest in using the wave number representation is that the matrix $\mathbf{D}^{\infty}$ is analytically expressed in terms of the plate properties.[21] For a thin plate, the bending freedoms w and θ are not coupled to the shear/extension freedoms u and v, and the 4×4 wave dynamic stiffness matrix is thus made of two 2×2 block matrices

$$\mathbf{D}^{\infty} = \begin{bmatrix} \mathbf{D}_{se}^{\infty} & \mathbf{0} \\ \mathbf{0} & \mathbf{D}_{b}^{\infty} \end{bmatrix}. \tag{A4}$$

Applying the inverse Fourier transform to Eq. (A3) yields the force distributions along the edge

$$\mathbf{f}_i(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \mathbf{D}^{\infty}(k)\boldsymbol{\Phi}_i(k)\exp(ikx)dk. \tag{A5}$$

Substituting Eq. (A5) into Eq. (A1), and inverting the integration order yields the direct field dynamic stiffness entry as a function of the wave number transforms of the shape functions and the wave dynamic stiffness matrix

$$\{D_{\mathrm{dir}}\}_{ji} = \frac{1}{2\pi}\int_{-\infty}^{\infty} \boldsymbol{\Phi}_j^{*\mathrm{T}}(k)\mathbf{D}^{\infty}(k)\boldsymbol{\Phi}_i(k)dk. \tag{A6}$$

For stress-free boundary conditions outside of the junction, an alternative approach also based on Fourier transform has been developed, which makes use of the semi-infinite plate wave dynamic receptance matrix $\{\mathbf{D}^{\infty}\}^{-1}$.

In the validation case with line junctions presented in the paper, only the bending motion of the panels is modeled with SEA, while the in-plane motion of the panels is modeled with FE. Hence, the direct field dynamic stiffness of the panels only involves the bending contribution $\mathbf{D}_b^{\infty}$ in Eq. (A4) and only concerns the out-of-plane FE degrees of freedoms w and θ along the junction.

[1] M. S. Kompella and B. J. Bernhard, "Measurement of the statistical variation of structural-acoustic characteristics of automotive vehicle," in *Proceedings of the SAE Noise and Vibration Conference*, Warrendale PA (1993).

[2] R. Cornish, "A novel approach to optimizing and stabilizing interior noise quality in vehicles," *Proceedings of the Institute of Mechanical Engineers, Part D* – J. Automobile Eng.**214**(D7), 685–692 (2000).

[3] R. H. Lyon and R. G. DeJong, *Theory and Application of Statistical Energy Analysis* (Butterworth-Heinemann, Boston 1995).

[4] R. S. Langley and V. Cotoni, "Response variance prediction in the statistical energy analysis of built-up systems," J. Acoust. Soc. Am. **115**(2), 706–718 (2004).

[5] A. K. Belyaev and V. A. Palmov, "Integral theories of random vibration of complex structures," in *Random Vibrations – Status and Recent Developments*, edited by I. Elishakoff and R. H. Lyon (Elsevier, Amsterdam, 1986).

[6] C. Soize, "A model and numerical method in the medium frequency range for vibroacoustic predictions using the theory of structural fuzzy," J. Acoust. Soc. Am. **94**(2), 849–865 (1993).

[7] R. S. Langley and P. Bremner, "A hybrid method for the vibration analysis of complex structural-acoustic systems," J. Acoust. Soc. Am. **105**(3), 1657–1671 (1999).

[8] V. Jayachandran and M. W. Bonilha "A hybrid SEA/modal technique for modeling structural-acoustic interior noise in rotorcraft," J. Acoust. Soc. Am. **113**(3), 1448–1454 (2003).

[9] R. M. Grice and R. J. Pinnington, "Analysis of the flexural vibration of a thin-plate box using a combination of finite element analysis and analytical impedances," J. Sound Vib. **249**(3), 499–527 (2002).

[10] L. Ji, B. R. Mace, and R. J. Pinnington, "A mode-based approach for the mid-frequency vibration analysis of coupled long- and short-wavelength structures," J. Sound Vib. **289**(1-2), 148–170 (2006).

[11] S. B. Hong, A. Wang, and N. Vlahopoulos, "A hybrid finite element formulation for a beam-plate system," J. Sound Vib. **298**(1-2), 233–256 (2006).

[12] P. J. Shorter and R. S. Langley, "Vibro-acoustic analysis of complex systems," J. Sound Vib. **288**(3), 669–699 (2005).

[13] P. J. Shorter and R. S. Langley, "On the reciprocity relationship between direct field radiation and diffuse reverberant loading," J. Acoust. Soc. Am. **117**(1), 85–95 (2005).

[14] R. S. Langley, "On the diffuse field reciprocity relationship and vibrational energy variance in a random subsystem at high frequencies," J. Acoust. Soc. Am. **121**(2), 913–921 (2007).

[15] R. S. Langley and V. Cotoni, "Prediction of the ensemble mean and variance of the response of uncertain structures using the Hybrid FE-SEA method," in *Proceedings of ISMA2006 Conference*, Leuven, Belgium, September 18–20 (2006).

[16] R. S. Langley and P. J. Shorter, "The wave transmission coefficients and coupling loss factors of point connected structures," J. Acoust. Soc. Am. **113**(4), 1947–1964 (2003).

[17] R. S. Langley, P. J. Shorter, R. H. Lande, and V. Cotoni, "Hybrid deterministic-statistical modeling of built-up structures," in *Proceedings of ICSV12 Conference*, Lisbon, Portugal, July 11–14 (2005).

[18] R. S. Langley, "Natural frequency statistics and universality," in *Proceedings of ICSV12 Conference*, Lisbon, Portugal, July 11–14 (2005).

[19] VA ONE 2006 User's Guide, The ESI Group Oct. 2006.

[20] B. R. Mace and P. J. Shorter, "Energy flow models from finite element analysis," J. Sound Vib. **233**(3), 369–389 (2000).

[21] R. S. Langley and K. H. Heron, "Elastic wave transmission through plate/beam junctions," J. Sound Vib. **143**(2), 241–253 (1990)

Structural acoustic control of plates with variable boundary conditions: Design methodology

Joseph D. Sprofera[a)]
Pratt School of Engineering, Duke University, Durham, North Carolina 27708

Randolph H. Cabell[b)]
NASA Langley Research Center, Structural Acoustics Branch, Hampton, Virginia 23681

Gary P. Gibbs[c)]
NASA Langley Research Center, Quiet Aircraft Technology Project, Hampton Virginia 23681

Robert L. Clark[d)]
Pratt School of Engineering, Duke University, Durham, North Carolina 27708

(Received 5 March 2006; revised 4 November 2006; accepted 19 April 2007)

A method for optimizing a structural acoustic control system subject to variations in plate boundary conditions is provided. The assumed modes method is used to build a plate model with varying levels of rotational boundary stiffness to simulate the dynamics of a plate with uncertain edge conditions. A transducer placement scoring process, involving Hankel singular values, is combined with a genetic optimization routine to find spatial locations robust to boundary condition variation. Predicted frequency response characteristics are examined, and theoretically optimized results are discussed in relation to the range of boundary conditions investigated. Modeled results indicate that it is possible to minimize the impact of uncertain boundary conditions in active structural acoustic control by optimizing the placement of transducers with respect to those uncertainties. © *2007 Acoustical Society of America.* [DOI: 10.1121/1.2739404]

PACS number(s): 43.40.Rj, 43.40.Dx, 43.40.At, 43.40.Vn [KAC] Pages: 271–279

I. INTRODUCTION

Active control of aircraft cabin noise through the utilization of smart structures is a research area that has been thoroughly investigated since the late 1980s.[1] Prior to the introduction of active control methods for this application, passive methods of noise reduction were the primary options for controlling cabin noise.[2,3] Many of the passive methods involve negative aspects ranging from weight penalties, in the case of structural stiffening,[4] to lack of performance at low frequency, long-wavelength applications when employing surface damping treatments.[5] The use of passively tuned vibration absorbers was also deemed ineffective due to problems maintaining proper tuning in the presence of varying environmental conditions, typical of an aircraft while changing altitudes.

At the time of its introduction active control was found to be an exciting and attractive option for controlling aircraft cabin noise because of its high level of adaptability. However, there were a number of limitations observed during the initial implementation of active control. The use of large numbers of transducers often restricted active control attempts to controlled laboratory environments. Microphone, speaker, and accelerometer arrays were not uncommon components of an early active control system.[6] The arrays had the potential to occupy substantial space, add undesired weight, and require considerable amounts of wiring, not to mention the potentially prohibitive costs. Many of these issues were addressed with the introduction of active materials as a transducer alternative. The beneficial properties of active materials (such as high bandwidth of operation, convenient surface mounting, and spatial filtering through distributed actuation/sensing) became a major selling point for active control of aircraft cabin noise.

Initial investigations into using piezoceramic actuators for the active control of structurally radiated sound were based upon the use of single and multiple transducers.[7,8] Much of this work built upon a prior investigation involving the active control of interior aircraft noise induced by propellers[1] and was complemented by an investigation involving the reduction of sound transmission from elastic plates through vibration inputs.[9] The work done in these investigations was instrumental in advancing the shift of structural acoustic control focus from the passive to active realm.

A later group of investigations on the active control of aircraft cabin noise was conducted in 1996 on a deHavilland Dash-8 aircraft.[10] While the placement of the transducers was not optimized in this work, the results of the tests proved that it was feasible to use piezoceramic actuators in the reduction of interior cabin noise and fuselage vibration. Though the primary source of disturbance in the deHavilland experiments was propeller noise, similar research that included broadband noise due to boundary layer turbulence was conducted in 1997.[6] Actuator locations used in that

[a)]Present address: 1602 North Rowell Avenue, Manhattan Beach, California 90266. Electronic mail: joe.sprofera@gmail.com
[b)]Electronic mail: r.h.cabell@larc.nasa.gov
[c)]Present address: 2020 Kraft Drive, Suite 3030, Blacksburg, Virginia 24060. Electronic mail: gary@aeqisound.com
[d)]Electronic mail: rclark@duke.edu

study were optimized based upon previously acquired analytical data. Results of the study showed that reduction of interior noise could be achieved through active techniques, but that a practical system would still take some time to develop.

Clark and Frampton[11] conducted additional investigations involving active control of turbulent boundary layer noise transmission through an aeroelastic plate. This was just one of several investigations by Frampton and Clark[12,13] that involved the control of TBL noise transmission and aeroelastic coupling, investigations that served to increase the interest in finding applicable active solutions to reducing interior aircraft noise. Research into the analysis of TBL noise transmission through aircraft sidewalls continues to be a major point of investigation.[14]

While there has been much research into the design of adaptive structures capable of reducing noise transmission to the interior of an aircraft cabin, the design of such structures able to handle a range of boundary conditions while still reducing transmission has yet to be fully explored. A model capable of aiding in this design process cannot be based upon ideal boundary conditions alone. Ideal conditions, such as simply-supported or clamped boundaries, are often found in basic engineering models and serve as simplistic bounds for realistic boundary conditions. While ideal conditions can have constraints such as zero displacement or slope at a boundary, realistic conditions cannot. Consideration of varying boundary conditions found in real-world structures is important for understanding the potential impact these variations can have on optimal transducer placement.

This paper outlines a process for identifying the optimal size and location of a distributed actuator and sensor pair to control sound radiation from a thin plate with variable boundary conditions. The process involves the creation of a structural model capable of representing the plate and the expected range of boundary condition variations. Piezoelectric transducers are added to the structural model to serve as actuators and sensors, and radiation filters as detailed by Gibbs *et al.*[15] are incorporated to provide an estimate of radiated sound power for the performance metric. The placement and size of the piezoelectric transducers are scored using a metric that promotes coupling to low frequency modes that radiate sound efficiently while minimizing coupling to modes that radiate inefficiently. The scoring metric is combined with a genetic algorithm to produce candidate solutions. Results from this study indicate that it is possible to optimize actuator/sensor size and location to minimize the impact associated with uncertainty in boundary conditions on structural acoustic control performance.

II. THE STRUCTURAL PLATE MODEL

The structural plate model is developed using a standard assumed modes approach.[16] Rotational springs, which can be assigned varying levels of stiffness, are distributed along each of the four edges of the plate,[17] building a model capable of varying from simply supported to a near clamped condition. Figure 1 shows a schematic of the assumed rotational springs distributed along a single edge of the plate. For

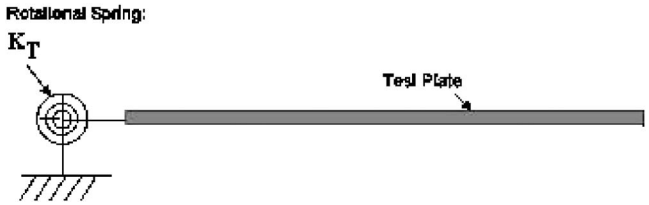


FIG. 1. Depiction of rotational spring along one edge of test plate.

such boundary conditions, the eigenfunctions of the pinned-boundary plate serve as admissible functions [Eq. (2)] for use in this model.[18]

From the expansion theorem, the response of the plate can be expressed as follows:

$$w(x,y,t) = \sum_{r=1}^{M} \sum_{s=1}^{N} \phi_{rs}(x,y)\,\eta_{rs}(t), \tag{1}$$

where $\eta_{rs}(t)$ is the time domain response in modal coordinates, and

$$\phi_{rs}(x,y) = \frac{2}{\sqrt{mab}} \sin\left(\frac{r\pi x}{a}\right)\sin\left(\frac{s\pi y}{b}\right). \tag{2}$$

From Eq. (2), ϕ_{rs} represents the mass normalized mode shape of a simply supported plate. Additionally, m is the mass per unit area of the plate, a is the length of the plate in the x direction, and b is the length of the plate in the y direction.

Since the purpose of this study is active control, an upper frequency limit (1250 Hz) is identified, above which it would be expected that passive methods may be more effective. Since the fundamental mode of the system is one of primary interest, a lower bound on the frequency range is not needed (it can be assumed to be 0 Hz). The plate model is generated for the first 30 assumed mode shapes in the x and y directions, respectively. This generates a total of 900 assumed modes.

The generation of state-space equations is accomplished using energy functions for the structural model development.[19] It is possible to create expressions for the kinetic (T) and potential (V) energy of the plate as follows:

$$T = \frac{1}{2}\rho h \int\int_{\text{area}} \left[\sum \phi_i(x,y)\dot{q}_i\right]\left[\sum \phi_j(x,y)\dot{q}_j\right] dx\,dy. \tag{3}$$

In the above equation for T, ρ is the density of the plate, h is the thickness, and $\dot{q}_i$ is the modal velocity.

$$V = \frac{1}{2}\int\int_{\text{area}} D_e\left[\left(\frac{\partial^2\phi}{\partial x^2}\right)^2 + \left(\frac{\partial^2\phi}{\partial y^2}\right)^2 + 2\nu\left(\frac{\partial^2\phi}{\partial x^2}\right)^2\left(\frac{\partial^2\phi}{\partial y^2}\right)^2 + 2(1-\nu)\left(\frac{\partial^2\phi}{\partial x\partial y}\right)^2\right] dx dy + \frac{1}{2}K_{T1}\int_0^{L_y}\left[\frac{\partial\phi(0,y)}{\partial x}\right]^2 dy + \frac{1}{2}K_{T2}\int_0^{L_y}\left[\frac{\partial\phi(L_x,y)}{\partial x}\right]^2 dy$$

$$+\frac{1}{2}K_{T3}\int_0^{L_x}\left[\frac{\partial\phi(x,0)}{\partial y}\right]^2 dx$$
$$+\frac{1}{2}K_{T4}\int_0^{L_x}\left[\frac{\partial\phi(x,L_y)}{\partial y}\right]^2 dx. \quad (4)$$

In the above equation for V, D_e is $E_s h_s^3/12(1-\nu_s^2)$, where E_s is Young's modulus, h_s is the plate's thickness, and ν_s is Poisson's ratio. The terms containing the K_T values represent the distributed rotational springs present along the four edges of the plate. Using the assumed modes and the expressions for kinetic and potential energy, it is possible to derive a mass and stiffness matrix for the system upon applying Lagrange's equations of motion in conjunction with Eq. (1).[16]

The eigenvectors and eigenvalues of the finite dimensional plate model are calculated after the formation of the mass and stiffness matrices. The stiffness matrix accrues off-diagonal terms due to the rotary springs at the boundaries. Each spring adds torque proportional to the rotational angle of the corresponding edge as well as the physical properties of the spring itself (contained in the K_T terms). This angle is determined by examining the rate of change of the deflection (or height) of the plate along each particular axis [$\partial\phi(0,y)/\partial x$ for example]. The springs contribute nothing to the kinetic energy since they are assumed massless.

The resulting finite-dimensional eigenvectors can be used to form the approximate mode shapes of the system from the admissible functions. The solution is known to converge asymptotically from above, so the greater the number of assumed modes, the better the estimate of the natural frequencies of the structure. Using the first 30 simply supported mode shapes in the x and y directions, it was possible to produce natural frequencies for the first 40 plate modes that when compared to an established source[20] exhibited less than a 3% error.

III. RADIATION MODEL

The goal of this project is to provide a design tool for optimizing transducer size and location for active control of sound transmission through a plate subject to a varying range of boundary conditions. Given this goal, the performance metric is radiated sound power. It is well known that some structural modes radiate more efficiently than others,[21] and thus the performance metric, constructed from radiation filters, is directed at control of these modes. Using discrete structural measurements, radiation modal expansion (RME)[15] can be used to quickly estimate the sound power radiated in the context of a state-space model, enabling rapid computation and comparison of actuator and sensor pairs during the optimization procedure.

The RME method is based on the concept of radiation modes of a structure, described by Elliott and Johnson,[22] whereby the sound radiation from a structure is assumed to consist of contributions from a finite number of discrete elemental radiators. The velocity of each elemental radiator is measured using an accelerometer placed at the center of the radiator. The spacing and number of elemental radiators, which corresponds to the number of accelerometers deployed, must satisfy constraints based on acoustic wavelength and structural wave number.[22] The radiated sound power is given by the product of the velocities of the elemental radiators $[a(j\omega)]$ and a radiation resistance matrix $[R(j\omega)]$:

$$\bar{P} = a(j\omega)^H R(j\omega) a(j\omega). \quad (5)$$

Radiation modal expansion uses an approximation to the frequency dependent radiation resistance matrix, R, that is compact and easy to compute. If the singular vectors of R at a frequency ω describe the radiation modes of R at that frequency, RME assumes that R can be described across a broad frequency range using a weighted summation of the radiation modes computed at that frequency. The accuracy of this approximation depends on the frequency where the radiation modes are computed, relative to the frequency where the approximation is used. For the work presented, the radiation modes are computed at the first natural frequency of the plate model for all possible boundary conditions. For applications where the bandwidth of the disturbance is limited, a finite number of radiation modes can be used to accurately estimate the radiated sound power from the plate.[15] The RME approach has been used and validated in past experimental studies.[23,15]

For the current application, the RME method is used with a single radiation mode. The plate model is divided into a five by three grid of 15 equal elemental radiators, and the acceleration is computed at the center of each radiator. The grid size limits accurate plate mode detection to the (5,3) mode; however, since the resonant frequency of this plate mode is well above the frequency range of interest, the number and arrangement of radiators is suitable for this investigation.

Using these discrete measures of acceleration on the plate to compute the response of the first radiation mode, the resulting estimate of radiated sound power is accurate to within 25% of that computed using the first 50 radiation modes. For the purpose of design optimization, the accuracy lost is deemed reasonable for the computational time gained. For more complex designs, one might iterate initially with a low-order radiation filter and then use the optimized results to confirm the solution with a higher order radiation filter.

IV. MODELING OF TRANSDUCERS

An important consideration in developing a piezoceramic actuator and sensor model was how to enable rapid evaluation of arbitrarily sized patches during the optimization process. Typical piezoceramic-structural models require the evaluation of two integrals over the domain of the model to determine the electromechanical coupling and the capacitance. To avoid these integrals, a discrete integration approach was used to model coupling between the patches and the plate. This discrete integration method was previously used to optimize patch locations for suppressing flutter on a delta wing.[24]

In this approach, a grid of 4-mm-square piezoceramic elements was distributed across the surface of the plate. The

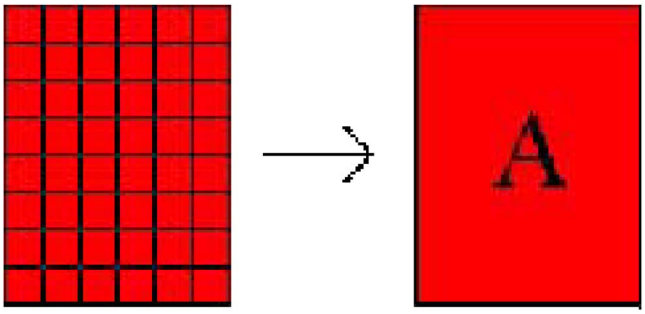


FIG. 2. (Color online) Element summation method.

electromechanical coupling and capacitance were calculated at the center of each element and multiplied by the element's area to obtain approximate values for each element. The capacitance and electromechanical coupling of larger rectangular patches are then computed by summing the capacitance and coupling values for all 4-mm elements whose centers fell within the larger patch's boundaries. Figure 2 shows an example of how a larger patch was created from the 4-mm elements.

The approach is accurate as long as the modal wavelength is larger than the discrete element used in the approximation. For the purpose of this study, modes with spatial wavelengths on the order of 4 mm are beyond the frequency range of interest by more than an order of magnitude. Additionally, the patches themselves are designed with the intent of providing coupling to out-of-plane motion.

It should be noted that the mass and stiffness effects of the piezoelectric patches are ignored. For the sensor patch, which is assumed to be polyvinylidene fluoride (PVDF), this is clearly acceptable since PVDF mass and stiffness properties are negligible in comparison to that of a steel plate. The likely actuator is lead zirconate titanate (PZT), which takes additional consideration because it can, depending on the relative thickness of the actuator to the plate, contribute to the mass and stiffness of the system at a significant level. For the structure under consideration, simple calculations revealed that the error in predicting the estimated resonant frequencies of the system would be less than 3%.

V. DESCRIPTION OF THE PLATE MODEL

The modeled test structure is a 0.508 m by 0.254 m steel plate with a thickness of 2.24 mm. A Young's modulus of 200×10^9 Pa and a Poisson's ratio of 0.29 were used. The mass per unit area of the plate was 17.433 kg/m^2, giving a density of 7782 kg/m^3.

VI. CREATING THE BOUNDARY CONDITIONS

In order to simulate a wide variety of boundary conditions, it is necessary to generate values for the distributed rotational springs present at the plate edges. The spring values used at the edges of the plate vary from 0 $(N\cdot m/rad)/m$, meant to represent a simply supported end condition, to 100 000 $(N\cdot m/rad)/m$, which was found to be a suitable stiffness in representing a near clamped end condition. Each edge of the plate is independently assigned a random value between 0 and 100 000 $(N\cdot m/rad)/m$ to create a boundary condition used in the optimization routine. Thirty different plate models were created. For each plate, random stiffness values (between 0 and 100 000) were independently assigned to each of the four edges. Each time the optimization routine is executed, a completely new set of 30 random plate models is created, meaning the optimization proceeds from a truly random set every time.

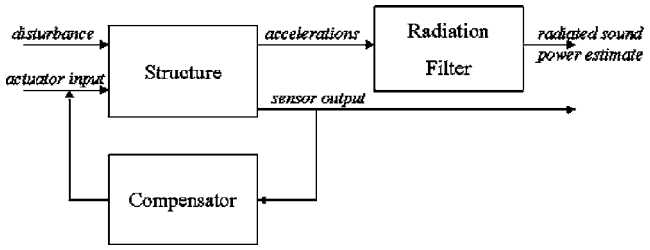


FIG. 3. System block diagram.

VII. DEFINING THE PERFORMANCE METRIC FOR TRANSDUCER PLACEMENT

The bandwidth of interest for structural acoustic control was limited to 800 Hz; however, the frequency range used in the performance metric extended to 1250 Hz. The purpose for extending the bandwidth in the performance metric is to identify actuator/sensor pairs that couple well to modes that radiate sound efficiently up to 800 Hz and couple poorly to modes between 800 and 1250 Hz to provide some natural roll-off in the response (spatial loop-shaping).[25] To increase the efficiency of the optimal patch pair, the in-bandwidth (0–800 Hz) coupling of a pair is rewarded while the out-of-bandwidth (800–1250 Hz) coupling is penalized. Due to the varying nature of the boundary conditions used in this study, the number of modes that fall below 800 Hz is model dependent.

In order to compare potential piezoceramic patch pairs to one another, a performance metric based on the calculation of the Hankel singular values (HSVs)[26] is used. HSVs define the energy of each state in the system relative to a specific system path and are numerically based on the observability and controllability Grammians of the system. The HSVs are useful in the design of performance metrics that quantify the level of controllability and observability related to a system path and have been used in the actuator/sensor design.[27,26,28–30]

The overall performance metric involves not only the actuator/sensor path but also the disturbance/performance path of the system (see Fig. 3). The performance path has been identified as the radiated sound power, and the disturbance path consists of stochastic point-force inputs applied to discrete points on the structure. Rather than using the HSVs of the disturbance/performance path, versions of the HSVs normalized with respect to the greatest disturbance/performance HSV were used.

The calculation of the HSV is accomplished using Eq. (6):

$$J_{\text{perf}} = \sum_{i=1}^{N} \Lambda_{\text{perf}} \gamma_{uy_i}^2 \bar{\gamma}_{wz_i}^2, \tag{6}$$

where Λ_{perf} is a binary selection vector used to specify the modes involved in the HSV calculation, and γ_{uy_i} and γ_{wz_i} are the approximate HSVs of the actuator/sensor (input-output) path and normalized HSVs of the disturbance/performance path normalized with respect to the greatest disturbance performance HSV path. The normalized γ_{wz_i} is used to remove any gains introduced in that path and instead create a simple weighting vector that is also physically easier to interpret.

The response of out-of-bandwidth modes can lead to spillover and stability issues, particularly when a controller with a high gain is being implemented. Typically, low-pass filters are used to attenuate system response at higher frequencies beyond the desired bandwidth for control. However, having poor observability and controllability of out-of-bandwidth modes designed into the open-loop system is the best method of attenuating the undesirable high frequency response[25] since no additional dynamics are introduced.

HSV scoring in the out-of-bandwidth region is not considered. Instead, a ratio between in- and out-of-bandwidth response levels was calculated. The numerator of the ratio was obtained from the minimum amplitude of response in the frequency domain corresponding to the resonant frequencies of the (1,1), (1,3), and (3,1) modes (the efficient radiators). The denominator of the ratio is the h_∞ norm of the response over the bandwidth from 800 to 1250 Hz. The result of this ratio is multiplied by the in-bandwidth score to produce the overall score for a single patch pair at one boundary condition. When the ratio is close to 1, indicating that the in- and out-of-bandwidth responses are near the same level, there is no positive weighting involved for the overall score. However, when the ratio is greater than one, which indicates that the in-bandwidth response is higher than the out-of-bandwidth response, the overall score will increase. This ratio establishes a physically intuitive relationship between the in-bandwidth coupling and out-of-bandwidth roll-off without requiring the calculation of HSVs:

$$\text{OUT}_{\text{ratio}} = \frac{\text{min.resonant response (1,1), (1,3), and (3,1) modes}}{h_\infty \text{ norm from 800 to 1250 Hz}}. \tag{7}$$

The in-bandwidth score is combined with the out-of-bandwidth ratio $\text{OUT}_{\text{ratio}}$ to create the overall score, as shown in Eq. (8):

$$J = J_{\text{perf}} * \log(\text{OUT}_{\text{ratio}}). \tag{8}$$

In the design process, the in- and out-of-bandwidth regions are chosen to afford loop-shaping prior to the implementation of a low-pass filter. When low-pass filters are added to a control system, the dynamics of the filters, particularly phase delay, can limit performance over the desired bandwidth of the control system since gain margin and phase margin are both critical with respect to stability margins. If actuator/sensor paths can be chosen so as to provide natural roll-off in the frequency response and thus afford the implementation of low-pass filters at a higher frequency, more aggressive controllers can be designed over the targeted bandwidth for performance.

When the initial pool of patches is created, the actuator/sensor pairs are scored using the previously described performance metric, but this scoring takes place for each of the different plate/boundary condition configurations. This produces several scores for each of the actuator/sensor pairs, which are accumulated in a simple summation to produce an overall score for that pair relative to all possible systems.

The scoring process takes into account all of the relevant inputs and outputs of the system: the actuator, input disturbance, sensor, and performance (radiated sound power as predicted by the radiation filters). After a patch pair is defined, a state space system of the combined piezo-structure is created. The creation of this system is expedited by the use of the piezoelectric element grid outlined previously. The resulting state space system incorporates the actuator input and sensor output specific to the patch pair being investigated.

The input disturbance used for these simulations is in the form of a 4 (long dimension) by 3 (short dimension) grid of equally spaced point forces. The input disturbance and radiation filter output do not change with the different actuator/sensor pairs, only with the changing boundary conditions. The result of this modeling step is a state space system combining the actuator/sensor pair and each of the plate model/boundary condition combinations. Each of these state space systems is then used in the scoring process, resulting in a score for each actuator/sensor pair and boundary condition grouping. These scores are then combined to create the overall score for a given actuator/sensor pair.

VIII. APPLICATION OF THE GENETIC ALGORITHM

A genetic algorithm was used to select an actuator/sensor pair that maximized the performance function when applied to several plates with different boundary conditions.[29] Genetic algorithms are adaptive search algorithms based upon the ideas of evolution, natural selection, and genetics. The basic concept of a genetic algorithm follows the principles of survival of the fittest. Genetic algorithms are optimization tools that perform mating and mutation operations on a population. For the case of this investigation, a single individual within the population is actually an actuator and sensor pair and not a single patch.

The parameters of each pair are placed in a row vector and converted into binary form, creating a parameter vector of ones and zeros. There are eight total parameters describing an individual relating to the height, width, x-center location, and y-center location of the actuator and sensor. Before the binary conversion of these parameters, they are normalized with respect to the maximum allowable values for each, resulting in all values being equal to or less than 1. This normalization prevents any possible future patch generations from obtaining parameter values above the set maximums, thus preventing invalid patch pairs from being created.

The mating process begins with a pool of 20 completely randomly generated individuals. The top ten individuals based on the cumulative performance score are selected for the creation of the next generation. This cumulative score is the result of summing the performance metric results for an individual in relation to each of the boundary conditions being investigated. Using the top scoring individuals increases the probability of producing high scoring offspring. Each of the ten individuals is mated with two other randomly selected individuals. The result of the initial mating step is 20 new offspring.

The mating of two individuals is accomplished by splitting the binary parameter vector of each and swapping the front and back ends to create new vectors. Before these offspring are used in the next generation of mating, they go through a mutation process. The mutation process simply switches a set percentage of random bits within the parameter vector of each individual from zeros to ones or vice versa. The mutation percentage can be varied for different results. A low percentage means there will likely be few drastic changes in the parameters while a high percentage will change the parameters to the extent that they will not maintain the generational characteristics desired from the parents. Finally, before beginning the mating process with the new generation the highest scoring individual from the parent generation is inserted among the offspring. This ensures that the highest scoring pair in each generation is at least equivalent to that of the previous generation, and provides for a quicker convergence on an optimal solution.

It is suggested that if the optimized results are similar at the end of the trials when starting with completely random conditions, then the impact of uncertain boundary conditions can be minimized by using this optimization routine. In this manner, the use of the performance metric and genetic optimization routine can provide an actuator/sensor pair that will produce a certain minimum level of radiated sound power reduction. This performance is considered independent of the boundary condition of the system, provided that condition is bounded by the random condition set used in the optimization process.

IX. INVESTIGATING THE RESULTS OF THE OPTIMIZATION PROCESS

One complete run of the genetic algorithm produces a single actuator/sensor pair that scores highly according to the performance metric described. Since it is possible to converge on a local maximum rather than a global maximum, several dozen trials of the genetic algorithm are executed. While there certainly exists one actuator/sensor pair that will result in the highest score, very slight variations in size and placement of the actuator/sensor pair will result in very similar scores. Due to the random nature of the mutation and mating process, limited generation runs will by definition yield suboptimal results, but further convergence would only serve to improve performance by tenths of a percent as designed for this study. A group of unlimited generation runs would theoretically always reach the same final actuator/sensor pair, but running trials to 100 generations yields repeatable results for the work presented here.

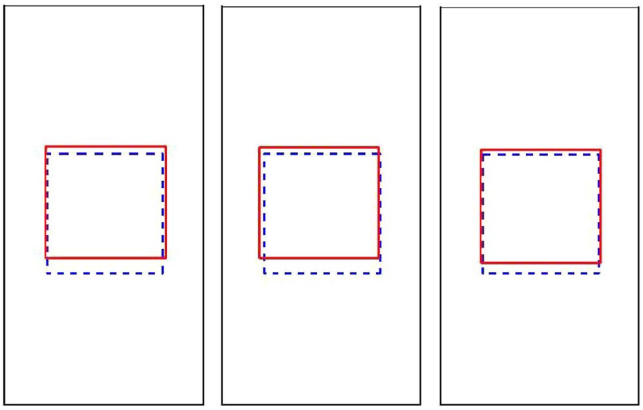


FIG. 4. (Color online) Top three scoring actuator (solid line) and sensor (dotted line) pairs.

The optimization process was executed several dozen times to confirm that 100 generations were sufficient to yield consistent convergence from an arbitrary starting point. The actuator/sensor pairs shown in Fig. 4 were found to be the highest scoring. The results (Fig. 4) appear intuitively correct in terms of what should be considered optimal for the requirements placed on the system. It is expected that actuator/sensor pairs will tend to move away from an edge that is nearly clamped since plate responses will be significantly lower in this region, and a central location is expected since the modal indices for modes that contribute to efficient acoustic radiation are odd.

Examining the top pairs, it is apparent that actuator size and location are very similar on each plate. The sensors have very similar locations as well, exhibiting differences slightly greater than that seen between the actuators. Detailed actuator and sensor center locations, heights, and widths for the top five scoring pairs are found in Table I. The average score for the top ten patch pairs resulting from the metric used was 125. Among those top ten pairs, the deviation from that average score was less than 1%, which was a desired result since it is well known that modal sensors can be quite sensitive to errors in placement.[31]

The optimal actuator/sensor pair yields a frequency response function that meets the bandwidth coupling and decoupling characteristics imposed by the performance metric. Figure 5 shows the high level of coupling for the in-bandwidth region (from 0 to 800 Hz), coupling desired of the optimal actuator/sensor pair to provide control over structural modes that contribute to efficient acoustic radiation. The magnitude of the coupling is lower in the out-of-

TABLE I. Actuator and sensor position details (all values in meters).

Actuator			Sensor		
Center	Height	Width	Center	Height	Width
(0.1245, 0.2510)	0.1404	0.1524	(0.1255, 0.2530)	0.1446	0.1518
(0.1335, 0.2570)	0.1416	0.1500	(0.1255, 0.2510)	0.1404	0.1518
(0.1165, 0.2550)	0.1416	0.1524	(0.1295, 0.2570)	0.1357	0.1524
(0.1195, 0.2470)	0.1452	0.1488	(0.1255, 0.2630)	0.1452	0.1524
(0.1255, 0.2550)	0.1428	0.1524	(0.1404, 0.2570)	0.1512	0.1518

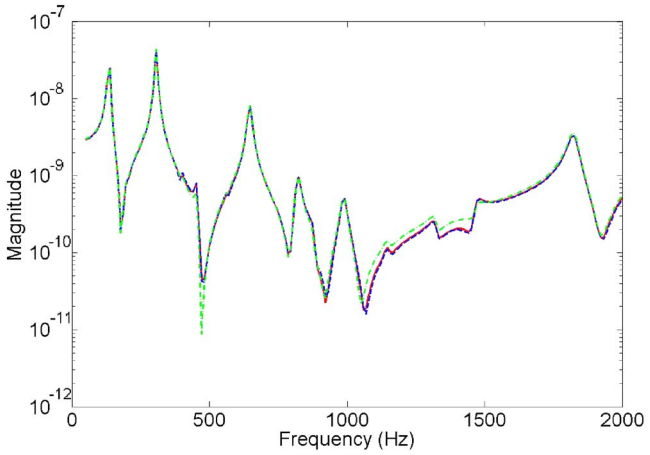


FIG. 5. (Color online) Optimal patch pair frequency response functions.

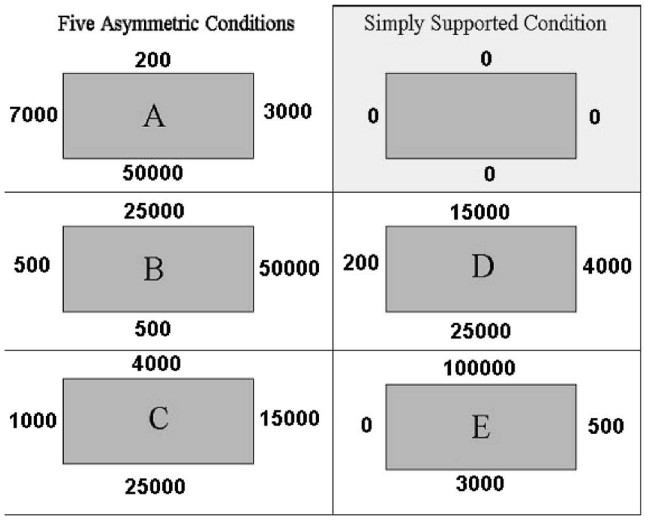


FIG. 7. Boundary condition sets.

bandwidth region, chosen between 800 and 1250 Hz, where the coupling of the response relative to the in-bandwidth response is effectively penalized in the performance metric. The lower level of coupling in this region can help prevent excessive spillover of control energy into modes that are not part of the targeted group, affording more aggressive controllers. By reducing the out-of-bandwidth response, the corner frequency of the low-pass filter used to reduce noise at high frequency can be increased relative to the targeted bandwidth for closed-loop performance, reducing the impact of phase delay in the targeted control bandwidth.

At higher frequencies, beyond 1250 Hz where coupling is ignored, the magnitude of the response again increases. Low-pass filters would be used to provide desired roll-off at approximately 1500 Hz for the example provided to attenuate the response and contribution of noise at high frequencies. The frequency response functions for one of the random boundary conditions investigated for each of the three orientations shown in Fig. 4 are plotted in Fig. 5. The specific random boundary condition associated with these response functions had stiffness values of 1000, 4000, 15 000, and 25 000 $(\mathrm{N}\cdot\mathrm{m}/\mathrm{rad})/\mathrm{m}$ at the four edges (condition D in Fig. 7). The frequency response functions from actuator to sensor of the three different pairs are very similar across the entire bandwidth of the plot from 0 to 2000 Hz, which is consistent with the fact that the scores from the performance metric differed less than 1%.

X. CONTROLLER DESIGN

Now that the optimization of a sensor/actuator pair has been investigated and potential solutions have been presented, discussing the design of an optimal controller is necessary. The actual design of the controller uses the linear quadratic Gaussian approach (LQG) for a single-input, single-output controller based on an identified system model. The linear quadratic regulator (LQR) is paired with a linear quadratic estimator (LQE) based on the identification of the actuator to sensor system path, in combination with a specific measurement noise floor that limits the modes targeted by the controller. The aggressiveness of the controller is determined by the values set for the measurement noise in the estimator and the control penalty in the regulator. Manipulation of the measurement noise and control penalty also alters the actuator effort level. The actuator effort level is computed from the H_2 norm between the applied disturbance and the measured output from the compensator (input to the actuator). Calculation of this effort level is performed to enable a mechanism for comparing the effort of controllers across the various boundary conditions examined. The overall system model contains two inputs and two outputs, but with only one of the outputs being fed back through the controller, the controller itself can be considered single-input, single-output (Fig. 6).

XI. PREDICTED CLOSED-LOOP RESULTS

Controllers were designed for each of the orientations shown in Fig. 4 for the six configurations of boundary conditions illustrated in Fig. 7. One of these conditions is the "near"-simply-supported case. Comparing the cumulative performance of each actuator/sensor pair using controllers with equivalent control signal energy over several boundary condition sets provides insight into the potential performance of each pair. The pair that consistently provides the highest level of performance across the conditions examined is likely to provide the best performance in the presence of an unexamined, random boundary condition. The first orientation shown in Fig. 5 demonstrated the highest level of performance after the design of the equivalent controllers and was chosen to be the patch pair used in the future experimental system.

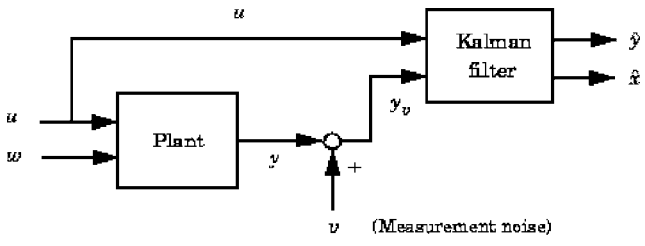


FIG. 6. Block diagram of controller design.

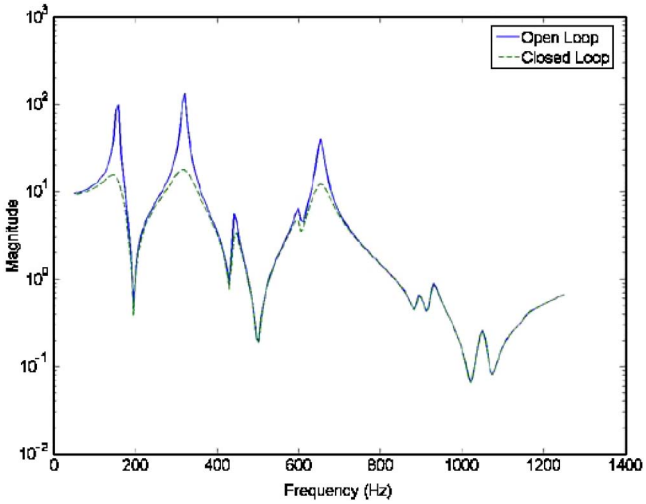


FIG. 8. (Color online) Open (solid) and closed (dashed) loop actuator-sensor FRF random BC A.

Figures 8–10 show the open and closed loop magnitude of the actuator-sensor frequency response function results for the first actuator/sensor pair when examined in relation to three of the random boundary conditions (conditions A, B, and C shown in Fig. 7) involved in the optimization process. In each figure, the dashed line represents the closed loop response. The actuator to sensor path was chosen for the frequency response function since the results of the control system and the effect of patch placement are easiest to observe within that path. The high level of coupling to the three major structural responses of interest, the roll-off of coupling after 800 Hz, and the clear reduction at the peak responses are all best observed within these frequency response functions.

It can be seen that in each case, despite the fact that the boundary conditions are completely randomized, the optimized actuator/sensor pair is capable of reducing the response of the plate. For each of the three conditions examined, the closed loop response demonstrated a reduction of approximately an order of magnitude for the first two dominant peaks and a factor of 5 for the third dominate peak. The dominant peaks appear at similar frequencies in the three plots despite different boundary conditions because the range that the dominant modes can vary over is not significant. The optimized size and position of the actuator/sensor pairs effectively capture responses over these frequency ranges regardless of the plate boundary conditions.

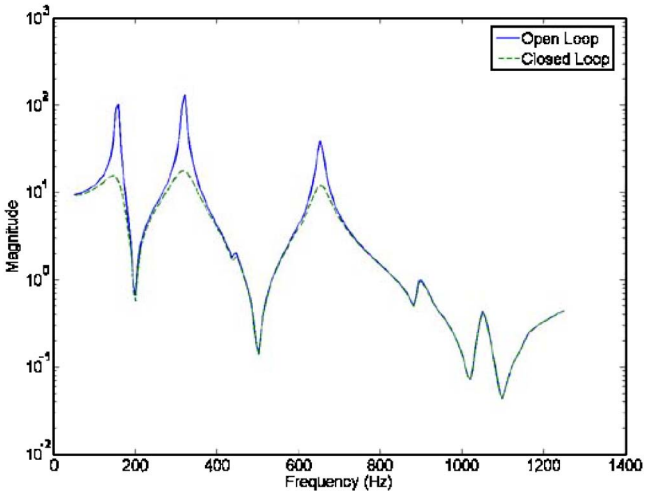


FIG. 9. (Color online) Open (solid) and closed (dashed) loop actuator-sensor FRF random BC B.

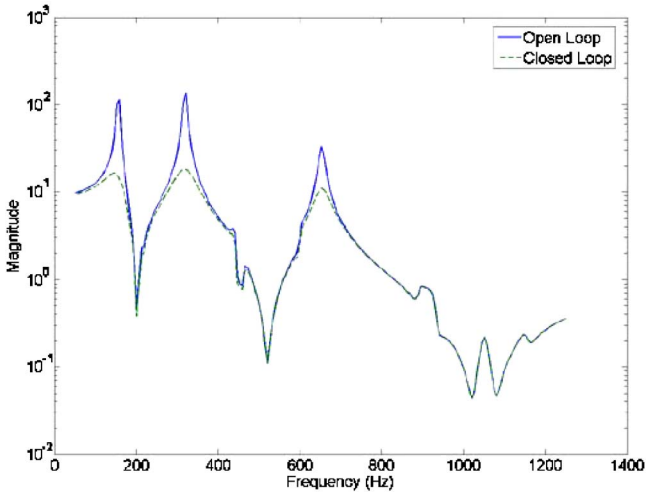


FIG. 10. (Color online) Open (solid) and closed (dashed) loop actuator-sensor FRF random BC C.

XII. CONCLUSION

This paper describes an optimization procedure for determining the location and size of a distributed actuator and sensor to reduce sound radiation from a plate with uncertain boundary conditions. Dimensions of the plate were chosen to be representative of typical aircraft panels. Results from this study indicate that it is possible to use a single actuator/sensor pair on panels of similar size and aspect ratio, with variations in boundary conditions ranging from any combination of simply supported to clamped, to consistently reduce the closed-loop response of modes that contribute significantly to acoustic radiation. In the practical application of active structural acoustic control, variations in boundary conditions are expected as a result of inconsistent assembly and mounting conditions. Providing a design tool that addresses this uncertainty and provides actuator and sensor locations and dimensions that meet desired performance objectives under such conditions is a positive step towards future commercial translation of this work.

ACKNOWLEDGMENTS

The authors gratefully acknowledge the Structural Acoustics Branch at NASA Langley Research Center and NASA for funding this research under Grant No. NAG-1-01014.

[1]C. R. Fuller and J. D. Jones, "Experiments on reduction of propeller induced interior noise by active control of cylinder vibration," J. Sound Vib. **112**(2), 389–395 (1987).
[2]G. SenGupta, "Reduction of cabin noise during cruise conditions by stringer and frame damping," AIAA J. **17**(3), 229–236 (1979).
[3]J. S. Mixson, L. A. Roussos, and C. K. Barton, "Laboratory study of

efficient add-on treatments for interior noise control in light aircraft," 7th AIAA Aeroacoustics Conference, Palo Alto, CA, 5–7 October, Paper No. AIAA-81-1969 (1981).
[4]R. Vaicaitis, M. Slazak, and M. T. Chang, "Noise transmission and attenuation by stiffened panels," 6th AIAA Aeroacoustics Conference, Hartford, CT, 4–6 June, Paper No. AIAA-80-1034 (1980).
[5]F. W. Grosveld and J. S. Mixson, "Noise transmission through an acoustically treated and honeycomb stiffened aircraft sidewall," 9th AIAA/NASA Aeroacoustics Conference, Williamsburg, VA, 15–17 October, Paper No. AIAA-84-2329 (1984).
[6]G. P. Mathur, B. N. Tran, M. A. Simpson, and D. K. Peterson, "Broadband active structural acoustic control of aircraft cabin noise—laboratory tests," American Institute of Aeronautics and Astronautics, Paper No. AIAA-97-1636-CP (1997), pp. 422–431.
[7]C. R. Fuller, C. H. Hansen, and S. D. Snyder, "Active control of structurally radiated noise using piezoceramic actuators," Inter-Noise **89**, 509–511 (1989).
[8]R. L. Clark and C. R. Fuller, "Experiments on active control of structurally radiated sound using multiple piezoceramic actuators," J. Acoust. Soc. Am. **91**(6), 3313–3320 (1992).
[9]C. R. Fuller, "Active control of sound transmission/radiation from elastic plates by vibration inputs: I analysis," J. Sound Vib. **136**(1), 1–3 (1990).
[10]A. Grewal, F. Nitzsche, D. G. Zimcik, and B. Leigh, "Active control of aircraft cabin noise using smart structures," AIAA/ASME/AHS Adaptive Structures Forum, Salt Lake City, UT, 18–19 April, Technical Papers A96-27071 06-39 (1996), pp. 91–100.
[11]R. L. Clark and K. D. Frampton, "Aeroelastic structural acoustic coupling: Implications on the control of turbulent boundary layer noise transmission," J. Acoust. Soc. Am. **102**(3), 1639–1647 (1997).
[12]K. D. Frampton and R. L. Clark, "Control of sound transmission through a convected fluid loaded plate with piezoelectric sensoriactuators," J. Intell. Mater. Syst. Struct. **8**(8), 686–696 (1997).
[13]K. Frampton and R. L. Clark, "Sound transmission through an aeroelastic plate into a cavity," AIAA J. **35**(7), 1113–1118 (1997).
[14]C. Maury, P. Gardonio, and S. J. Elliott, "Analysis of the boundary layer noise transmitted through aircraft sidewalls," 7th AIAA/CEAS Aeroacoustics Conference, Maastricht, Netherlands, 28–30 May, Paper No. AIAA-2001-2111 (2001).
[15]G. P. Gibbs, R. L. Clark, D. E. Cox, and J. S. Vipperman, "Radiation model expansion: Application to active structural acoustic control," J. Acoust. Soc. Am. **107**(1), 332–339 (1999).
[16]R. L. Clark, W. R. Saunders, and G. P. Gibbs, *Adaptive Structures, Dynamics and Control* (Wiley, New York, 1998).
[17]R. L. Clark, "Advanced sensing techniques for active structural acoustic control," Ph.D. thesis, Virginia Polytechnic Institute and State University, Blacksburg, Virginia, 1992.
[18]R. D. Blevins, *Formulas for Frequency and Mode Shape* (Krieger, Malabar, FL, 1987).
[19]L. Meirovitch, *Analytical Methods in Vibrations* (MacMillan, New York, 1967).
[20]A. Leissa, *Vibration of Plates* (The Acoustical Society of America, Melville, 1993).
[21]K. A. Cunefare, "Effect of modal interaction on sound radiation from vibrating structures," AIAA J. **30**(12), 2819–2828 (1992).
[22]S. J. Elliott and M. E. Johnson, "Radiation modes and the active control of sound power," J. Acoust. Soc. Am. **94**(4), 2194–2204 (1993).
[23]D. Cox, G. Gibbs, R. Clark, and J. Vipperman, "Experimental robust control of structural acoustic radiation," J. Vibr. Acoust. **121**, 433–439 (1999).
[24]R. E. Richard and R. L. Clark, "Genetic spatial optimization of active elements on an aeroelastic delta wing," in *ASME Winter Annual Meeting*, Orlando, FL, 5–10 November (ASME, Orlando, FL, 2000).
[25]R. L. Clark and D. E. Cox, "Band-limited actuator and sensor selection for disturbance rejection," J. Guid. Control Dyn. **22**(5), 740–743 (1999).
[26]K. B. Lim, "Disturbance rejection approach to actuator and sensor placement," J. Guid. Control Dyn. **20**(1), 202–204 (1997).
[27]W. Gawronski and K. B. Lim, "Balanced actuator and sensor placement for flexible structures," Int. J. Control **65**(1), 131–145 (1996).
[28]K. B. Lim and W. Gawronski, "Hankel singular values of flexible structures in discrete time," J. Guid. Control Dyn. **19**(6), 1370–1377 (1996).
[29]R. E. Richard, J. A. Rule, and R. L. Clark, "Genetic spatial optimization of active elements on an aeroelastic delta wing," J. Vibr. Acoust. **123**(4), 466–471 (2001).
[30]G. C. Smith and R. L. Clark, "Adaptive structure design through optimum spatial compensation," in *ACTIVE 99*, Fort Lauderdale, FL, 2–4 December (ACTIVE, Fort Lauderdale, FL, 1999).
[31]R. L. Clark and S. E. Burke, "Practical limitations in achieving shaped modal sensors with induced strain materials," J. Vibr. Acoust. **118**, 668–675 (1996).

An improved multimodal method for sound propagation in nonuniform lined ducts

WenPing Bi,[a] Vincent Pagneux, Denis Lafarge, and Yves Aurégan
Laboratoire d'Acoustique de I'Université du Maine, UMR CNRS 6613, Av. O Messiaen, 72085 Le Mans Cedex 9, France

(Received 6 December 2006; revised 6 April 2007; accepted 13 April 2007)

An efficient method is proposed for modeling time harmonic acoustic propagation in a nonuniform lined duct without flow. The lining impedance is axially segmented uniform, but varies circumferentially. The sound pressure is expanded in term of rigid duct modes and an additional function that carries the information about the impedance boundary. The rigid duct modes and the additional function are known *a priori* so that calculations of the true liner modes, which are difficult, are avoided. By matching the pressure and axial velocity at the interface between different uniform segments, scattering matrices are obtained for each individual segment; these are then combined to construct a global scattering matrix for multiple segments. The present method is an improvement of the multimodal propagation method, developed in a previous paper [Bi *et al.*, J. Sound Vib. **289**, 1091–1111 (2006)]. The radial rate of convergence is improved from $O(n^{-2})$, where n is the radial mode indices, to $O(n^{-4})$. It is numerically shown that using the present method, acoustic propagation in the nonuniform lined intake of an aeroengine can be calculated by a personal computer for dimensionless frequency K up to 80, approaching the third blade passing frequency of turbofan noise. © *2007 Acoustical Society of America.* [DOI: 10.1121/1.2736785]

PACS number(s): 43.50.Gf, 43.20.Mv, 43.20.Fn, 43.20.Hq [LLT] Pages: 280–290

I. INTRODUCTION

Numerous methods have been proposed to study sound propagation in ducts with locally reactive liners which are mathematically represented by an impedance boundary condition. In the presence of circumferential variations of the lining impedance (e.g., hard walled splices in lined intakes of aeroengine), the problem to solve is fully three dimensional. When dimensionless frequency K is high, where $K=kR$, $k=2\pi f/c$, f is frequency, c is sound velocity in air, and R is the radius of duct, it turns out to be challenging to model it efficiently.

The finite element method (FEM),[1,2] perturbation method,[3] and point matching method[4] have all been proposed to model the sound propagation in nonuniform lined ducts. An equivalent surface source method[5] and kinetic theory[6] were also employed to analyze the acoustic field in ducts lined with complicated distributions of impedance.

The hybrid analytical/numerical methods are always interesting, in which analysis is taken as far as possible. One of the hybrid analytical/numerical methods may be the mode matching method. Modes are first calculated in segmented uniform lined ducts and then matched between different uniform segments. When the lining impedance is circumferentially nonuniform, the sound pressure and particle velocity field cannot be separated in the $r-\theta$ plane, the dispersion relation in one segment cannot be written explicitly, and it is not possible to use classical root finding routines to determine the eigenmodes. Watson[7] used a hard walled duct mode expansion series to numerically evaluate the eigenmodes without flow. The Galerkin method was employed to force the series to satisfy the true boundary condition in the lined segment. The complex modal output amplitudes for a specified source distribution were then obtained by applying a mode matching technique at the discontinuity between rigid and lined segments. Fuller[8,9] expanded the circumferential admittance function as a Fourier series and the eigenmodes over the separable components adapted to the cylindrical coordinates. He obtained the eigenequation set without flow to be solved. The axial wave numbers in the lined segment were then calculated by solving this set of equations using the method of Muller. Campos *et al.*[10] studied acoustic modes in a cylindrical duct with an arbitrary wall impedance distribution with flow. When the wall impedance varies along the circumference, they calculate the acoustic modes in a similar way as Fuller.[8] Wright[11] used FEM to calculate the modes and then analytically matched them to the rigid duct modes. Wright *et al.*[12] also extended it to include uniform flow. Astley *et al.*[13] proposed a finite element mode matching method for propagation in lined ducts with flow in which the modes are calculated by FEM and numerically matched by a modified Galerkin method.

In Ref. 14, we proposed a multimodal propagation method (MPM) to study sound propagation in a nonuniform lined duct without flow. The sound pressure is expressed as a double series of the rigid duct modes which are known *a priori*. By matching the pressure and axial velocity at the interface between different segments, scattering matrices are obtained for each individual segment and then combined to construct a global scattering matrix for multiple segments. Different kinds of sources can be easily integrated without recalculating the scattering matrix. The full three-dimensional (3D) problem is reduced to a two-dimensional

[a]Electronic mail: wenping.bi@univ-lemans.fr

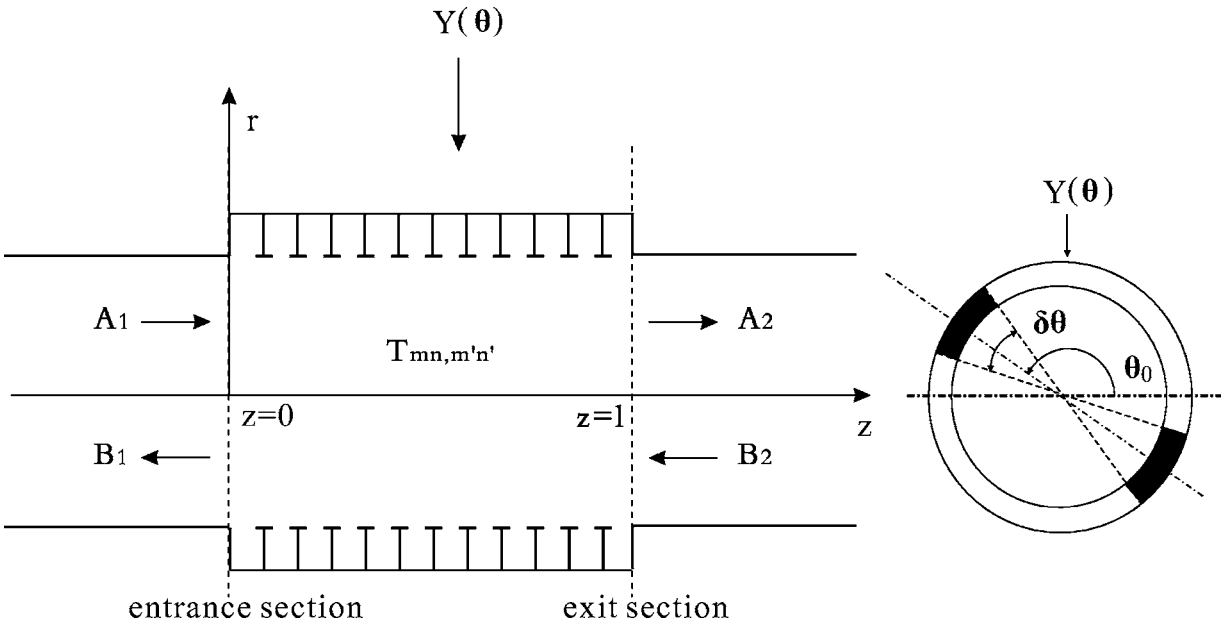


FIG. 1. Configuration of one axial segment nonuniform lined duct.

one which is better suited for the study of acoustic propagation in the lined intake of an aeroengine. The boundary condition is satisfied in the integral sense. Unlike the methods cited earlier, calculations of the eigenmodes of nonuniform lined ducts, which are very difficult, are avoided. This method is also extended to include uniform flow.[15] Because the individual rigid duct mode does not satisfy the impedance boundary condition, the radial convergence rate is only $O(n^{-2})$, where n is the radial mode indices.

In this paper, we improve the MPM[14] by accelerating the radial convergence rate. The sound pressure is expressed as a double series of the rigid duct modes and an additional function which carries the information of the impedance boundary. This is motivated by Ref. 17 which studies water waves over variable bathymetry regions, and Refs. 18 and 19 which study sound propagation in rigid waveguides with varying cross section. It is shown that the radial convergence rate of the double infinite series is improved from $O(n^{-2})$ to $O(n^{-4})$. This improvement can be extended to include uniform flow.[16]

The paper is organized as follows. In Sec. II, the derivation of equations are presented for circumferentially uniform and nonuniform boundary conditions, respectively. The convergence properties are then shown analytically and numerically in Sec. III. Finally, in Sec. IV, we present some numerical examples to show the robustness and capability of the method.

II. DERIVATION OF THE MULTIMODAL EQUATIONS

We consider an infinite rigid duct with circular cross section lined with a region of nonuniform liner. The liner properties are assumed to be given by a distribution of locally reacting impedance. Without significant loss of generality, the distribution may be assumed axially segmented, i.e., the impedance is set piecewise constant along the duct, while being arbitrarily variable along the circumference of each segment. In Fig. 1 the configuration of one axial segment of lining impedance is depicted, the circumferential variation of impedance is presented as two acoustically rigid splices, which is a typical configuration in the intake of an aeroengine. Linear and lossless sound propagation in air is assumed. With time dependence $\exp(j\omega t)$ omitted, the equation of mass conservation combined with the equation of state, and the equation of momentum conservation are written as

$$\nabla \cdot \mathbf{v} = -\frac{j\omega}{\rho_0 c_0^2} p, \tag{1}$$

$$j\omega \mathbf{v} = -\frac{1}{\rho_0} \nabla p, \tag{2}$$

where $\mathbf{v}$ is the particle velocity, p is the acoustic pressure, and ρ_0 and c_0 are the ambient density and speed of sound in air. Pressures, velocities, and lengths are, respectively, divided by $\rho_0 c_0^2$, c_0 and R (the duct radius) to reduce Eqs. (1) and (2) to the dimensionless form

$$\nabla \cdot \mathbf{v} = -jKp, \tag{3}$$

$$-jK\mathbf{v} = \nabla p, \tag{4}$$

where $K = \omega R / c_0$ is the dimensionless wave number. This yields the 3D wave equation

$$\nabla_\perp^2 p + \frac{\partial^2 p}{\partial z^2} + K^2 p = 0, \tag{5}$$

where

$$\nabla_\perp^2 = \frac{1}{r}\frac{\partial}{\partial r}\left(r\frac{\partial}{\partial r}\right) + \frac{1}{r^2}\frac{\partial^2}{\partial \theta^2}. \tag{6}$$

The radial boundary condition is

$$\frac{\partial p}{\partial r} = Y(\theta)p \quad \text{at } r = 1, \tag{7}$$

where $Y(\theta) = -jK\beta(\theta)$, and $\beta(\theta)$ is the liner admittance.

For the sake of clarity, we first consider a problem with a circumferentially uniform boundary condition. After this "warm up," the circumferentially nonuniform problem is investigated.

A. Circumferentially uniform impedance boundary condition

When the lining impedance is circumferentially uniform, the boundary condition is written as

$$\frac{\partial p}{\partial r} = Y_0 p \quad \text{at } r = 1, \tag{8}$$

where $Y_0 = -jK\beta_0$, and β_0 is the liner admittance and it is a complex constant.

In contrast to Ref. 14, the solution of Eq. (5) with boundary condition (8) is expressed as an infinite series and an additional function in order to satisfy the boundary condition

$$p(r,\theta,z) = \sum_{n=0}^{\infty} P_{mn}(z)\Psi_{mn}(r,\theta) + A_m(z)\chi_m(r)e^{-jm\theta}, \tag{9}$$

where P_{mn} are the expansion coefficients and m and n refer to azimuthal and radial mode indices, respectively. It is noted that because the lining impedance is circumferentially uniform, there is no coupling between azimuthal modes m, Eq. (9) involves only the coupling between radial modes n. The basis functions

$$\Psi_{mn} = \frac{1}{\sqrt{\pi\Lambda_{mn}}}\frac{J_m(\alpha_{mn}r)}{J_m(\alpha_{mn})}e^{-jm\theta} \tag{10}$$

are the eigenfunctions of the hard walled cylindrical circular duct which obey the transverse Laplacian eigenproblem

$$\left[\frac{1}{r}\frac{\partial}{\partial r}\left(r\frac{\partial}{\partial r}\right) - \frac{m^2}{r^2}\right]\Psi_{mn} = -\alpha_{mn}^2\Psi_{mn}, \tag{11}$$

with hard walled boundary condition

$$\frac{\partial \Psi_{mn}}{\partial r} = 0 \quad \text{at } r = 1, \tag{12}$$

and the orthogonality relation

$$\int \Psi_{mn}(r,\theta)\Psi^*_{m'n'}(r,\theta)dS = \delta_{m,m'}\delta_{n,n'}, \tag{13}$$

where the asterisk denotes the complex conjugate and δ denotes the Kronecker delta. The normalization coefficients Λ_{mn} are as follows:

$$\Lambda_{mn} = 1 - \frac{m^2}{\alpha_{mn}^2}. \tag{14}$$

In order to choose $A_m\chi_m$, Eq. (9) is substituted into the boundary condition (8) and using Eq. (12),

$$A_m(z)\frac{\partial \chi_m(r)}{\partial r}e^{-jm\theta} = Y_0\left[\sum_{n=0}^{\infty} P_{mn}(z)\Psi_{mn}(r,\theta) + A_m(z)\chi_m(r)e^{-jm\theta}\right] \quad \text{at } r = 1. \tag{15}$$

Function $\chi_m(r)$ can be chosen freely. From Eq. (15), it is shown that for deciding $A_m(z)$, two conditions can be imposed on $\chi_m(r)$,

$$\chi_m(r)|_{r=1} = 0, \tag{16a}$$

$$\left.\frac{d\chi_m(r)}{dr}\right|_{r=1} = 1. \tag{16b}$$

Substitution of the conditions (16) into Eq. (15) yields

$$A_m(z) = Y_0\sum_{n=0}^{\infty} P_{mn}(z)\Psi_{mn}(1,\theta)e^{jm\theta} = Y_0\sum_{n=0}^{\infty}\frac{P_{mn}(z)}{\sqrt{\pi\Lambda_{mn}}}. \tag{17}$$

Functions which satisfy the conditions (16) may be not unique. One choice may be

$$\chi_m(r) = B_m J_m(\beta_{m,0}r), \tag{18}$$

where B_m is constant, J_m is the m order first kind Bessel function, and $\beta_{m,0}$ refers to the roots of

$$J_m(\beta_{m,0}) = 0. \tag{19}$$

To obtain the constant B_m, substitution of Eq. (18) into Eq. (16b) yields

$$B_m = \frac{-1}{\beta_{m,0}J_{m+1}(\beta_{m,0})}. \tag{20}$$

Substitution of Eqs. (17), (18), and (20) into Eq. (9), yields

$$p(r,\theta,z) = \sum_{n=0}^{\infty} P_{mn}(z)\Psi_{mn}(r,\theta) + \sum_{n=0}^{\infty} P_{mn}(z)\frac{-Y_0}{\sqrt{\pi\Lambda_{mn}}}\frac{J_m(\beta_{m,0}r)e^{-jm\theta}}{\beta_{m,0}J_{m+1}(\beta_{m,0})}. \tag{21}$$

For calculating $P_{mn}(z)$, we project $p(r,\theta,z)$ on the basis Ψ_{mn}. Following the matricial terminology, Eq. (9) is written

$$p(r,\theta,z) = \mathbf{\Psi}^T\mathsf{M}\mathbf{P}, \tag{22}$$

where $\mathbf{P}$ and $\mathbf{\Psi}$ are column vectors, the superscript "T" indicates the transpose. M is a matrix, it is equal to

$$\mathsf{M} = \mathsf{I} + 2\pi Y_0 \mathsf{N}\mathbf{\Phi}^*\mathbf{\Phi}^T, \tag{23}$$

where I refers to the identity matrix, N is a diagonal matrix, its elements in the main diagonal are $1/(\alpha_{mn}^2 - \beta_{m,0}^2)$. They come from the projection of χ_m over the rigid mode eigenfunctions as shown in Appendix A. $\mathbf{\Phi}$ is a column vector, its elements are $1/\sqrt{\pi\Lambda_{mn}}$.

Using Eqs. (22) and (23), we project Eq. (5) to yield

$$\mathsf{M}\mathbf{P}'' + \mathsf{A}\mathbf{P} = 0, \tag{24}$$

where matrix A is

$$\mathsf{A} = (K^2\mathsf{I} - \mathsf{L})\mathsf{M} + 2\pi Y_0 \boldsymbol{\Psi}^*(1,\theta)\boldsymbol{\Psi}^T(1,\theta), \tag{25}$$

L is a diagonal matrix, its elements in the main diagonal are α_{mn}^2, and the double prime refers to the second derivative with respect to axial coordinate z. The projection of $\boldsymbol{\nabla}_\perp^2 p$ of Eq. (5) is shown in Appendix A.

It is noted that because the boundary condition is circumferentially uniform, there is no coupling between azimuthal modes. The indices m and n of the above-mentioned vectors and matrices are $m=m_0$ and $0 \leqslant n < \infty$.

B. Circumferentially nonuniform impedance boundary condition

When the boundary condition is circumferentially nonuniform as in Eq. (7), we have to solve a full 3D problem. The sound pressure cannot be separated in the $r-\theta$ plane. We do not succeed in finding a function to exactly satisfy the boundary condition (7). On the other hand, a function is found to satisfy the nonuniform boundary condition in the sense of $\partial p/\partial r = (\Sigma_m Y_m e^{-jm\theta})p$, where Y_m is the Fourier transformation coefficients of $Y(\theta)$. Similar to Eq. (9), the sound pressure is expressed as

$$p(r,\theta,z) = \sum_{m=-\infty}^{\infty}\sum_{n=0}^{\infty} P_{mn}(z)\Psi_{mn}(r,\theta) + \sum_{m=-\infty}^{\infty} A_m(z)\chi_m(r)e^{-jm\theta}. \tag{26}$$

As in Sec. II A, substitution of Eq. (26) into the boundary condition (7) yields

$$A_m(z) = \frac{1}{2\pi}\int_0^{2\pi} Y(\theta)e^{jm\theta}d\theta \sum_{m'=-\infty}^{\infty}\sum_{n'=0}^{\infty} P_{m'n'}(z)\Psi_{m'n'}(1,\theta) = \sum_{m'=-\infty}^{\infty}\frac{1}{2\pi}\int_0^{2\pi} Y(\theta)e^{-j(m'-m)\theta}d\theta \sum_{n'=0}^{\infty}\frac{P_{m'n'}(z)}{\sqrt{\pi\Lambda_{m'n'}}}, \tag{27}$$

where we have used Eq. (10) and imposed the conditions (16). Function χ_m, satisfying conditions (16), is the same as in Eqs. (18) and (20).

Following the matricial terminology, Eq. (26) is written

$$p(r,\theta,z) = \boldsymbol{\Psi}^T\mathsf{M}\mathbf{P}, \tag{28}$$

where M is

$$\mathsf{M} = \mathsf{I} + \mathsf{N}\mathbf{Y}\boldsymbol{\Phi}^*\boldsymbol{\Phi}^T, \tag{29}$$

N and $\boldsymbol{\Phi}$ are the same as in Eq. (23), respectively, and the elements of the matrix $\mathbf{Y}$ are $\int_0^{2\pi} Y(\theta)e^{-j(m'-m)\theta}d\theta$.

Using Eqs. (28) and (29), we project Eq. (5) to yield

$$\mathsf{M}\mathbf{P}'' + \mathsf{A}\mathbf{P} = 0, \tag{30}$$

where matrix A is

$$\mathsf{A} = (K^2\mathsf{I} - \mathsf{L})\mathsf{M} + \int_0^{2\pi} Y(\theta)\boldsymbol{\Psi}^*(1,\theta)\boldsymbol{\Psi}^T(1,\theta)d\theta, \tag{31}$$

where L is the same as in Eq. (25), a diagonal matrix, its elements in the main diagonal are α_{mn}^2, the double prime refers to the second derivative with respect to axial coordinate z. The projection of $\boldsymbol{\nabla}_\perp^2 p$ of Eq. (5) is shown in Appendix A.

It is noted that because the boundary condition is circumferentially nonuniform, modes are coupled between azimuthal orders. The indices m and n of the above-mentioned vectors and matrices are $-\infty < m < \infty$ and $0 \leqslant n < \infty$.

Equations (24) and (30) are constant coefficient matrix differential equations when the axial lining impedance is uniform in one segment. Their solutions can be directly written as

$$\mathbf{P} = \mathsf{X}\mathsf{D}(z)\mathbf{C_1} + \mathsf{X}\mathsf{D}(l-z)\mathbf{C_2}, \tag{32}$$

where $\mathbf{C_1}$ and $\mathbf{C_2}$ are amplitude vectors of dimension N_t ($M \times N$, where M and N refer to the truncated dimensions of mode indices m and n), X is the $N_t \times N_t$ matrix whose columns are the generalized eigenvectors $\mathbf{X}_n$ of matrix $\mathsf{M}^{-1}\mathsf{A}$, and $\mathsf{D}(z)$ and $\mathsf{D}(l-z)$ are diagonal matrices with $\exp(-j\nu_n z)$ and $\exp(-j\nu_n(l-z))$, respectively, on the main diagonal, with $\nu_n = \sqrt{d_n}$, d_n being the generalized eigenvalues of matrix $\mathsf{M}^{-1}\mathsf{A}$. In the form of Eq. (32), numerical stability is ensured because the propagation matrices $\mathsf{D}(z)$ and $\mathsf{D}(l-z)$ have only positive arguments and contain no exponentially diverging terms due to the evanescent modes. By matching the pressure and axial velocity at the interfaces of the segment, the coefficients of transmission and reflection are yielded. Scattering matrices are then obtained for each individual segment; these are combined to construct a global scattering matrix for multiple segments. This procedure is the same as in Ref. 14 and outlined in detail in Appendix B.

III. CONVERGENCE ANALYSIS

In Ref. 14, the sound pressure is expressed as a double series of the rigid duct modes, which are known *a priori*. It is numerically shown that the convergence rates are $O(n^{-2})$ when m is fixed, and $O(m^{-3})$ when n is fixed. The convergence rate for n is slow because the rigid duct modes do not satisfy individually the impedance boundary condition. This slow convergence rate is improved in this paper by adding a function which satisfies the boundary condition. In this section, the behaviors of the convergence rate are shown analytically and numerically.

In general, the MPM[14] and the method presented in this paper are the generalized Fourier series method. Their convergence rate can be estimated by the divergence theorem or integration by parts. Let us take a function g on the segment $[0,1]$ with $g'(0)=0$ and $g'(1)=a$ and with its second derivative g'' integrable. If we project this function on the Neumann basis $u_n(x) = \sqrt{2-\delta_{n0}}\cos(n\pi x)$ with $u_n'(0) = u_n'(1) = 0$, then $g(x) = \Sigma_{n=0}^{n=+\infty} G_n u_n(x)$ and by integration by parts

$$G_n = \int_0^1 g u_n dx = -\frac{1}{n^2\pi^2}\left[-a u_n(1) + \int_0^1 g'' u_n dx\right]. \tag{33}$$

Since g'' is integrable, by the Riemann-Lebesgue lemma we know that $\lim_{n\to\infty}\int_0^1 g'' u_n dx = 0$. Consequently

$$G_n = \frac{\sqrt{2}a(-1)^{n+1}}{\pi^2 n^2} + o\left(\frac{1}{n^2}\right), \tag{34}$$

so that the leading term for G_n is given by the derivative of g at the boundary.

Consider sound pressure $p(r,\theta,z)$ to satisfy Helmholtz equation (5) with impedance boundary condition (7) in an infinite lined duct. Because we are interested in the radial convergence rate, without loss of generality, the axial lining impedance can be assumed as uniform. In the following, we assume also that the $p(r,\theta,z)$ is sufficient differential. Equations (9) and (26) can be expressed in a generalized Fourier series

$$p(r,\theta,z) = \sum_i P_i(z)\Phi_i(r,\theta), \tag{35}$$

where Φ_i is any set of functions which are complete and orthogonal. If the basis functions Φ_i satisfy the transverse Laplace eigenproblem

$$\nabla_\perp^2 \Phi_i = -\gamma_i^2 \Phi_i, \tag{36}$$

using the divergence theorem, the expansion coefficients P_i can be written as

$$P_i = \int p\Phi_i dS = \frac{-1}{\gamma_i^2}\int \nabla_\perp^2 p\Phi_i dS + \frac{1}{\gamma_i^2} \times \oint \left(\Phi_i \frac{\partial p}{\partial r} - p\frac{\partial \Phi_i}{\partial r}\right) dC, \tag{37}$$

where γ_i are the transverse Helmholtz wave numbers corresponding to Φ_i. When the basis functions Φ_i satisfy the boundary condition

$$\frac{\partial \Phi_i}{\partial r} = Y(\theta)\Phi_i \quad \text{at } r = 1, \tag{38}$$

individually, the second term in Eq. (37) is equal to zero. This is the so-called eigenfunction expansion, i.e., Φ_i and γ_i are the eigenfunctions and eigenvalues of the Helmholtz equation with corresponding impedance boundary conditions in the nonuniform lined ducts. Repeated use of the divergence theorem and Helmholtz equation (5) leads to an exponential convergence rate of P_i.

When the basis functions Φ_i are not the eigenfunctions of lined ducts, e.g., $\Phi_i = \Psi_{mn}$, where Ψ_{mn} are the rigid duct eigenfunctions, which do not satisfy the boundary condition (38) individually, the convergence rate will be slow. In this case, the expansion coefficients P_{mn} are

$$\begin{aligned} P_{mn} &= \frac{-1}{\alpha_{mn}^2}\int \nabla_\perp^2 p\Psi_{mn}^* dS + \frac{1}{\alpha_{mn}^2} \times \oint \left(\Psi_{mn}^* \frac{\partial p}{\partial r} - p\frac{\partial \Psi_{mn}^*}{\partial r}\right) dC \\ &= \frac{-1}{\alpha_{mn}^2}\int \nabla_\perp^2 p\Psi_{mn}^* dS + \frac{1}{\alpha_{mn}^2}\sum_{m'n'} \frac{P_{m'n'}}{\pi\sqrt{\Lambda_{m'n'}\Lambda_{mn}}} \\ &\quad \times \int_0^{2\pi} Y(\theta) e^{-j(m'-m)\theta} d\theta, \end{aligned} \tag{39}$$

where α_{mn} are the eigenvalues of a rigid duct as mentioned earlier. The divergence theorem can be used a second time to show that the first term is $o(1/\alpha_{mn}^2)$. For the second term, it is evident that at fixed m, $\Sigma_{m'n'} P_{m'n'}/(\pi\sqrt{\Lambda_{m'n'}\Lambda_{mn}})\int_0^{2\pi} Y e^{-j(m'-m)\theta} d\theta$ does not affect the radial convergence rate of P_{mn}. Hence, at fixed m, Eq. (39) yields

$$P_{mn} = O\left(\frac{1}{\alpha_{mn}^2}\right). \tag{40}$$

The asymptotic forms of α_{mn} at fixed m and $n > N_0$, where N_0 is a sufficiently large constant, is[21]

$$\alpha_{mn} = \left(n + \frac{2m-3}{4}\right)\pi + O\left(\frac{1}{n}\right). \tag{41}$$

We therefore obtain the rates of convergence of P_{mn} with respect to index n at fixed m,

$$P_{mn} = O\left(\frac{1}{\left(n + \frac{m}{2}\right)^2}\right), \quad n > N_0. \tag{42}$$

For the method presented in this paper, an additional function is involved in the expression of $p(r,\theta,z)$, the P_{mn} is

$$\begin{aligned} P_{mn} &= \int p\Psi_{mn}^* dS - \int \sum_{m'=-\infty}^{\infty} A_{m'}\chi_{m'} e^{-jm'\theta}\Psi_{mn}^* dS \\ &= \frac{-1}{\alpha_{mn}^2}\int \nabla_\perp^2 p\Psi_{mn}^* dS + \frac{1}{\alpha_{mn}^2}\oint \left(\Psi_{mn}^* \frac{\partial p}{\partial r} - p\frac{\partial \Psi_{mn}^*}{\partial r}\right) dC \\ &\quad - \frac{1}{\alpha_{mn}^2 - \beta_{m,0}^2}\sum_{m'n'} \frac{P_{m'n'}}{\pi\sqrt{\Lambda_{m'n'}\Lambda_{mn}}}\int_0^{2\pi} Y(\theta) e^{-j(m'-m)\theta} d\theta, \end{aligned} \tag{43}$$

where we have used Eq. (A1). Using the divergence theorem a second time for the first term we find that the first term is at least $O(1/\alpha_{mn}^4)$. Equation (43) is written as

$$\begin{aligned} P_{mn} &= O\left(\frac{1}{\alpha_{mn}^4}\right) + \frac{1}{\alpha_{mn}^2}\sum_{m'n'} \frac{P_{m'n'}}{\pi\sqrt{\Lambda_{m'n'}\Lambda_{mn}}} \\ &\quad \times \int_0^{2\pi} Y(\theta) e^{-j(m'-m)\theta} d\theta - \frac{1}{\alpha_{mn}^2} \\ &\quad \times \left[1 + \frac{\beta_{m,0}^2}{\alpha_{mn}^2} + O\left(\frac{\beta_{m,0}^2}{\alpha_{mn}^4}\right)\right] \\ &\quad \times \sum_{m'n'} \frac{P_{m'n'}}{\pi\sqrt{\Lambda_{m'n'}\Lambda_{mn}}}\int_0^{2\pi} Y(\theta) e^{-j(m'-m)\theta} d\theta \\ &= O\left(\frac{1}{\alpha_{mn}^4}\right) \end{aligned}$$

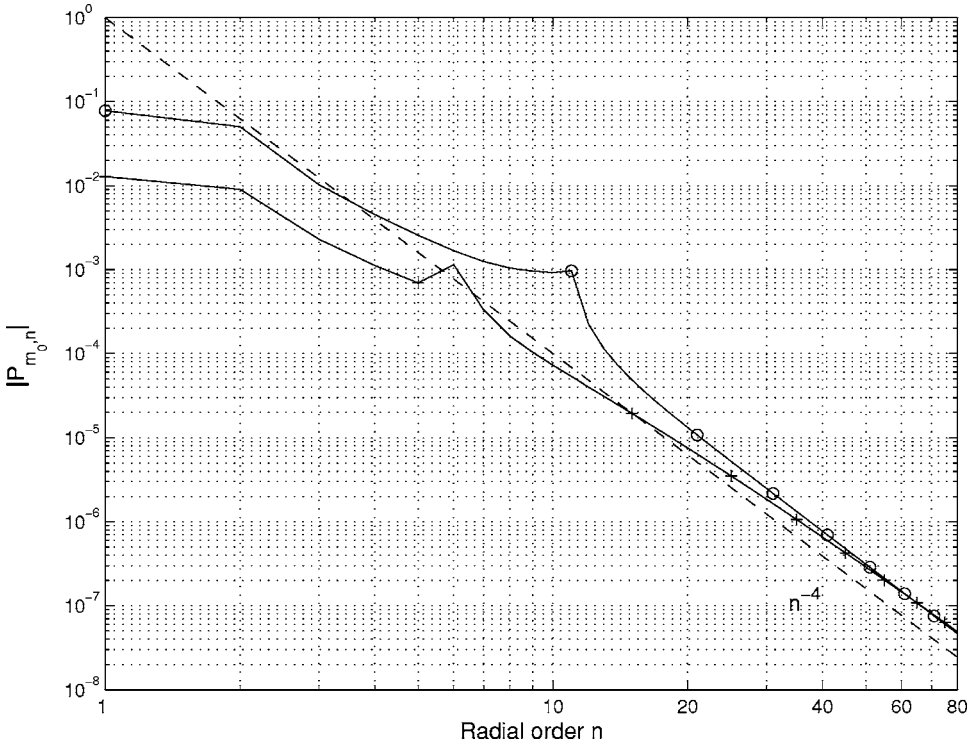


FIG. 2. Convergence rate of radial order n for fixed $m=m_0$. Solid line with circles refers to $m_0=0$, solid line with plus refers to $m_0=10$, dashed line refers to n^{-4}. Truncation dimension is $M=30$, $N=80$. $K=31.26$, $Z/\rho c=2-j$, two splices, $\delta\theta=0.06$, mode (2, 0) is incident.

$$+ O\left(\frac{\beta_{m,0}^2}{\alpha_{mn}^4}\right) \sum_{m'n'} \frac{P_{m'n'}}{\pi\sqrt{\Lambda_{m'n'}\Lambda_{mn}}} \int_0^{2\pi} Y(\theta) e^{-j(m'-m)\theta} d\theta. \tag{44}$$

When m is fixed, using Eq. (41), Eq. (44) is written as

$$p_{mn} = O\left(\frac{1}{\left(n+\frac{m}{2}\right)^4}\right), \quad n > N_0. \tag{45}$$

This convergence rate is valid for both circumferentially uniform and circumferentially nonuniform boundary conditions.

From Eq. (45), it is shown that

(1) The radial convergence rate of the present method is $O(n^{-4})$ as $n>N_0$, where N_0 is a sufficiently large constant, when m is fixed.
(2) N_0 depends on the azimuthal order m. For a small m, N_0 is small, P_{mn} converges as $O(n^{-4})$ after a few terms of n. For a large m, however, P_{mn} converges as $O(n^{-4})$ after a large n.
(3) It is important to note that the behavior of convergence rate Eq. (45) is independent of the variation of admittance $Y(\theta)$. It means that whether Y is constant, i.e., no circumferential mode scattering, or Y varies circumferentially, i.e., there is circumferential mode scattering, the convergence behavior does not change for fixed m.

Now, the radial convergence properties are numerically shown in Figs. 2–4 for fixed m. The configuration is the same as in Fig. 1. An infinite rigid duct is lined with one axial segment impedance with two acoustically rigid splices distributed oppositely. The splice angles are 0.06 rad, dimensionless frequency is $K=31.26$, lining impedance is $Z/\rho c=2-j$, mode ($m=2$, $n=0$) is incident. Because of the two rigid splices, the incident mode (2,0) is scattered into different m and n modes. In Fig. 2, the radial convergence is plotted for $m=0$ and $m=10$. The truncated dimension is $M=30$, $N=80$. A reference curve n^{-4} is also plotted in Fig. 2. It is shown that the convergence rates turn to $O(n^{-4})$ after $n>40$. The effects of azimuthal order m on the transient region are shown in Figs. 3 and 4. For $m=26$, until $n=80$, P_{mn} are still in the transient region. The convergence rate is between $O(n^{-3})$ and $O(n^{-4})$. The same behavior takes place for $m=58$ and $m=98$ as shown in Fig. 4. The convergence rates are about $O(n^{-2})$ when $n<20$ and between $O(n^{-2})$ and $O(n^{-3})$ when $20<n<40$. Note that in Figs. 3 and 4, only the transient regions are plotted because of the limitation of the PC memory. Following the above-mentioned statement (3), the radial asymptotic convergence rate can be seen in Fig. 5 in which circumferential admittance is uniform $Y(\theta)=Y_0$ (no splice). The other parameters are the same as in Figs. 3 and 4 and noted in Fig. 5. In this example, incident modes ($m=2$, $n=0$) and ($m=98$, $n=0$) are only scattered in corresponding radial modes of $m=2$ and $m=98$, respectively. The truncated dimension is $N=1000$. It is clearly shown that for incident mode (2,0), the convergence rate is $O(n^{-4})$, when $n>10$. However, the rate $O(n^{-4})$ takes place after $n>200$ for incident mode (98,0). In the transient region $n<200$, the convergence rate is slower, about $O(n^{-3})$ in the region $20<n<200$, and $O(n^{-2})$ in the region $n<20$.

It is important to point out that in Figs. 2 and 4, although the transient region is longer for the modes with $m=98$ than for the modes with $m=0$, and in the transient region the convergence rate is slower, the amplitudes $|P_{98,n}|$ are already much smaller than $|P_{0,n}|$.

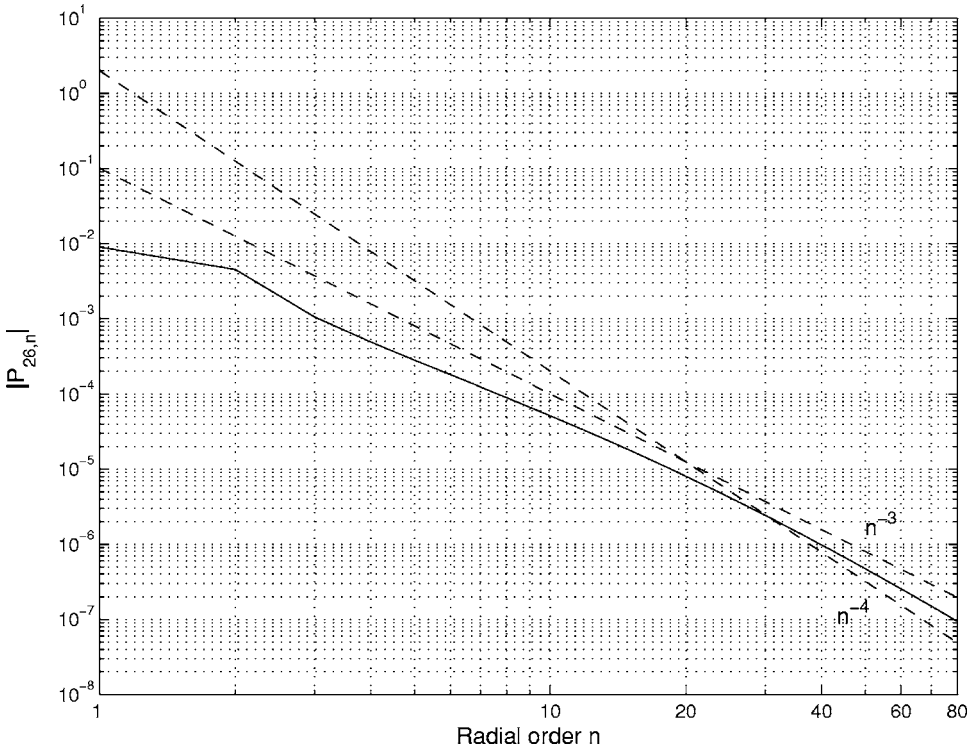


FIG. 3. Convergence rate of radial order n for $m=m_0=26$. Truncation dimension is $M=30$, $N=80$. $K=31.26$, $Z/\rho c=2-j$, two splices, $\delta\theta=0.06$, mode (2, 0) is incident.

IV. NUMERICAL EXAMPLES

In this section, some examples corresponding to the problem of Fig. 1 are presented to show the capability of the method. In Fig. 6, $\text{abs}(W_N^o - W_{1000}^o)/W_{1000}^o$ versus different truncations N is shown, where W_N^o refers to the output sound power at the exit plane (Fig. 1) for truncation N. W_{1000}^o is the converged value. The lining admittance is circumferentially uniform, i.e., $Y(\theta)=Y_0$. The parameters are noted in Fig. 6. Incident amplitudes are equal to 1. For an incident mode $(m=2, n=0)$ and with the truncation $N=10$, relative error 1% is obtained compared with truncation $N=1000$. For incident mode $(m=98, n=0)$, when the truncation $N=25$, relative error 1% is obtained compared with truncation $N=1000$. Another example is shown in Fig. 7. The lining impedance is $Z/\rho c=j(0.01K/R-\cot(0.016K/R))$, where $K=10, R=0.2$. When mode $(m=0, n=0)$ is incident, surface modes are invoked. Such surface modes are located near the duct wall and exponentially decay away from the duct wall.[20] This is

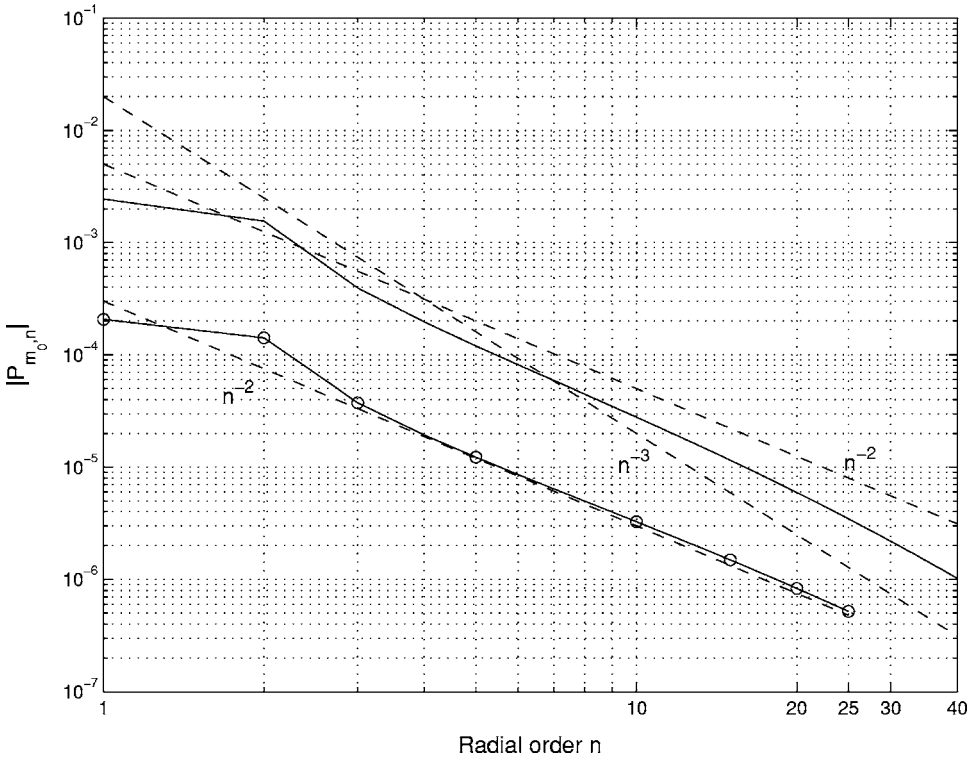


FIG. 4. Convergence rates of radial order n for fixed $m=m_0$. Solid line refers to $m_0=58$, truncation dimension $M=60$, $N=40$. Solid line with circles refers to $m_0=98$, truncation dimension $M=100$, $N=25$. $K=31.26$, $Z/\rho c=2-j$, two splices, $\delta\theta=0.06$, mode (2, 0) is incident.

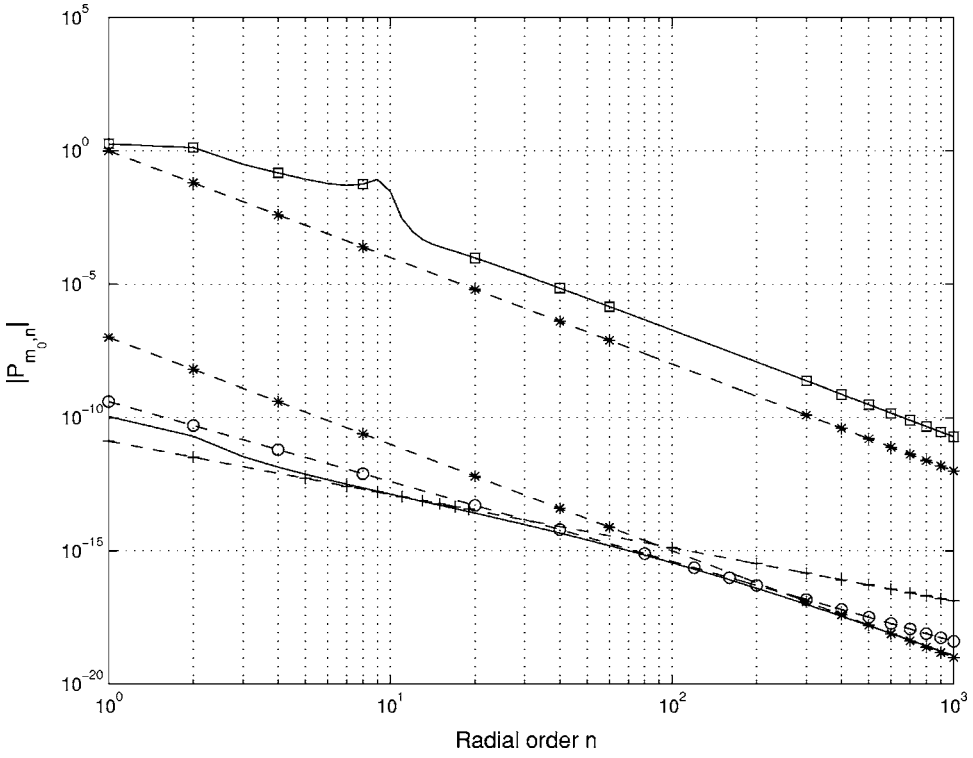


FIG. 5. Convergence rates of radial order n for different incident modes $(m_0,0)$. Solid line refers to $m_0=98$, solid line with squares refers to $m_0=2$, dashed line with stars refers to n^{-4}, dashed line with circles refers to n^{-3}, and dashed line with plus refers to n^{-2}. $K=31.26$, $Z/\rho c=2-j$, no splice.

the most difficult case for the MPM[14] and the improved method presented in this paper. For MPM,[14] only rigid duct modes are used to express this surface mode, about $N=200$ rigid duct modes are needed to converge to the surface mode. However, only $N=30$ terms are needed for the improved method in this paper to converge to the surface mode.

A full 3D example with high dimensional frequency K is shown in Figs. 8 and 9. The configuration is the same as in Fig. 1. An infinite rigid duct is lined with one axial segment impedance with two rigid splices distributed oppositely. The splice angles are 0.06 rad. The dimensionless frequency is $K=80$, the lining impedance is $Z/\rho c=2+j$. The parameters are typical of aeroengine intakes. The dimensionless frequency $K=80$ approximately corresponds to the third BPF of turbofan noise. However they do not exactly correspond to any true aeroengine. Due to the lined segment, the incident mode is scattered to different azimuthal and radial modes. The sound power of every rigid mode at the exit plane is

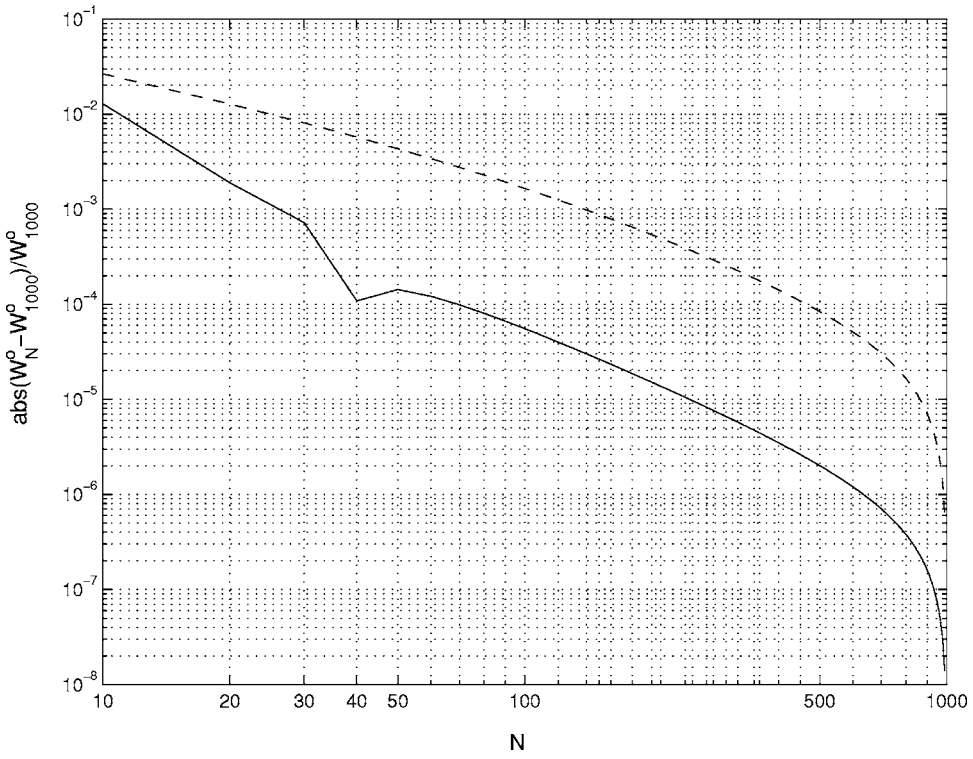


FIG. 6. $\text{abs}(W^o_N-W^o_{1000})/W^o_{1000}$ vs different truncation N. Solid line refers to incident mode $(m=2,\ n=0)$, dashed line refers to incident mode $(m=98,\ n=0)$. $K=110$, $Z/\rho c=2+j$, $L/R=0.48$, no splice.

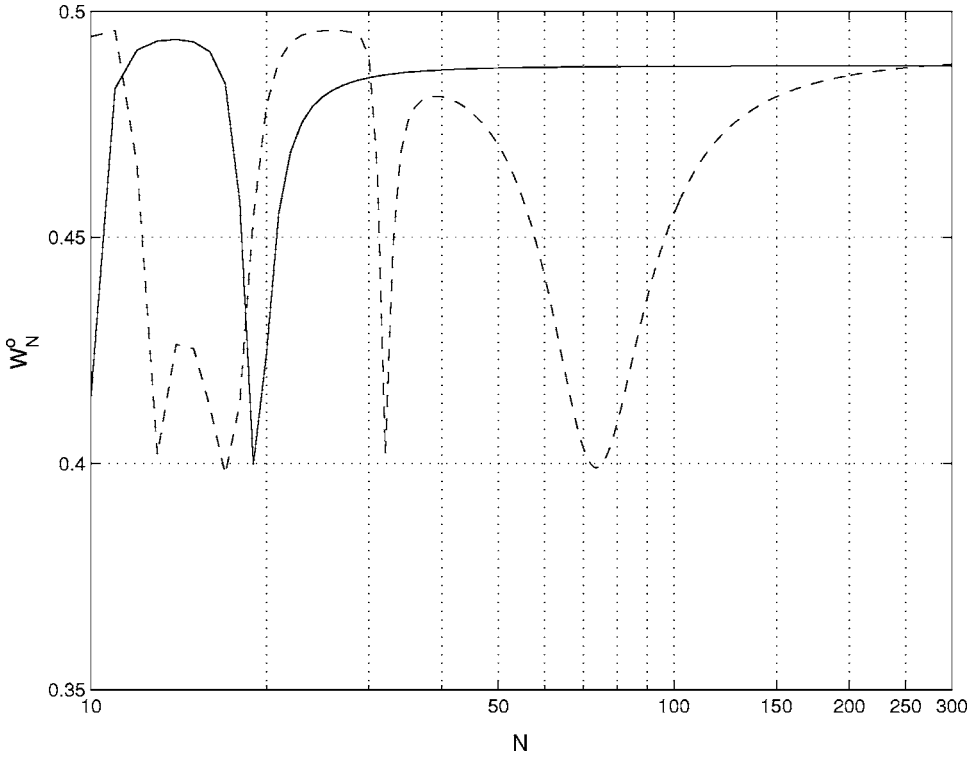


FIG. 7. W_N^o vs N for different methods. Solid line refers to the improved method, dashed line refers to the MPM, $K=10$, $R=0.2$, $Z/\rho c = j(0.01K/R - \cot(0.016K/R))$, $L/R = 2.5$, mode (0, 0) is incident, $|P_{0,0}| = 1$, no splice. There exist surface waves, which is the most difficult case for the MPM or the improved method.

shown in the figures. This example may be difficult for purely numerical methods, e.g., FEM. For the present method, it takes hours on a personal computer (about 2 h on a PC with processor P IV 2.4 GHz, 512M physical RAM and 3G virtual memory, without optimizing the MATLAB code). In Fig. 8, mode ($m=2$, $n=0$) is incident. The scattered modes are distributed nearly symmetrically for $+m$ and $-m$. In Fig. 9, mode ($m=76$, $n=0$), the last propagating mode, is incident. The incident mode is scattered to all modes with even m that have lower azimuthal orders. The performance of the liner will be reduced by the presence of rigid splices.

V. CONCLUSIONS

An efficient method is developed to model acoustic propagation in a nonuniform lined duct. The sound pressure is expressed as a double series of rigid duct modes and an additional function carrying the information about the impedance boundary, which are known *a priori*. Calculations of the eigenmodes of a lined duct, which are difficult, are avoided. The radial convergence rate is accelerated from $O(n^{-2})$ to $O(n^{-4})$ for fixed m, where n and m are the radial and azimuthal indices, respectively. Numerical examples show that this method can deal with a full 3D problem with high dimensionless frequency K, e.g., $K=80$ corresponding to the third BPF in an aeroengine.

APPENDIX A: PROJECTIONS OF P AND $\nabla_\perp^2 P$

Using Eqs. (26), (27), (18), and (20), the projection of p over the basis $\mathbf{\Psi}^*$ for circumferentially nonuniform boundary condition is

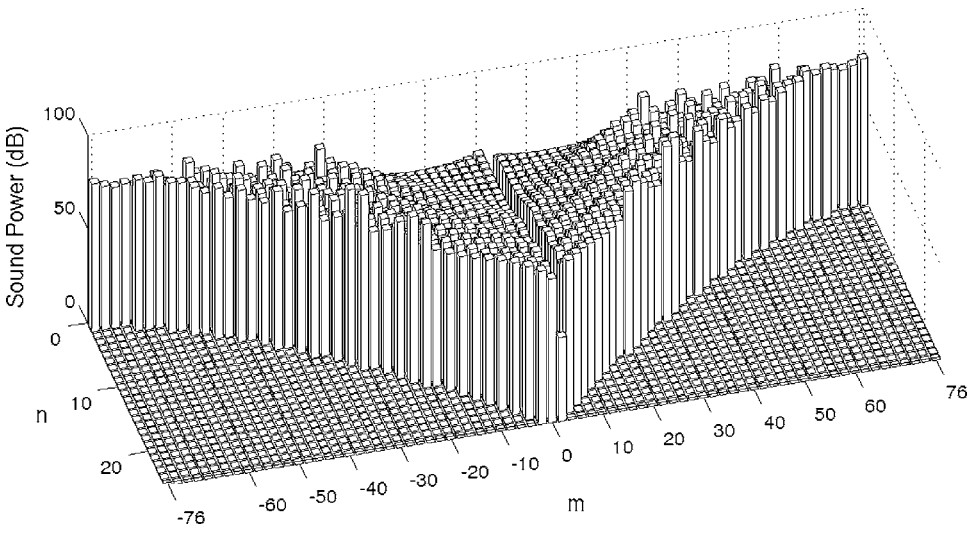


FIG. 8. Output sound power of rigid modes at the exit section. $K=80$, $Z/\rho c = 2+j$, $L/R=0.48$, mode ($m=2$, $n=0$) is incident, $|P_{2,0}|=1$, two splices with angle 0.06 rad. Only modes with even m are excited and shown.

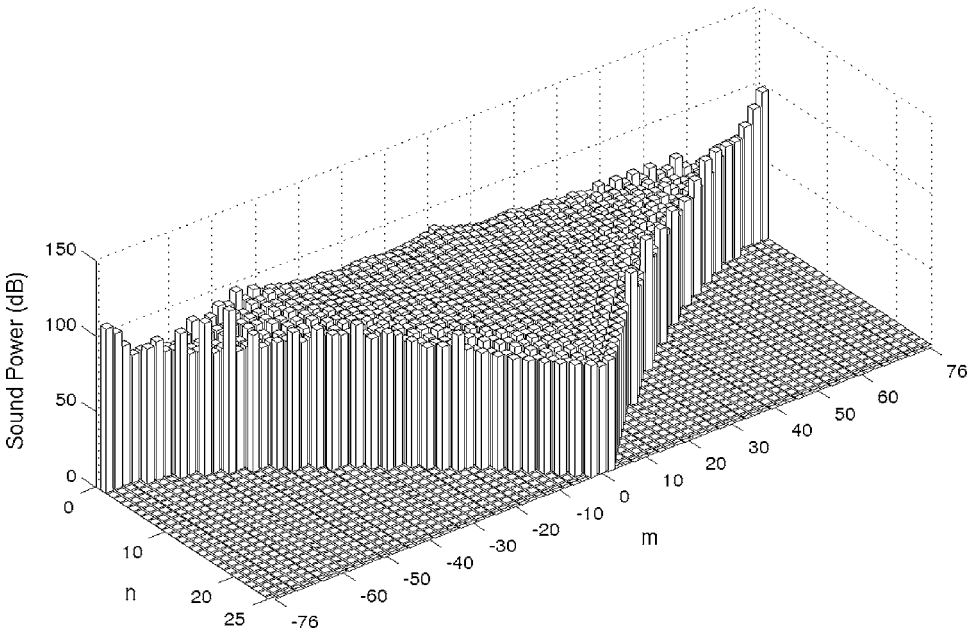


FIG. 9. Output sound power of rigid modes at the exit section. $K=80$, $Z/\rho c=2+j$, $L/R=0.48$, mode (76, 0) is incident, $|P_{76,0}|=1$, two splices with angle 0.06 rad. Only modes with even m are excited and shown.

$$\sum_{m'n'} M_{mnm'n'}P_{m'n'} = \int p(r,\theta,z)\Psi_{mn}^* dS = \int \left(\sum_{m'n'} P_{m'n'}\Psi_{m'n'} + \sum_{m'} A_{m'}\chi_{m'}e^{-jm'\theta}\right)\Psi_{mn}^* dS = P_{mn}$$

$$+ \frac{1}{\sqrt{\pi\Lambda_{mn}}}\frac{1}{J_m(\alpha_{mn})}\int_0^{2\pi}\sum_{m'} A_{m'}e^{-jm'\theta}e^{jm\theta}d\theta\int_0^1 \chi_{m'}J_m(\alpha_{mn}r)rdr$$

$$= P_{mn} + \frac{1}{\sqrt{\pi\Lambda_{mn}}}\frac{1}{J_m(\alpha_{mn})}2\pi A_m B_m\int_0^1 J_m(\beta_{m,0}r)J_m(\alpha_{mn}r)rdr = P_{mn}$$

$$+ \frac{1}{\alpha_{mn}^2-\beta_{m,0}^2}\sum_{m'}\int_0^{2\pi} Y(\theta)e^{-j(m'-m)\theta}d\theta\frac{1}{\sqrt{\pi\Lambda_{mn}}}\sum_{n'}\frac{P_{m'n'}}{\sqrt{\pi\Lambda_{m'n'}}}$$

$$= \sum_{m'n'}\left(\delta_{mm'}\delta_{nn'} + \frac{1}{\alpha_{mn}^2-\beta_{m,0}^2}\frac{\int_0^{2\pi} Y(\theta)e^{-j(m'-m)\theta}d\theta}{\pi\sqrt{\Lambda_{mn}\Lambda_{m'n'}}}\right)P_{m'n'}, \tag{A1}$$

where we have used the relation[21]

$$\int_0^x tJ_\mu(at)J_\mu(bt)dt = \frac{x}{a^2-b^2}\left[J_\mu(ax)\frac{dJ_\mu(bx)}{dx} - J_\mu(bx)\frac{dJ_\mu(ax)}{dx}\right].$$

When the boundary condition is circumferentially uniform, i.e., $Y(\theta)=Y_0$, there is no coupling between azimuthal modes. For every azimuthal mode m, Eq. (A1) is written as

$$\sum_{n'} M_{mnn'}P_{mn'} = \sum_{n'}\left(\delta_{nn'} + \frac{1}{\alpha_{mn}^2-\beta_{m,0}^2}\times\frac{2\pi Y_0}{\pi\sqrt{\Lambda_{mn}\Lambda_{mn'}}}\right)P_{mn'}. \tag{A2}$$

The projection of $\nabla_\perp^2 p$ is

$$\int \nabla_\perp^2 p\Psi_{mn}^* dS = \int p\nabla_\perp^2\Psi_{mn}^* dS + \oint\left(\Psi_{mn}^*\frac{\partial p}{\partial r} - p\frac{\partial\Psi_{mn}^*}{\partial r}\right)dC$$

$$= \int p(-\alpha_{mn}^2)\Psi_{mn}^* dS + \oint \Psi_{mn}^*(1,\theta)Y(\theta)pdC$$

$$= -\alpha_{mn}^2\sum_{m'n'} M_{mn,m'n'}P_{m'n'} + \sum_{m'n'}\int_0^{2\pi} Y(\theta)\Psi_{mn}^*(1,\theta)\times\Psi_{m'n'}(1,\theta)d\theta P_{m'n'}. \tag{A3}$$

APPENDIX B: CALCULATIONS OF T AND R

The reflection and transmission matrices are easily identified by writing the general field solution in the lined section, and the continuity conditions at the interfaces.

Using the solutions (32) in the lining segment and the continuity of pressure and axial velocity leads to

$$\mathbf{A_1} + \mathbf{B_1} = \mathsf{MX}(\mathbf{C_1} + \mathrm{D}_l\mathbf{C_2}),$$

$$\mathsf{K}_0(\mathbf{A_1} - \mathbf{B_1}) = \mathsf{MXK}_\mathsf{Y}(\mathbf{C_1} - \mathrm{D}_l\mathbf{C_2}),$$

$$\mathbf{A_2} + \mathbf{B_2} = \mathsf{MX}(\mathrm{D}_l\mathbf{C_1} + \mathbf{C_2}),$$

$$\mathsf{K}_0(\mathbf{A_2} - \mathbf{B_2}) = \mathsf{MXK}_\mathsf{Y}(\mathrm{D}_l\mathbf{C_1} - \mathbf{C_2}), \tag{B1}$$

where $\mathrm{D}_l = \mathrm{D}(l)$, K_0 and K_Y are diagonal matrices with the axial wave numbers on the main diagonal in the rigid and lined sections (respectively, the $K_{0z,mn}$ and d_n). $\mathbf{A_1}$, $\mathbf{B_1}$, $\mathbf{A_2}$, and $\mathbf{B_2}$ are the modal amplitudes in the rigid duct respectively as shown in Fig. 1.

By denoting

$$\mathsf{F} = \mathsf{MX} + \mathsf{K}_0^{-1}\mathsf{MXK}_\mathsf{Y}, \tag{B2}$$

$$\mathsf{G} = \mathsf{MX} - \mathsf{K}_0^{-1}\mathsf{MXK}_\mathsf{Y}, \tag{B3}$$

Eq. (B1) is reduced to

$$2\mathbf{A_1} = \mathsf{F}\mathbf{C_1} + \mathsf{G}\mathrm{D}_l\mathbf{C_2},$$

$$2\mathbf{B_1} = \mathsf{G}\mathbf{C_1} + \mathsf{F}\mathrm{D}_l\mathbf{C_2},$$

$$2\mathbf{A_2} = \mathsf{F}\mathrm{D}_l\mathbf{C_1} + \mathsf{G}\mathbf{C_2},$$

$$2\mathbf{B_2} = \mathsf{G}\mathrm{D}_l\mathbf{C_1} + \mathsf{F}\mathbf{C_2}, \tag{B4}$$

The reflection and transmission matrices, which completely characterize the segmented liner, are then given by

$$\mathsf{T} = t = (\mathsf{F}\mathrm{D}_l - \mathsf{G}\mathsf{F}^{-1}\mathsf{G}\mathrm{D}_l)(\mathsf{F} - \mathsf{G}\mathrm{D}_l\mathsf{F}^{-1}\mathsf{G}\mathrm{D}_l)^{-1},$$

$$\mathsf{R} = r = (\mathsf{G} - \mathsf{F}\mathrm{D}_l\mathsf{F}^{-1}\mathsf{G}\mathrm{D}_l)(\mathsf{F} - \mathsf{G}\mathrm{D}_l\mathsf{F}^{-1}\mathsf{G}\mathrm{D}_l)^{-1}, \tag{B5}$$

where we have used the definition of T, R, and scattering matrix S of a single lining segment

$$\begin{pmatrix} \mathbf{A_2} \\ \mathbf{B_1} \end{pmatrix} = \mathsf{S}\begin{pmatrix} \mathbf{A_1} \\ \mathbf{B_2} \end{pmatrix}$$

where

$$\mathsf{S} = \begin{bmatrix} \mathsf{T} & \mathsf{R} \\ \mathsf{R} & \mathsf{T} \end{bmatrix}. \tag{B6}$$

The global scattering matrix of multiple lining segments are then easily obtained as given in Ref. 14.

[1] R. J. Astley, N. J. Walkington, and W. Eversman, "Transmission in flow ducts with peripherally varying liners," AIAA Pap. 80–1015 (1980).
[2] B. Regan and J. Eaton, "Modeling the influence of acoustic liner nonuniformities on duct modes," J. Sound Vib. **219**, 859–879 (1999).
[3] B. Tester, N. Baker, A. Kempton, and M. Wright, "Validation of an analytical model for scattering by intake liner splices," AIAA Pap. 2004–2906 (2004).
[4] T. Elnady, H. Bodén, and R. Glav, "Application of the point matching method to model circumferentially segmented non-locally reacting liners," AIAA Pap. 2001–2202 (2001).
[5] M. Namba and K. Fukushige, "Application of the equivalent surface source method to the acoustics of duct systems with non-uniform wall impedance," J. Sound Vib. **73**, 125–146 (1980).
[6] M. S. Howe, "The attenuation of sound in a randomly lined duct," J. Sound Vib. **87**, 83–103 (1983).
[7] W. R. Watson, "Circumferentially segmented duct liners optimized for axisymmetric and standing-wave sources," NASA Rep. No. 2075, 1982.
[8] C. R. Fuller, "Propagation and radiation of sound from flanged circular ducts with circumferentially varying wall admittances, I. Semi-infinite ducts," J. Sound Vib. **93**, 321–340 (1984).
[9] C. R. Fuller, "Propagation and radiation of sound from flanged circular ducts with circumferentially varying wall admittances. II. Finite ducts with sources," J. Sound Vib. **93**, 341–351 (1984).
[10] L. M. B. C. Campos and J. M. G. S. Oliveira, "On the acoustic modes in a cylindrical duct with an arbitrary wall impedance distribution," J. Acoust. Soc. Am. **116**, 3336–3346 (2004).
[11] M. C. M. Wright, "Hybrid analytical/numerical method for mode scattering in azimuthally nonuniform ducts," J. Sound Vib. **292**, 583–594 (2006).
[12] M. C. M. Wright and A. McAlpine, "Calculation of modes in azimuthally non-uniform lined ducts with uniform flow," J. Sound Vib. **302**, 403–407 (2007).
[13] R. J. Astley, V. Hii, and G. Gabard, "A computational mode matching approach for propagation in three-dimensional ducts with flow," AIAA Pap. 2006–2528 (2006).
[14] W. P. Bi, V. Pagneux, D. Lafarge, and Y. Aurégan, "Modelling of sound propagation in a non-uniform lined duct using a multi-modal propagation method," J. Sound Vib. **289**, 1091–1111 (2006).
[15] Y. Aurégan, M. Leroux, and V. Pagneux, "Measurement of liner impedance with flow by an inverse method," AIAA Pap. 2004–2838 (2004).
[16] W. P. Bi, V. Pagneux, D. Lafarge, and Y. Aurégan, "Efficient modeling sound propagation in nonuniform lined intakes," AIAA Pap. 2007–3522 (2007).
[17] G. A. Athanassoulis and K. A. Belibassakis, "A consistent coupled-mode theory for the propagation of small-amplitude water waves over variable bathymetry regions," J. Fluid Mech. **389**, 275–301 (1999).
[18] C. Hazard and V. Pagneux, "Improved multimodal approach in waveguides with varying cross-section," *Proceedings of the International Congress on Acoustics*, Rome, 2001, Vol. **25**, pp. 3,4.
[19] C. Hazard and E. Lunéville, "Multimodal approach and optimum design in non uniform waveguides," Workshop on the Method of Numerical and Electromagnetism, Toulouse, 2002.
[20] S. W. Rienstra, "A classification of duct modes based on surface waves," Wave Motion **37**, 119–135 (2003).
[21] G. N. Watson, *A Treatise on the Theory of Bessel Function* (Cambridge University Press, New York, 1952).

A multiple degree of freedom electromechanical Helmholtz resonator[a)]

Fei Liu and Stephen Horowitz
Department of Mechanical and Aerospace Engineering, University of Florida, Gainesville, Florida 32611-6250

Toshikazu Nishida
Department of Electrical and Computer Engineering, University of Florida, Gainesville, Florida 32611-6130

Louis Cattafesta and Mark Sheplak[b)]
Department of Mechanical and Aerospace Engineering, University of Florida, Gainesville, Florida 32611-6250

(Received 14 March 2007; revised 29 March 2007; accepted 4 April 2007)

The development of a tunable, multiple degree of freedom (MDOF) electromechanical Helmholtz resonator (EMHR) is presented. An EMHR consists of an orifice, backing cavity, and a compliant piezoelectric composite diaphragm. Electromechanical tuning of the acoustic impedance is achieved via passive electrical networks shunted across the piezoceramic. For resistive and capacitive loads, the EMHR is a 2DOF system possessing one acoustic and one mechanical DOF. When inductive ladder networks are employed, multiple electrical DOF are added. The dynamics of the multi-energy domain system are modeled using lumped elements and are represented in an equivalent electrical circuit, which is used to analyze the tunable acoustic input impedance of the EMHR. The two-microphone method is used to measure the acoustic impedance of two EMHR designs with a variety of resistive, capacitive, and inductive shunts. For the first design, the data demonstrate that the tuning range of the second resonant frequency for an EMHR with non-inductive shunts is limited by short- and open-circuit conditions, while an inductive shunt results in a 3DOF system possessing an enhanced tuning range. The second design achieves stronger coupling between the Helmholtz resonator and the piezoelectric backplate, and both resonant frequencies can be tuned with different non-inductive loads. © *2007 Acoustical Society of America.* [DOI: 10.1121/1.2735116]

PACS number(s): 43.50.Gf, 43.38.Fx, 43.58.Bh, 43.55.Ev [AJZ] Pages: 291–301

I. INTRODUCTION

Acoustic liners are used to line the engine nacelle duct wall to suppress the radiation of propulsion-related noise.[1] The acoustic liners provide a complex impedance boundary condition for the sound propagating within the duct. The liners can either be passive or active in terms of their noise suppression characteristics. A typical passive single degree of freedom (SDOF) acoustic liner cell is essentially a Helmholtz resonator that consists of a solid backplate, perforated face sheet, and honeycomb core. The conventional passive SDOF liners are most effective over a narrow frequency range, i.e., one octave.[1] Multiple degree of freedom (MDOF) passive liners have additional perforated septum sheets and offer a wider suppression bandwidth but represent a tradeoff in terms of size, weight, and cost.[2] The greatest limitation of passive liner technology is the constraint of fixed impedance for a given geometry. However, for a given aircraft propulsion system, there may be different desired impedance distributions associated with take-off, cut-back, and approach conditions.[2]

Active liners have attracted attention from researchers because of their potential to suppress engine noise under different operating conditions, but the associated drawbacks in terms of cost, complexity, and weight are non-trivial. Active liners configurations typically modify the performance *in situ* by changing the geometry characteristics of the liner or using bias flow through the liner resistive elements.[3–13] Note that such a system, as well as all active/adaptive systems, seeks to improve the attenuation characteristics of an acoustic liner by directly modifying the impedance of one or more of the components of the liner. Therefore, actuator(s), sensor(s), and a feedback controller are required. Such a system is often complex and potentially expensive to implement. The liner concept presented in this paper uses an alternative method of impedance tuning. The primary element of the liner is an electromechanical Helmholtz resonator (EMHR), as shown in Fig. 1, which consists of a narrow neck, cavity and compliant piezoelectric backplate.[14] In addition, a passive electrical network is attached to the diaphragm. The acoustic impedance of the EMHR is adjusted by modifying

[a)]Preliminary portions of this work were presented in "A tunable electromechanical Helmholtz resonator," AIAA Paper 2003-3145, at the 9th AIAA/CEAS Aeroacoustics Conference and Exhibit, Hilton Head, South Carolina, May 2003 and "A transfer matrix formulation of an electromechanical Helmholtz resonator," at ASA Fall Meeting, Minneapolis, Minnesota, 17–21 October 2005.

[b)]Author to whom correspondence should be addressed. Electronic mail: sheplak@ufl.edu

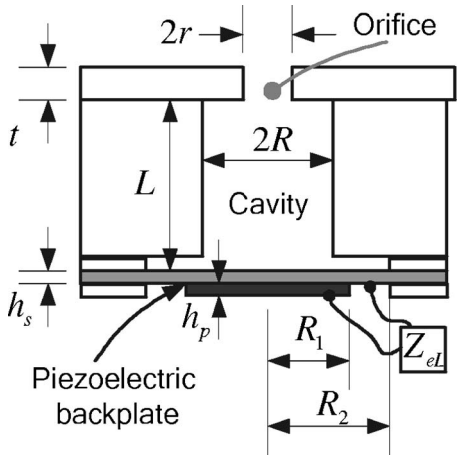


FIG. 1. Cross-section of an electromechanical Helmholtz resonator.

the shunt network. Furthermore, the EMHR is a 2DOF or a 3DOF system when the shunts are non-inductive and inductive, respectively.

This paper presents the theoretical aspects of the EMHR and demonstrates the concept via experiments using a variety of shunts for two different designs. The paper is organized as follows. Section II introduces the lumped-element model (LEM) of the EMHR. Section III presents an analysis of the equivalent circuit representation of the EMHR. Section IV summarizes analytical models for each of the lumped elements. Sections V and VI discuss the experimental setup and results, respectively. Section VII presents conclusions and offers guidance for future work.

II. LUMPED ELEMENT MODEL AND EQUIVALENT CIRCUIT REPRESENTATION

At low frequencies, the dimensions of the EMHR are much smaller than the wavelength of interest. The device components can thus be lumped into idealized discrete circuit elements.[15] An EMHR can be lumped into two parts. The first part is similar to the conventional Helmholtz resonator, where the neck of the resonator possesses both dissipative (R_{aN}) and inertial (M_{aN}) components and the cavity is modeled as a compliance (C_{aC}). The second part includes the short-circuit acoustic mass (M_{aD}), damping loss (R_{aD}), and compliance (C_{aD}) of the piezoelectric diaphragm and the electroacoustic transduction between the acoustical and electrical energy domains. The electroacoustic transduction is represented by an ideal transformer representing the effective acoustic piezoelectric transduction ratio (ϕ) and the electrical blocked capacitance of the piezoelectric diaphragm (C_{eB}). The equivalent circuit representation for the EMHR is shown in Fig. 2, where P and P' represent the incident and diaphragm acoustic pressures, respectively. Similarly, Q and Q' are incident and diaphragm volumetric flow rates, respectively. In the present notation, the first subscript denotes the domain (e.g., "a" for acoustic), while the second subscript describes the element (e.g., "D" for diaphragm). Z_{eL} is the electrical load impedance across the piezoelectric backplate, and V is the voltage output generated across the piezoelectric backplate under the excitation of an incident acoustic wave.

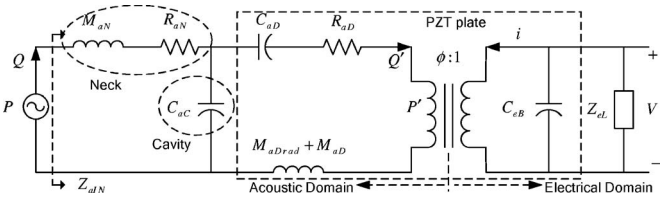


FIG. 2. Equivalent electroacoustic circuit representation of the EMHR (Ref. 29). Used with permission.

From a lumped element modeling perspective, the EMHR schematically shown in Fig. 1 is very similar to the compliant isotropic backplate Helmholtz resonator presented in Horowitz *et al.*[16] The primary difference is that the isotropic backplate impedance is replaced by a two-port network representing the electroacoustic transduction of the piezoelectric composite backplate.[17] The EMHR is similar in nature to a synthetic jet actuator presented in Gallas *et al.*[18] However, the piezoelectric backplate is actively driven in the latter case, while the EMHR is a passive device whose acoustic impedance is adjusted by changing the passive loads.

III. EQUIVALENT CIRCUIT MODEL ANALYSIS

A. Acoustic input impedance of the EMHR

Before estimating each parameter discussed above, it is instructive to analyze the equivalent circuit representation of the EMHR to obtain the acoustic input impedance, Z_{aIN}, using network theory.[19] As shown in Fig. 2, an ideal lossless transformer converts energy between the acoustical and electrical domains and "transforms" impedances between the two ports (energy domains), i.e., electrical impedance, Z_{eL}, to an acoustic impedance, Z_{aL}, by

$$Z_{aL} = \phi^2 Z_{eL}. \tag{1}$$

This enables a simplified equivalent circuit representation entirely in the acoustic domain as shown in Fig. 3 for various passive electrical loads. For the general case, the complex acoustic input impedance of the Helmholtz resonator, Z_{aIN}, shown in Fig. 3 is

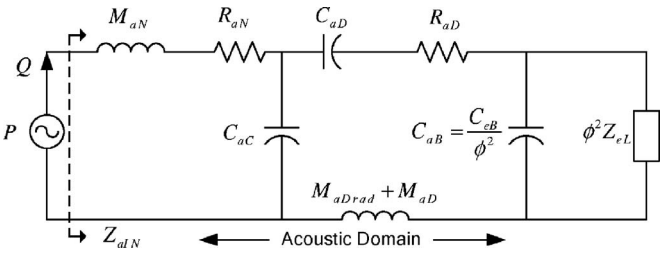


FIG. 3. Simplified equivalent acoustic circuit representation of the EMHR (Ref. 29). Used with permission.

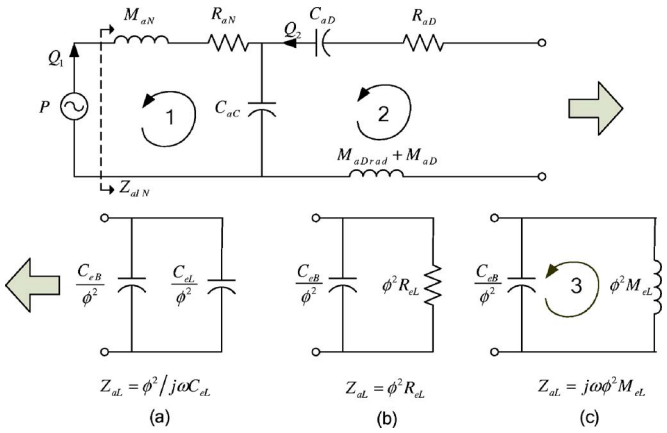


FIG. 4. (Color online) EMHR with passive electrical loads is analogous to a 2DOF system for (a) capacitive and (b) resistive loads and a 3DOF system with an (c) inductive load.

$$Z_{aIN}=\frac{P}{Q}=R_{aN}+sM_{aN}+\frac{\frac{1}{sC_{aC}}\left(\frac{1}{sC_{aD}}+R_{aD}+s(M_{aD}+M_{aDrad})+\frac{\phi^2 Z_{eL}}{1+sC_{eB}Z_{eL}}\right)}{\frac{1}{sC_{aC}}+\frac{1}{sC_{aD}}+R_{aD}+s(M_{aD}+M_{aDrad})+\frac{\phi^2 Z_{eL}}{1+sC_{eB}Z_{eL}}}, \tag{2}$$

where $s=j\omega$, $j=\sqrt{-1}$, and Z_{eL} represents different resistive, capacitive, and inductive loads.

Equation (2) clearly demonstrates that the acoustic impedance and the resonant frequencies, defined when $\mathrm{Im}(Z_{aIN})=0$, of the EMHR are a function of the passive electrical load. When the shunt loads are capacitive or resistive, the EMHR is analogous to a 2DOF system with two resonant frequencies, as shown in Figs. 4(a) and 4(b), respectively. The EMHR with inductive loads in Fig. 4(c) has 3DOF and thus has three resonant frequencies. As indicated by Eq. (2), the resonant frequencies of the EMHR can be tuned via adjusting the load Z_{eL}. The tuning behavior of the EMHR with various electrical loads is analyzed below. In particular, the tuning of weakly coupled, lightly damped EMHR with capacitive and inductive loads are both analyzed because these cases are mathematically tractable. Unfortunately, the explicit analytic solution for the resonant frequencies of the general EMHR without the above assumptions is more complicated. In these cases, it is more convenient to solve Eq. (2) numerically to obtain the resonant frequencies of the EMHR.

B. Capacitive tuning of the EMHR

As mentioned above, the EMHR with capacitive shunts ($Z_{eL}=1/sC_{eL}$) has 2DOF and, thus, has two resonant frequencies. To simplify the problem without loss of generality, it is assumed that the EMHR with capacitive shunts has negligible damping. The two undamped resonant frequencies are

$$f_{1,2}^2=\frac{(f_{L1}^2+f_{L2}^2)\pm\sqrt{(f_{L1}^2+f_{L2}^2)^2-4(1-\kappa^2)(f_{L1}f_{L2})^2}}{2}, \tag{3}$$

where

$$f_{L1}=1/(2\pi\sqrt{M_{aN}C_{aC}}) \tag{4}$$

is the resonant frequency of loop 1 (Fig. 4), and

$$f_{L2}=\frac{1}{2\pi\sqrt{\frac{(M_{aD}+M_{aDrad})(C_{aB}+C_{aL})C_{aD}C_{aC}}{(C_{aB}+C_{aL})C_{aD}+(C_{aB}+C_{aL}+C_{aD})C_{aC}}}} \tag{5}$$

is the resonant frequency of loop 2. The electromechanical coupling factor of the system is defined as

$$\kappa^2=\frac{C_{aC}}{C_{aC}^2}\frac{(C_{aB}+C_{aL})C_{aD}C_{aC}}{(C_{aB}+C_{aL})C_{aD}+(C_{aB}+C_{aL}+C_{aD})C_{aC}}=\kappa_1\kappa_2, \tag{6}$$

where κ_1 and κ_2 are the coupling coefficients of loop 1 and loop 2, respectively.[20] The coupling coefficient defines the ratio of the energy stored in the coupling elements to that stored in the total capacitance for each loop. For loop 1, the coupling coefficient is

$$\kappa_1=\frac{\left(\int Q_1 dt\right)^2/C_{aC}}{\left(\int Q_1 dt\right)^2/C_1}=\frac{C_{aC}}{C_{aC}}=1, \tag{7}$$

where Q_1 is the volume velocity, as shown in Fig. 4, and $C_1=C_{aC}$ is the total acoustic compliance of loop 1. Similarly, for loop 2, the coupling coefficient is

$$\kappa_2=\frac{(C_{aB}+C_{aL})C_{aD}}{(C_{aB}+C_{aL})C_{aD}+(C_{aB}+C_{aL}+C_{aD})C_{aC}}. \tag{8}$$

If the system is weakly coupled, e.g., the piezoelectric backplate is less flexible, such that $C_{aC}\gg C_{aD}$, $\kappa^2\to 0$, and Eq. (3) can be rewritten as

$$f_{1,2}=\sqrt{\frac{(1+\alpha)f_{HR}^2+f_D^2\pm\sqrt{[(1-\alpha)f_{HR}^2-f_D^2]^2}}{2}}, \tag{9}$$

where

$$f_{HR}=\frac{1}{2\pi}\sqrt{\frac{1}{M_{aN}C_{aC}}} \tag{10}$$

is the resonant frequency of the Helmholtz resonator with a rigid wall instead of a piezoelectric backplate, and

$$f_D=\frac{1}{2\pi}\sqrt{\frac{C_{aB}+C_{aL}+C_{aD}}{(M_{aD}+M_{aDrad})(C_{aB}+C_{aL})C_{aD}}} \tag{11}$$

is the resonant frequency of the piezoelectric backplate, and

$$\alpha=\frac{M_{aN}}{M_{aD}+M_{aDrad}} \tag{12}$$

is the mass ratio between the neck and piezoelectric backplate. If the mass ratio of the EMHR is very small, $\alpha\to 0$, then

$$f_{1,2} \approx \sqrt{\frac{f_{HR}^2 + f_D^2 \pm \sqrt{[f_{HR}^2 - f_D^2]^2}}{2}} = \begin{cases} f_{HR} \\ f_D. \end{cases} \tag{13}$$

In other words, the resonant frequencies of a lightly damped EMHR with capacitive shunts possessing weak coupling are approximately the resonant frequency of the solid-walled Helmholtz resonator and the piezoelectric backplate. As indicated by Eq. (11), the resonant frequency of the piezoelectric backplate is adjusted via changing the loads and summarized as follows:

- Open-circuit $Z_{aL}=1/sC_{aL}\to\infty$ or $C_{aL}=C_{eL}=0$,

$$[f_D]_{\text{open}} = \frac{1}{2\pi}\sqrt{\frac{C_{aB}+C_{aD}}{(M_{aD}+M_{aDrad})C_{aB}C_{aD}}}; \tag{14}$$

- Short-circuit $Z_{aL}=1/sC_{aL}\to 0$ or $C_{aL}=C_{eL}=\infty$,

$$[f_D]_{\text{short}} = \frac{1}{2\pi}\sqrt{\frac{1}{(M_{aD}+M_{aDrad})C_{aD}}}; \tag{15}$$

- General capacitive load $Z_{aL}=1/sC_{aL}$, where $C_{aL}=C_{eL}/\phi^2$,

$$[f_D]_{\text{capacitive}} = \frac{1}{2\pi}\sqrt{\frac{C_{aB}+C_{aL}+C_{aD}}{(M_{aD}+M_{aDrad})(C_{aB}+C_{aL})C_{aD}}}. \tag{16}$$

Inspection of Eq. (16) reveals that the resonant frequency of the piezoelectric backplate with capacitive loads is confined to short- and open-circuit limits,

$$[f_D]_{\text{short}} < [f_D]_{\text{capacitive}} < [f_D]_{\text{open}}, \tag{17}$$

or by the use of Eq. (13),

$$[f_2]_{\text{short}} < [f_2]_{\text{capacitive}} < [f_2]_{\text{open}}. \tag{18}$$

Equation (18) demonstrates how the second resonant frequency of the EMHR is tuned with different capacitive shunts. The second resonant frequency shifts between the short- and open-circuit limits as the capacitance changes.

The aforementioned analysis focuses on capacitive tuning behavior of the EMHR with weak coupling. In fact, when such coupling is not weak, the first resonant frequency of the system can be tuned *in situ* as well, as indicated by Eq. (3).

C. Inductive tuning of the EMHR

As shown in Fig. 4(c), the EMHR with inductive loads is a 3DOF system. To simplify the problem without loss of generality, it is assumed that the EMHR with inductive shunts has negligible damping. The effective impedance of the loop 3 is thus given by

$$Z_{\text{L3}} = \frac{(\phi^2/j\omega C_{eB})j\omega\phi^2 M_{eL}}{\phi^2/j\omega C_{eB} + j\omega\phi^2 M_{eL}}, \tag{19}$$

where the electrical resonance is given by

$$f_E = \frac{1}{2\pi}\sqrt{\frac{1}{C_{eB}M_{eL}}}. \tag{20}$$

In the frequency range $f \ll f_E$, the impedance of loop 3 can be further approximated as

$$Z_{\text{L3}} \simeq j\omega\phi^2 M_{eL}. \tag{21}$$

Physically, this approximation means that the impedance of loop 3 effectively adds a mass $\phi^2 M_{eL}$ to loop 2 [Fig. 4(c)]. The resonant frequencies of the weakly coupled EMHR are given by

$$f_{1,2} \approx \sqrt{\frac{f_{HR}^2 + f_D^2 \pm \sqrt{[f_{HR}^2 - f_D^2]^2}}{2}} = \begin{cases} f_{HR} \\ f_D \end{cases}, \tag{22}$$

where f_{HR} is given by Eq. (10), while f_D is

$$f_D = \frac{1}{2\pi}\sqrt{\frac{1}{(M_{aD}+M_{aDrad}+\phi^2 M_{eL})C_{aD}}}. \tag{23}$$

Clearly, $f_D \leqslant [f_D]_{\text{short}}$ in Eq. (15), which illustrates how inductive loading enables a tuning range greater than that defined by the open- and short-circuit limits.

IV. MODEL PARAMETER ESTIMATION

In this section, the methods and assumptions used to estimate each of the parameters in Eq. (2) are outlined. Table I provides equations to evaluate the parameters of the LEM of the neck (radius r and thickness t) and cavity (radius R and depth L) of the EMHR, as shown in Fig. 1. For more details on the lumped element modeling of the neck and cavity, the reader is referred to Horowitz *et al.*[16] The extraction of the model parameters for the piezoelectric backplate (Fig. 1) is more complex due to composite plate mechanics. The clamped piezoelectric backplate consists of an axisymmetric piezoceramic of radius R_1 and thickness h_p bonded in the center of a metal shim of radius R_2 and thickness h_s. Up to and just beyond the first resonant mode, the one-dimensional piezoelectric electroacoustic coupling is given by[17]

$$\begin{Bmatrix} \Delta\tilde{V} \\ q \end{Bmatrix} = \begin{Bmatrix} C_{aD} & d_a \\ d_a & C_{eF} \end{Bmatrix} \begin{Bmatrix} P \\ V \end{Bmatrix}, \tag{24}$$

where $\Delta\tilde{V}$ is the volume displacement of the piezoelectric backplate due to the application of the pressure P and voltage V. Additionally, q is the charge stored on the piezoelectric electrodes, d_a is the effective acoustic piezoelectric coefficient, and C_{eF} is the electrical free capacitance of the piezoelectric material. The volume displacement is calculated by integrating the transverse displacement, $w(r)$, over the whole plate

$$\Delta\tilde{V} = \int_0^{R_2} 2\pi r w(r) dr. \tag{25}$$

Thus, the short-circuit ($V=0$) acoustic compliance, C_{aD}, and the effective acoustic piezoelectric coefficient ($P=0$), d_a, can be determined following Eq. (24) and are listed in Table I, which also provides equations for the other parameters of the piezoelectric backplate. For more details on the lumped element modeling of the piezoelectric backplate, the reader is referred to Prasad *et al.*[17]

It is worth noting that the radiation mass, M_{aDrad}, of the vibrating diaphragm is modeled by treating the vibrating backplate as a baffled piston. This assumption is reasonable

TABLE I. Lumped elements modeling parameter estimation for the neck and cavity of the EMHR[a,b,c,d,e,f].

	LEM parameters	Description
Neck[c,d]	R_{aN}	$R_{aN}=\frac{t}{r}\frac{\sqrt{2\mu\rho_0\omega}}{\pi r^2}+2\frac{\sqrt{2\mu\rho_0\omega}}{\pi r^2}+\frac{\rho_0 c_0}{\pi r^2}\left[1-\frac{2J_1(2kr)}{2kr}\right]$, where J_1 is the Bessel function of the first kind and $k=\omega/c_0$ is the wave number
	M_{aN}	$M_{aN}=\frac{\rho_0(t+\Delta t)}{\pi r^2}$, where $\Delta t=0.85r\left(1-0.7\frac{r}{R}\right)+0.85r$
Cavity[e,f]	C_{aC}	$C_{aC}=\frac{\hat{V}}{\rho_0 c_0^2}$, where $\hat{V}$ is volume of the cavity
Piezoelectric backplate[b]	C_{aD}	$C_{aD}=\int_0^{R_2} 2\pi r w(r)dr \Big/ P\Big\vert_{V=0}$
	M_{aD}	$M_{aD}=\frac{2\pi}{(\Delta\tilde{V})^2}\int_0^{R_2}\rho_A(w(r))^2 r dr\Big\vert_{V=0}$
	M_{aDrad}	$M_{aDrad}=\frac{8\rho_0}{3\pi^2 R_2}$ (see footnote g below)
	R_{aD}	$R_{aD}=2\xi\sqrt{\frac{M_{aD}+M_{aDrad}}{C_{aD}}}$ where ξ is the experimentally determined damping factor[g]
	ϕ	$\phi=\frac{-d_a}{C_{aD}}=\frac{-\int_0^{R_2} 2\pi r w(r)dr \Big/ V\Big\vert_{P=0}}{C_{aD}}$
	C_{eB}	$C_{eB}=C_{eF}(1-\kappa^2)=\frac{\varepsilon_r\varepsilon_0\pi R_1^2}{h_p}(1-\kappa^2)$ where ε_r is the relative dielectric constant of the piezoelectric material, ε_0 is the permittivity of free space, and $\kappa^2=d_a^2/C_{eF}C_{aD}$ is the electroacoustic coupling factor

[a]Reference 16.
[b]Reference 17.
[c]Reference 21.
[d]Reference 22.
[e]Reference 23.
[f]Reference 24.
[g]Reference 25.

TABLE II. Dimensions of the EMHRs (Unit: mm, resolution 0.01 mm).

			Case I	Case II
Neck		Radius r	2.42	2.42
		Length t	3.16	3.16
Cavity		Radius R	6.34	6.34
		Depth L	16.42	9.38
Piezoelectric backplate	Piezoceramic	Radius R_1	10.05	10.05
		Thickness h_p	0.13	0.13
	Shim	Radius R_2	12.41	12.41
		Thickness h_s	0.19	0.19

for the EMHR in this paper because the piezoelectric backplate is mounted in a plate that is larger in extent than the radius of the backplate. However, such an assumption breaks down when the piezoelectric backplate is not baffled.

The acoustic resistance, R_{aD}, of the PZT backplate models acoustic resistance and structural damping in the backplate. The damping may arise from thermoelastic dissipation, compliant boundaries, and other intrinsic loss mechanisms. The acoustic resistance is given by

$$R_{aD} = 2\xi\sqrt{\frac{M_{aD} + M_{aDrad}}{C_{aD}}}, \tag{26}$$

where ξ is an experimentally determined damping factor determined using, for example, the logarithmic decrement method.[25] In this study, the measured value of ξ is 0.01 for the piezoelectric backplate for the conditions and geometry listed in Table II.

Each parameter in the LEM is now defined in terms of device dimensions and fluid and material properties. The device dimensions are listed in Table II, and material properties are listed in Table III. By substituting each parameter into Eq. (2), one can obtain the acoustic impedance of the resonator, as shown in Fig. 5.

V. EXPERIMENTAL SETUP

The experimental setup implements the standard two-microphone method (TMM) for measuring acoustic impedance,[26] shown in Fig. 6. The plane wave tube (PWT) consists of a 965 mm long, 25.4 mm by 25.4 mm square duct that permits a plane wave acoustic field at frequencies up to 6.7 kHz. The normalized specific acoustic impedance of the Helmholtz resonator is defined as

$$\zeta = \mathrm{Re}\left(\frac{Z_{aIN}A_o}{\rho_0 c_0}\right) + \mathrm{Im}\left(\frac{Z_{aIN}A_o}{\rho_0 c_0}\right) = \theta + j\chi, \tag{27}$$

where ρ_0 is the mean air density, c_0 is the isentropic sound speed in the air, A_o is the area of the orifice of the EMHR,

TABLE III. Materials parameters for the piezoelectric backplate.

Piezoceramic (APC 850)	Young's modules (N/m^2)	6.3E10
	Poisson 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 modules (N/m^2)	11.0E10
	Poisson ratio	0.38
	Density (kg/m^3)	8530

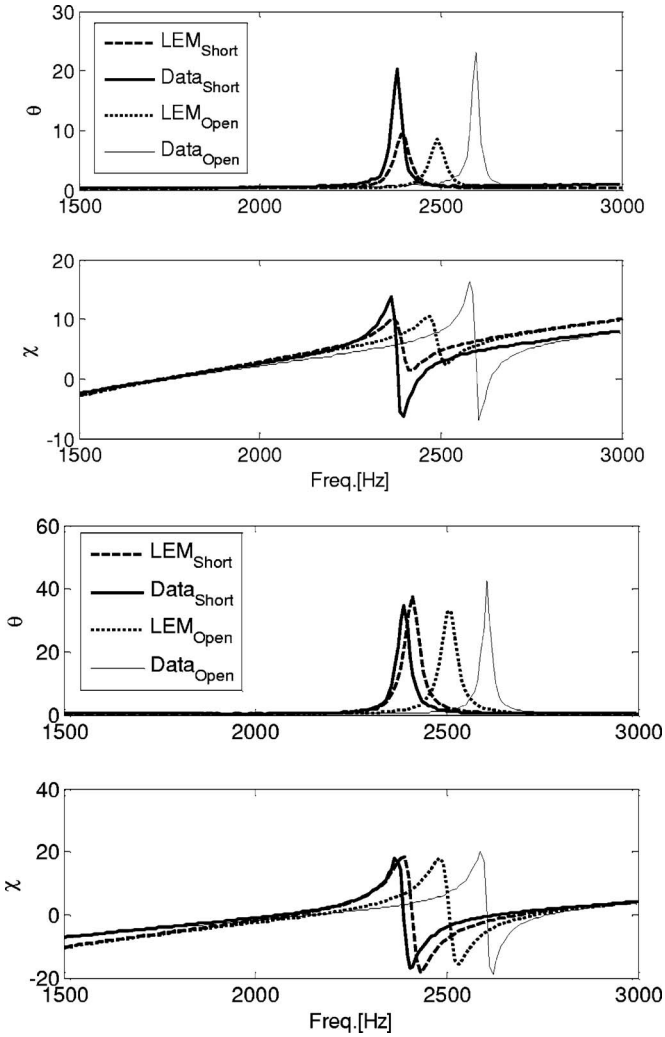


FIG. 5. Comparison between model prediction and experimental results for the normalized specific acoustic impedance of the EMHRs for the short- and open-circuit cases (upper: Case I, lower: Case II). Note that f_{1s} is 1732 Hz for Case I and 2064 Hz for Case II.

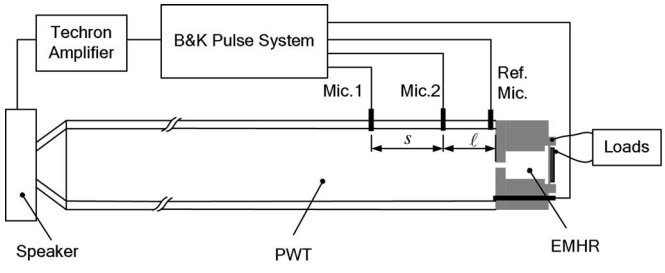


FIG. 6. Schematic of the plane wave tube terminated by an EMHR.

TABLE IV. Selected loads matrix used in the experiment to tune the EMHR.

Resistive loads (Ω)		Capacitive loads (nF)		Inductive loads (mH)	
Nominal	Measured	Nominal	Measured	Nominal	Measured
Short	Short	Short	Short	Short	Short
200	199.1	10	10.3	100	101.9(65.4)[a]
2k	2.0k	47	37.8	300	304.6(191.8)
7.5k	7.4k	100	89.7	500	507.8(319.8)
Open	Open	Open	Open	Open	Open

[a]The value in ( ) represents the resistance of the inductive loads in Ω.

and θ and χ are the normalized specific resistance and reactance, respectively.

Three Brüel and Kjær (B&K) type 4138 microphones were used. Two microphones, labeled as Mic. 1 and Mic. 2, were flush mounted in a rotating plug to the side of the impedance tube, as shown in Fig. 6. The rotating plug is used to remove the phase mismatch between microphones. The third microphone was flush mounted 6.3 mm away from the end face of the impedance tube to measure the total acoustic pressure at the entrance of the EMHR. This microphone also served as a reference to monitor sound pressure level (SPL) at the entrance of the resonator. In the experiment, the SPL is set to be low enough (<100 dB re 20 μPa) for nonlinear behavior of the EMHR to be neglected.

All microphones are calibrated with a B&K 4228 Pistonphone when connected to a B&K PULSE Multi-Analyzer System Type 3560. The PULSE system served as the power supply and data acquisition and processing system for the microphones as well as the signal source. A pseudo-random waveform generated from the PULSE system was fed through a Techron Model 7540 power supply amplifier to drive a BMS H4590P compression driver which can produce acoustic waves between 200 Hz and 22 kHz. The periodic random noise waveform was bandpass filtered from 300 Hz to 6.7 kHz ($\Delta f=2$ Hz). Meanwhile, a zoom FFT (frequency span 3.2 kHz, center frequency 1.9 kHz, 400 lines, and 1500 ensemble averages) was performed on each microphone signal.

The specimen used in the acoustic impedance measurement using TMM is an EMHR, as shown in Fig. 1. The piezoelectric backplate is a commercially available piezoceramic circular bender disk from APC International, Ltd. The piezoceramic material employed in this bender disk is APC850 lead zirconate titanate (PZT). The dimensions and material properties of the piezoelectric backplate Helmholtz resonator are listed in Tables II and III, respectively.

VI. EXPERIMENTAL RESULTS AND DISCUSSION

A variety of loads were used in the experiment to investigate the tuning ability of the resonator. Table IV lists the different loads used in the experiment. Experimental results for the normalized specific acoustic impedance, defined in Eq. (2), of the piezoelectric backplate Helmholtz resonator are presented below.

The EMHR with a short- or open-circuit is analogous to a 2DOF system, and has two resonant frequencies. The resonance of the EMHR is due to an oscillatory conversion between kinetic and potential energy. In comparison with the short-circuit case, an EMHR with an open-circuit has a decreased acoustic compliance, and thus increases potential energy storage and increases the resonant frequency of the system, as shown in Fig. 5. A comparison between the numerical and experimental results of the short-circuit case ($Z_{eL}=0$) and the open-circuit case ($Z_{eL}=\infty$) is also shown in Fig. 5. The comparison indicates that the model prediction for the normalized specific acoustic impedance matches the experimental data quite well for the short-circuit case, although the predictions for the normalized specific acoustic impedance do not match the experimental data very well for the open-circuit case. One possible explanation for the observed discrepancy is that, for the short-circuit case, the piezoelectric backplate is electrically shorted and, thus, C_{eB} and ϕ do not affect the acoustic impedance of EMHR as they do for the open-circuit case. Hence, any inaccuracies in either C_{eB} or ϕ will impact the predicted results. Some factors do impact the piezoelectric backplate model, such as the bond layer between the piezoceramic patch and the brass shim (the model assumes a negligible bond layer), any asymmetry in the piezoceramic patch geometry, and imperfect clamped boundary condition.[17] The results also reveal that the measured resistance is larger than the predicted results in the low and high frequency ranges. The deviations may be caused by the fact that the reflection coefficient is close to unity in those regions. When the reflection coefficient is near unity, the uncertainty in the acoustic impedance measurement using the TMM will become very large.[27] The uncertainty in the acoustic impedance arises from the measurement uncertainties in microphone locations (ℓ and s in Fig. 6), the frequency response function between the two measurement microphones, ambient temperature, and pressure.[27]

Improvement in the predictive performance of the LEM is obtained via parameter extraction of the piezoelectric backplate. The parameter extraction is performed by measuring the transverse deformation of the piezoelectric backplate under voltage loading.[17] Figure 7 shows the comparison between the measurement data and analytical result for the deformation of the piezoelectric plate. The actual displacement of the piezoelectric plate is larger than that predicted using the analytical method.[17] This is likely due to non-idealities in the composite plate boundary conditions. A clamped boundary is difficult to achieve in practice and any compliance in the boundaries or in-plane compressive stress due to mounting will result in an enhanced d_a. Figure 8 shows prediction results for short- and open-circuited EMHRs using the experimentally extracted d_a. The model prediction for the normalized specific acoustic impedance matches the experimental data quite well for both short- and open-circuit cases. The results indicate that the functional form of the LEM accurately captures the physical behavior of the EMHR.

Figure 9 shows only the measured acoustic impedance for the resonator (Case I in Table II) with a variety of capacitive loads for clarity, the measured reflection coefficient magnitude is shown in Fig. 10. As discussed above, the EMHR with capacitive loads has two resonant frequencies located where the reactance of the EMHR crosses zero with

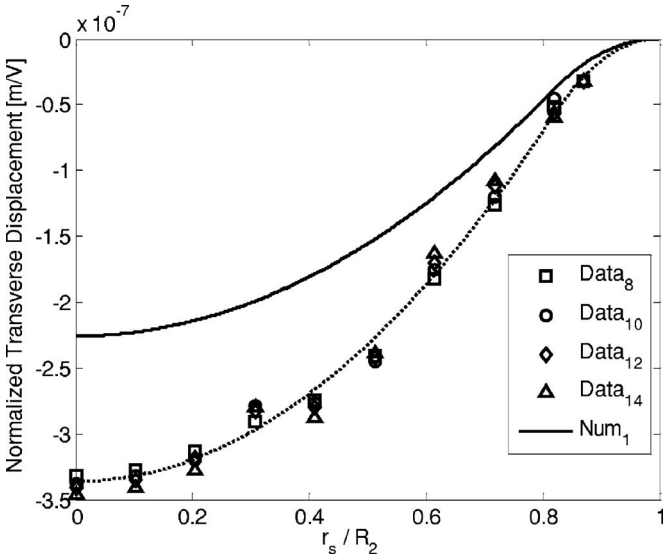


FIG. 7. Comparison between the analytical result and measured deformation of the piezoelectric backplate due to the application of various voltages. Note that $Data_8$ denotes the measurement result where the magnitude of applied voltage is 8 volt.

positive slope, as shown in Fig. 9. There is also an anti-resonant frequency where the reactance of the EMHR crosses zero with negative slope. Near the resonant frequencies, there is a sharp dip in the reflection coefficient magnitude (Fig. 10). Clearly, as the capacitance is decreased, the second resonant frequency shifts towards the open-circuit case from the short-circuit case, while the first resonant frequency barely changes for the EMHR (Case I) studied in this paper. As described in Sec. III B, this is because of the weak coupling between the Helmholtz resonator and the PZT backplate for the tested EMHR. There is approximately an 8% capacitive tuning range of the second resonant frequency of the EMHR for the conditions and geometry listed in Table II.

Figure 11 shows only the measured results for different resistive loads across the Case I resonator for clarity. Again, the EMHR with resistive loads has 2DOF [Fig. 4(b)]. The resistive loads effectively change the acoustic resistance of the PZT backplate and, thus, shift the resonant frequency of the EMHR. A trend similar to the capacitive tuning can be observed for the resistive tuning, in which the tuning range is also restricted between the short- and open-circuit limits. When the resistance is increased, the resonant frequency shifts towards the open-circuit case as expected. Moreover, resistive loads reduce the amplitude of the impedance peaks via system damping. Maximum damping (i.e., minimized peaks) occur when the magnitude of the output resistance of the EMHR and external load resistance are matched.

Figure 12 shows the results for inductive tuning of the Case I EMHR, and the measured reflection coefficient magnitude is shown in Fig. 13. Unlike capacitive and resistive tuning, inductive tuning provides an additional DOF for the EMHR, resulting in a 3DOF EMHR with three resonant frequencies [Figs. 4(c) and 13]. The third resonant frequency shifts closer to the second resonant frequency open-circuit case as the inductance is increased. However, as a result of

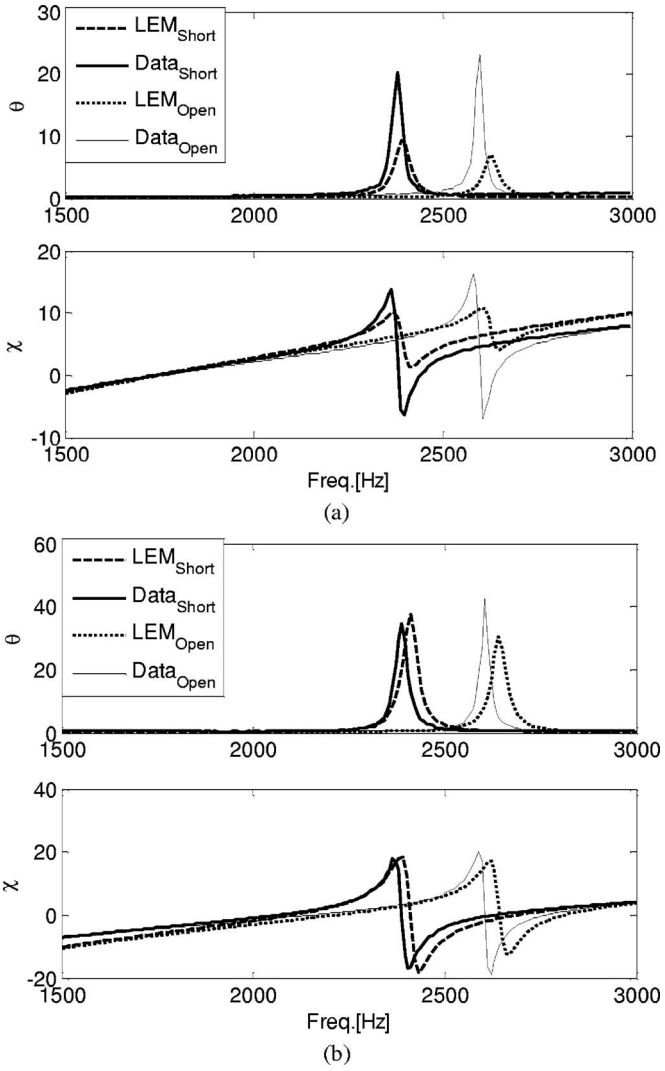


FIG. 8. Predictions of the LEM and TM for short- and open-circuited EMHRs using the experimentally extracted d_a. (a) EMHR (Case I), (b) EMHR (Case II).

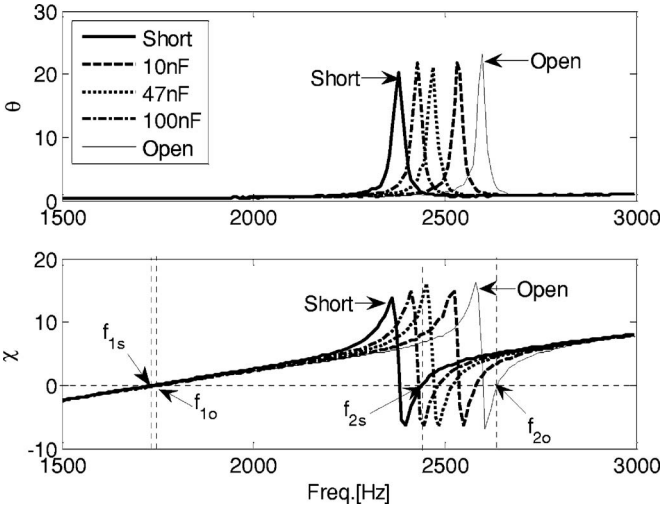


FIG. 9. Experimental results for the normalized specific acoustic impedance of the EMHR (Case I) as function of the capacitive load.

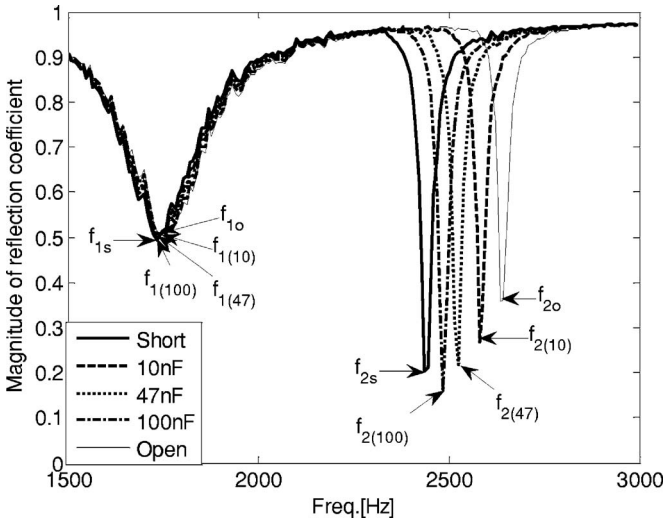


FIG. 10. Experimental results for the reflection coefficient of the EMHR (Case I) as function of the capacitive load. Note that $f_{1(10)}$ denotes the first resonant frequency under a capacitive loading of 10 nF.

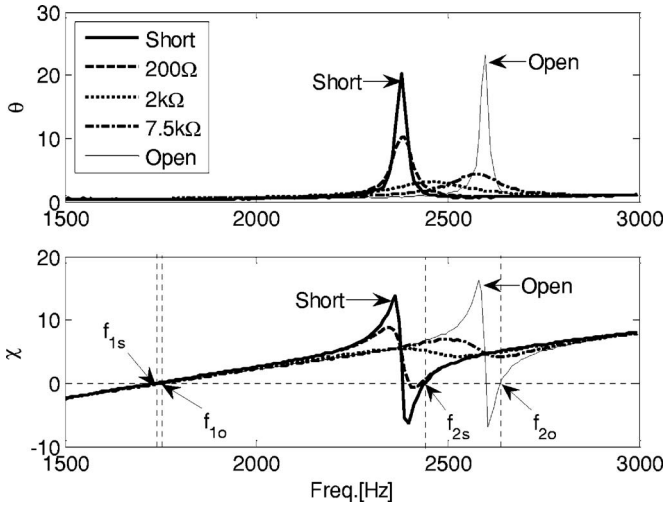


FIG. 11. Experimental results for the normalized specific acoustic impedance of the EMHR (Case I) as a function of the resistive load.

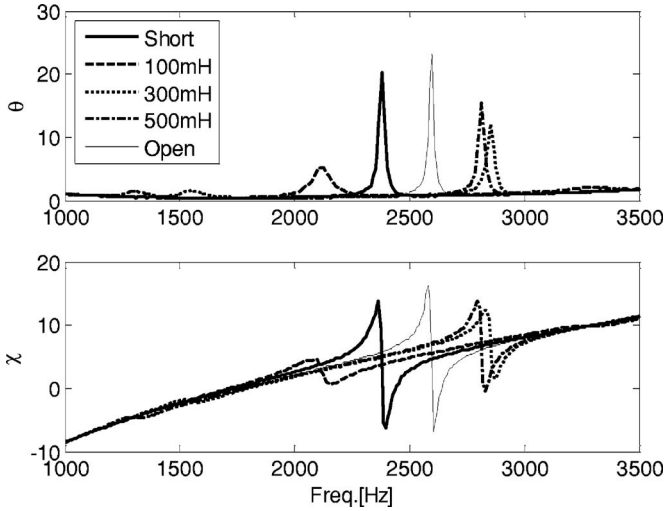


FIG. 12. Experimental results for the normalized specific acoustic impedance of the EMHR (Case I) as function of the inductive load.

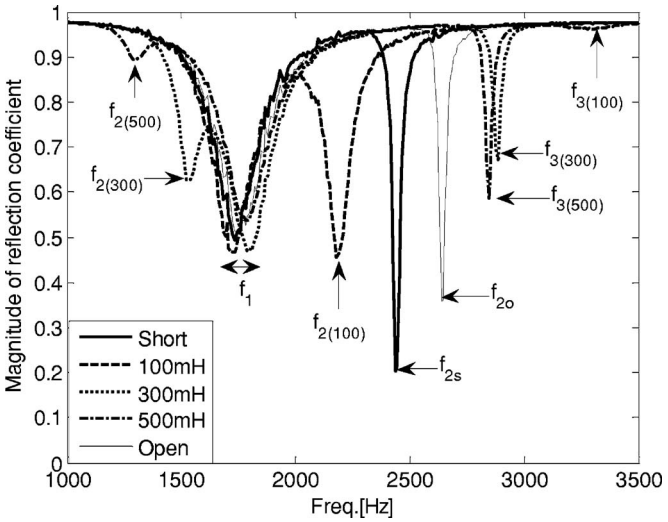


FIG. 13. Experimental results for the reflection coefficient of the EMHR (Case I) as function of the inductive load. Note that $f_{3(300)}$ denotes the third resonant frequency under an inductive loading of 300 mH.

the 3DOF, inductive tuning is not restricted to lie between the short- and open-circuit limits. Rather, the second resonant frequency always lies outside the short- to open-circuit tuning range as shown in Figs. 12 and 13. Notice that the second resonant frequency of the resonator shifts to lower frequencies as the inductance is increased. Moreover, when the inductive load is large enough, the second resonant frequencies is smaller than the first resonant frequency. Table V lists the tuning results of the EMHR (Case I) with inductive loads. There is approximately a 47% inductive tuning range ($\Delta f_2/f_{2s}$) of the EMHR (Case I) under the conditions and geometry listed in Table II. Again, the first resonant frequency only changes slightly due to weak coupling between the piezoelectric backplate and the Helmholtz resonator. Furthermore, a real inductor possesses finite resistance as well (see Table IV). Typically, the resistance increases with increasing inductance, so the larger inductor also has a larger resistance.

The experimental results of the Case II EMHR with short- and open-circuits are shown in Fig. 14. It indicates that the first resonant frequency of the EMHR shifts due to non-weak coupling between the piezoelectric backplate and the Helmholtz resonator. In comparison with the Case I EMHR, the depth of the cavity of the Case II EMHR is shorter, so the acoustic compliance of the cavity consequently decreases. Hence, the coupling factor given by Eq. (6) is larger than that of Case I.

TABLE V. Inductive tuning of the EMHR (Case I).[a]

Inductive load (mH)	f_1 (Hz)	f_2 (Hz)	f_3 (Hz)	$\Delta f_1/f_{1s}$ (%)	$\Delta f_2/f_{2s}$ (%)
Short	1732	2436	⋯	⋯	⋯
100	1716	2180	3300	−0.9	−10.5
300	1788	1532	2884	3.2	−37.1
500	1780	1300	2844	2.8	**−46.6**
Open	1756	2636	⋯	1.4	**8.2**

[a]The tuning ranges are defined as $\Delta f_1 = f_{1(\text{shunts})} - f_{1s}$ and $\Delta f_2 = f_{2(\text{shunts})} - f_{2s}$.

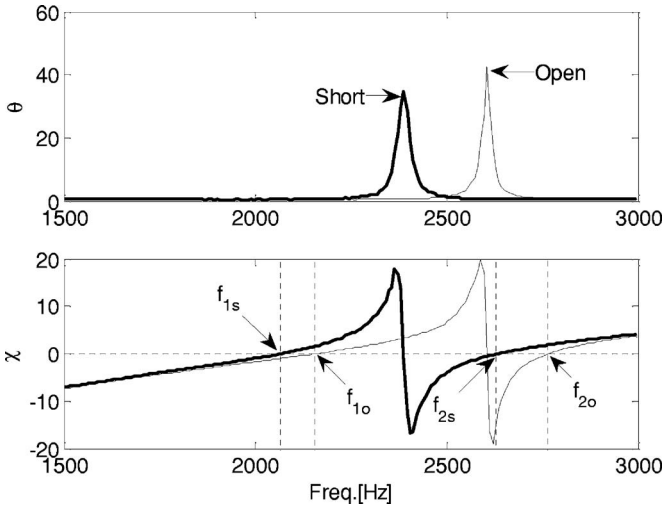


FIG. 14. Experimental results of the normalized acoustic impedance of the EMHR (Case II) for the short- and open-circuit.

VII. CONCLUSIONS AND FUTURE WORK

Two EMHRs are theoretically and experimentally studied as a means to passively tune the noise suppression frequency range of a Helmholtz resonator. The acoustic impedance as well as the resonant frequencies of the EMHR are electromechanically adjusted by changing the shunted electrical loads without changing any physical dimensions. For the first EMHR (Case I) studied in this paper, there is only weak coupling between the Helmholtz resonator and the piezoelectric backplate. The second resonant frequency of the EMHR can thus be tuned, while the first resonant frequency barely changes with different shunt loads. There is an approximately 8% tuning range for the second resonant frequency of the EMHR via capacitive or resistive loads, while the tuning range of the inductive load reaches up to 47%. Moreover, the range of resistive and capacitive tuning is limited between the short- and open-circuit limits, while an inductive shunt greatly enhances the tuning range and is not theoretically limited to the 2DOF short- and open-circuit limits. For the EMHR (Case II), there is stronger coupling between the Helmholtz resonator and the piezoelectric backplate, therefore both resonant frequencies can thus be adjusted with different loads.

The LEM provided an effective means to analyze the tuning behavior of the EMHR. In particular, the model prediction provides a good match with the experimental data for short-circuit case, while there is a deviation between model estimation and data when the piezoelectric backplate of the EMHR is not shorted, where LEM parameters ϕ and C_{eB} are important. The experimentally extracted parameter d_a illustrated that the piezoelectric composite model underpredicted the electroacoustic coupling. This was likely due to non-ideal boundary conditions. Clearly, the practical implementation of this technology must provide a well-defined mechanical mounting scheme with known boundary conditions for the backplates.

Additional efforts are aimed at the development of an optimal design methodology for the EMHR to satisfy specific design requirements. The optimization of the EMHR being currently undertaken follows two paths. One is to maximize the tuning ranges of the EMHR with capacitive or resistive loads; the other is to optimize the EMHR with inductive loads. For an optimized EMHR with non-inductive shunts, the first and second resonant frequency can be tuned simultaneously by adjusting shunt loads. However, the maximizing tuning range of both resonant frequencies cannot be achieved simultaneously. Optimizing the tuning range of one resonant frequency will degrade the tuning range of the other.[28] For an optimized EMHR with inductive shunts, optimal tuning of all three resonant frequencies has not yet been addressed.

ACKNOWLEDGMENTS

Financial support for this project was provided by NASA Langley Research Center (Grant No. NAG-1-2261), monitored by Michael G. Jones.

[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: Noise Control*, edited by H. H. Hubbard (Acoustical Society of America, New York, 1995), Chap. 14, pp. 165–206.
[2] G. W. Bielak, J. W. Premo, and A. S. Hersh, "Advanced turbofan duct liner concepts," NASA/CR-1999-209002 (National Aeronautics and Space Administration, Hampton, VA, 1999).
[3] E. Little, R. Kashani, J. Kohler, and F. Morrison, "Tuning of an electrorheological fluid-based Helmholtz resonator as applied to hydraulic engine mounts," in *Transportation Systems*, edited by R. Kashani (American Society of Mechanical Engineers, New York, 1994), pp. 43–51.
[4] J. M. De Bedout, 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).
[5] G. B. Thurston, L. E. Hargrove, Jr., and B. D. Cook, "Nonlinear properties of circular orifices," J. Acoust. Soc. Am. **29**, 992–1001 (1957).
[6] M. S. Howe, "On the theory of unsteady high Reynolds number flow through a circular cylinder," J. Fluid Mech. **61**, 109–127 (1979).
[7] F. G. Leppington and H. Levine, "Reflexion and transmission at a plane screen with periodically arranged circular or elliptical apertures," J. Fluid Mech. **61**, 109–127 (1973).
[8] I. J. Hughes and A. P. Dowling, "The absorption of sound by perforated linings," J. Fluid Mech. **218**, 299–355 (1990).
[9] X. D. Jing and X. F. Sun, "Experimental investigations of perforated liner with bias flow," J. Acoust. Soc. Am. **106**, 2436–2441 (1999).
[10] H. W. Zhao and X. F. Sun, "Active control of wall acoustic impedance," AIAA J. **37**, 825–831 (1999).
[11] B. E. Walker, A. S. Hersch, L. J. Heidelberg, D. L. Sutliff, and M. E. Spencer, "Active resonators for control of multiple spinning modes in an axial flow fan inlet," AIAA Paper 1999-1853, in *5th AIAA/CEAS Aeroacoustics Conference and Exhibit* (Bellevue, WA, 1999).
[12] M. A. Galland, B. Mazeaud, and N. Sellen, "Hybrid passive/active absorbers for flow ducts," Appl. Acoust. **66**, 691–708 (2005).
[13] J. P. Smith and R. A. Burdisso, "Experiments with fixed and adaptive Herschel-Quincke waveguides on the Pratt and Whitney JT15D engine," NASA/CR-2002-211430 (National Aeronautics and Space Administration, Hampton, VA, 2002).
[14] M. Sheplak, L. Cattafesta, T. Nishida, and S. B. Horowitz, "Electromechanical acoustic liner," U.S. Patent No. 6,782,109 (2004).
[15] M. Rossi, *Acoustics and Electroacoustics* (Artech House, Norwood, MA, 1988), pp. 245–373.
[16] S. B. Horowitz, T. Nishida, L. Cattafesta, and M. Sheplak, "Characterization of a compliant-backplate Helmholtz resonator for an electromechanical acoustic liner," Int. J. Aeroacoust. **1**, 183–205 (2002).
[17] S. Prasad, Q. Gallas, S. Horowitz, B. Homeijer, B. Sankar, L. Cattafesta, and M. Sheplak, "An analytical electroacoustic model of a piezoelectric composite circular plate," AIAA J. **44**(10), 2311–2318 (2006).
[18] Q. Gallas, R. Holman, T. Nishida, B. Carrol, M. Sheplak, and L. Cattafesta, "Lumped element modeling of piezoelectric-driven synthetic jet actuators," AIAA J. **41**, 240–247 (2003).
[19] M. E. Van Valkenburg, *Network Analysis* (Prentice-Hall, Englewood

Cliffs, 1964), Chaps. 3 and 9.
[20]F. A. Fischer, *Fundamentals of Electroacoustics* (Interscience, New York, 1955), pp. 1–51.
[21]I. B. Crandall, *Theory of Vibrating Systems and Sound* (Van Nostrand, New York, 1926), pp. 229–241.
[22]U. Ingard, "On the theory and design of acoustic resonators," J. Acoust. Soc. Am. **25**, 1037–1061 (1953).
[23]D. T. Blackstock, *Fundamentals of Physical Acoustics* (Wiley, New York, 2000).
[24]L. E. Kinsler, *Fundamentals of Acoustics* (Wiley, New York, 2000), pp. 272–274.
[25]L. Meirovitch, *Fundamentals of Vibrations* (McGraw-Hill, New York, 2001), pp. 94–98.
[26]ASTM-E1050-98, "Impedance and absorption of acoustical materials using a tube, two microphones, and a digital frequency analysis system," ASTM International.
[27]T. Schultz, M. Sheplak, and L. Cattafesta, "Uncertainty analysis of the two-microphone method," J. Sound Vib. **304**, 91–109 (2007).
[28]F. Liu, S. B. Horowitz, L. Cattafesta, and M. Sheplak, "Optimization of an electromechanical Helmholtz resonator," AIAA Paper 2006-2524, in *12th AIAA/CEAS Aeroacoustics Conference* (Cambridge, Massachusetts, 2006).
[29]F. Liu, S. B. Horowitz, L. Cattafesta, and M. Sheplak, "A tunable electromechanical Helmholtz resonator," AIAA Paper 2003-3145, in *9th AIAA/CEAS Aeroacoustics Conference* (Hilton Head, South Carolina, 2003).

On the modeling of the diffuse field sound transmission loss of finite thickness apertures

Franck Sgard[a)]
Laboratoire des Sciences de l'Habitat, DGCB URA CNRS 1652, Ecole Nationale des Travaux Publics de l'Etat, 69518 Vaulx-en-Velin Cedex, France

Hugues Nelisse
Institut de Recherche Robert-Sauvé en Santé et en Sécurité du Travail, 505, Boul. de Maisonneuve Ouest, Montréal Quebec H3A 3C2, Canada

Noureddine Atalla
Department of Mechanical Engineering, Université de Sherbrooke, 2500 Boulevard Université, Sherbrooke, Quebec J1K 2R1, Canada

(Received 5 December 2006; revised 2 April 2007; accepted 3 April 2007)

The modeling of the diffuse field sound transmission loss (TL) of apertures has been rarely considered in the literature. The aims of this paper are (i) to give a comprehensive review of the existing models, (ii) to propose a general efficient and rigorous numerical method to predict the diffuse field TL of apertures of rectangular and circular cross section, (iii) to provide the reader with numerical results regarding this indicator together with its relation with the normal incidence case for various geometrical configurations, and (iv) to conclude on the relevance of using such a sophisticated model compared to more classical normal incidence ones. The proposed approach is based on the description of the sound field inside the aperture in terms of propagating and evanescent acoustic modes. The radiation of the aperture is accounted for using a modal radiation impedance matrix. The coupled problem is solved in terms of modal contribution factors. The convergence of the approach is then investigated and the model is validated by comparisons with existing models for various configurations and excitations. Several numerical examples are provided regarding the normal incidence and diffuse field TL for various apertures and the relationship between these two indicators is discussed.
[DOI: 10.1121/1.2735109]

PACS number(s): 43.50.Jh, 43.55.Ti, 43.55.Rg, 43.20.El [NX] Pages: 302–313

I. INTRODUCTION

The airborne acoustic transmission performance of enclosures is strongly affected by the presence of openings. These openings might be designed purposely to allow for the circulation of matter through the structure or could be the result of bad assembly and/or mounting conditions. When the size of the opening is small compared to the acoustic wavelength, it is typically referred to as a leak while the term "opening" is generally used for larger dimensions. In this paper, the term *aperture* is used for both leaks and openings. During the acoustic design of enclosures, predicting the effect of an aperture on the overall acoustic transmission can become crucial, particularly when its size becomes important as the aperture may become the main transmission path as the frequency increases. Despite its importance, relatively few authors have dealt with this problem and there is consensus on which formulations work best for predicting the diffuse field transmission of an aperture. In previous studies,[1–15] apertures were considered as rectangular, slit, or circular shapes with negligible or finite thickness since these geometries are representative of most practical cases. In order to better understand the various proposed methods, this paper presents a comprehensive literature review and comparisons of the principal approaches available in the literature. The theory and approximations behind each approach are briefly presented, compared, and discussed. The usefulness and range of validity of these approaches are discussed and some remaining interrogations are stated. To help in answering some of these questions, a new general and rigorous formulation is then proposed to predict the transmission loss (TL) of rectangular and circular apertures. The developed tool is simple, user-friendly, efficient, and can easily be utilized to provide input data to Statistical Energy Analysis (SEA) based codes for instance. The approach convergence is investigated and the model is validated by comparison with existing models. Finally, the normal incidence and diffuse field sound transmission loss are compared for various configurations.

II. LITERATURE REVIEW

Most existing models dealing with the calculation of normal incidence and more recently oblique incidence plane wave sound TL of apertures consider rectangular, slit-shaped, or circular cross section of negligible or finite thickness. The

[a)]Invited researcher, Institut de Recherche Robert-Sauvé en Santé et en Sécurité du Travail, 505, Boul. de Maisonneuve Ouest, Montréal Quebec H3A 3C2, Canada; electronic mail: frasga@irsst.qc.ca

transmission through apertures with negligible thickness has been first studied by Bouwkamp,[1] Spence,[2] and Mulholland and Parbrook.[3] The main models dedicated to the acoustic transmission through finite thickness apertures are those of Gomperts,[4,7] Wilson and Soroka,[5] Sauter and Soroka,[6] Mechel,[8] Nomura's model,[9] Thomé,[10] Chen,[11] Park,[12] Serizawa,[13] Huang,[14] and Furue.[15] The first four authors assume a plane wave field within the circular and/or rectangular aperture.

Gomperts[4,7] relates the velocity potential in a circular or slit-shaped aperture to the velocity potential of spherical or cylindrical waves, respectively, to get the transmission coefficient in terms of trigonometric functions and "end corrections" associated to the aperture shapes. The author compares his model with experimental results obtained from measurements in two coupled reverberant rooms. In his calculation, it is assumed that the diffuse field transmission coefficient is twice the normal incidence one and the position of the aperture in the wall of the room is taken into account using reciprocity considerations. In the case of circular apertures, the model compares well to experimental measurements up to a value of $k_0a=1$ (a being the radius of the aperture). For slit-shaped cross sections, the author concludes that average trends of theoretical and experimental results are in good agreement, the largest discrepancies occurring around resonance frequencies associated to the aperture depth.

Wilson and Soroka,[5] Sauter and Soroka,[6] and Mechel[8] assume the field in the aperture to be the sum of two plane waves propagating in opposite directions. The boundary conditions over the surface of the apertures are imposed in an average sense which allows them to use the radiation impedance functions of rigid massless baffled pistons to account for the coupling between the aperture and the outer environment. Wilson and Soroka's together with Sauter and Soroka's models consider the normal incidence transmission loss of circular and rectangular apertures of finite thickness, respectively. Wilson and Soroka[5] successfully compared their approach with existing exact solutions available for the diffraction of sound at normal incidence by a finite thickness circular aperture. They also confronted their numerical simulations with experimental results obtained in reverberant rooms. It was found that even if the assumption of planes waves in the aperture is not true for wavelengths shorter than the dimension of the aperture, the experimental results indicate that the solution is exploitable up to $k_0a=8\pi$, the error being on average less than 2 dB. Sauter and Soroka[6] showed numerical results for the normal incidence transmission loss of rectangular cross sections for various aspect ratios and various ratios of depth to equivalent radius of aperture cross section. Experimental measurements were carried out in reverberant rooms and showed good agreement with the calculated normal incidence transmission loss for the tested configurations up to a value of $k_0a=5.6$ where a represents the equivalent radius of the rectangular cross section. Mechel[8] proposed similar theoretical expressions for a circular-shaped aperture filled with an absorbing material sealed by rigid masses at the aperture extremities and excited at oblique incidence. It was found, unexpectedly, that the diffuse field TL was equal to π times the normal incidence TL. The author also provides an analytical expression for the oblique incidence transmission loss of a slit-shaped aperture. Chen[11] investigated the prediction of the normal incidence and diffuse field transmission loss of slits. A combination of Rayleigh's formula and a wave number transform of the fields were used to solve the problem. This is similar to Mechel's work.[8] Measurements of the diffuse field sound transmission loss of several slits using a reverberant source room and a sound intensity probe in the receiver room were performed. These measurements were compared with both the normal incidence and diffuse field models for frequencies ranging from 100 to 1250 Hz. Differences between the calculation and the experiment are found to be smaller than 4 dB. It is interesting to note that larger discrepancies are observed when the diffuse field theory is used compared to the normal incidence which suggests questionable results.

Independent of the excitation, the assumption of a one-dimensional field within the aperture is valid below its cutoff frequency, imposed by the lateral dimensions of the aperture, for locations inside the aperture not too close from the boundaries (front and rear faces). As a consequence, the approach may be limited in frequency as soon as the aperture is not small compared to the acoustic wavelength; that is, if transverse motion occurs. For oblique incidence, the pressure field varies over the surface of the aperture and transverse modes can contribute to the transmitted pressure field. Several authors have proposed models to deal with this issue. Thomé[10] proposed a simplified model to account for the spatial variation of the pressure field across rectangular and slit-shaped aperture due to oblique incidence excitation. He assumes that the oblique incidence sound field modifies the radiation directivity rather than the amount of transmitted energy, this hypothesis being especially true at low frequencies. A discretization scheme of the aperture surface in multiple small elementary pistons vibrating with the same absolute value of the velocity but with different phases imposed by the incident sound field is used. Sauter and Soroka's model is then applied to each of those elementary pistons and it is assumed that the phase relationships are preserved after going through the aperture. This amounts to calculating a radiation impedance modulated by the incident wave. Experimental results regarding the transmission loss measured between a reverberant box and an anechoic room proved to match globally in a satisfactory way the results predicted by the model.

Park and Eom's model[12] has been developed for the calculation of the oblique incidence sound transmission loss and reflection coefficient of a rectangular aperture of finite thickness taking into account the three-dimensional (3D) nature of the sound field within the aperture. In their approach, the acoustic field inside the aperture is expanded in terms of propagating and evanescent modes of a duct with rigid walls. The scattered fields are expressed in terms of their spatial Fourier transform. The Fourier transform of boundary conditions at the front and back face of the aperture together with a projection on the modal basis of the duct lateral modes leads to a linear system of equations in terms of modal contributions. The difficulty lies in the calculation of the modal

impedance radiation matrix in the wave number domain. Using a change of variables, the authors propose an improved numerical algorithm to calculate this impedance matrix. Once the system is solved, the reflection and transmission coefficients together with the diffraction patterns are obtained. The authors present numerical results at normal incidence for several apertures sizes which are in good agreement with Sauter and Soroka's.

Nomura's model[9] proposes an exact solution for the normal incidence transmission loss of a circular aperture in a thick wall taking into account the 3D wave field in the aperture using Weber-Schafheitlin integrals. Serizawa and Hongo's model[13] is based on the same methodology as Normura but applies to rectangular apertures. Their model is based on the use of a modal expansion for the sound field within the aperture and a series of wave functions for the sound fields radiated at the front and the rear of the aperture. The problem is then solved by the Kobayashi potential method. This technique consists in using the discontinuous properties of the Weber-Schafheitlin integrals to enforce the Neuman-type boundary conditions for the wave functions and projecting the remainder of the boundary conditions on each side of the aperture on a basis of Jacobi's polynomials and trigonometric functions. Matrix equations are then obtained in terms of the amplitude of the function sets. The authors derive the transmission loss from the transmitted acoustic power calculated in the far field. They successfully compared their results with Park and Eom's and Sauter and Soroka's models in the case of normal incidence only. Also presented are results for the diffracted patterns and velocity distributions on the aperture for different incidence angle of the exciting plane wave. Despite its relative completeness, this model is however difficult to implement because of the calculation of double infinite series and integrals of Bessel functions.

General numerical techniques such as boundary element methods (BEM) can be used to predict the diffracted acoustic field by complicated shaped apertures. Huang *et al.* solved Helmholtz integral equation to calculate the sound diffracted through circular apertures in soft and hard plane baffles using boundary elements.[14] Furue combined both geometrical acoustics and Kirchhoff's diffraction formula to calculate the high frequency sound field radiated from a baffled rectangular aperture excited by a point source set in an anechoic chamber. At low frequencies, he applied the BEM to calculate the sound field diffracted by the extremity of an unbaffled circular aperture excited by a point source.[15] These BEM methods, though accurate, require special numerical treatments of singularities and make use of meshing tools together with postprocessing modules. For simple aperture shapes, these techniques are more cumbersome to use than the previous techniques.

To summarize the above presented discussion, Tables I–III recapitulate the different analytical or semi-analytical models classically used in the design of noise control solutions. Expressions for the transmission loss are provided only when simple enough to fit in the tables. The aperture geometry is illustrated in Fig. 1.

Four main conclusions can be drawn from the literature review.

1. *Range of validity in the frequency domain*. As mentioned before, for models assuming plane wave propagation in the aperture, the sound transmission behavior is only correctly represented up to the aperture cut-off frequency. For leaks (small dimensions compared to the acoustic wavelength), these models can be sufficient except at higher frequencies. However, for openings (large dimensions compared to the acoustic wavelength), this becomes obviously incorrect. However, in practice, these models are often used beyond their domain of validity. While some authors[10,12–15] have considered the 3D nature of the sound field within the aperture to circumvent this issue, their models were either debatable,[10] too time consuming,[12] very complicated to implement,[13] required meshing tools,[14,15] or did not provide results regarding diffuse field sound transmission loss.[12–15] There is clearly a lack for a model that would predict accurately and rapidly the diffuse field sound transmission loss in the acoustical frequency range whatever the size of the aperture.
2. *Normal incidence vereus diffuse field predictions*. Based on the comparison between normal incidence calculation and diffuse field sound transmission loss measurements, many authors suggest that it is sufficient to use the normal incidence TL with an acceptable error or to use the normal incidence TL with a correction factor to simulate the diffuse field transmission loss. There is no real consensus on how the diffuse field TL should be related to the normal incidence. For example, for circular apertures, some authors suggest that the diffuse field TL can be taken equal to the normal incidence transmission loss based on experimental observations (Wilson and Soroka[5]) while others suggest that it is equal to the normal incidence one minus a 3 dB correction (Gomperts[4]). The same approximations are also found for rectangular shaped cross sections. To the authors' knowledge, no rigorous study has shown the validity of such approximations. In other words, to which extent is it relevant to use the normal incidence transmission loss (with or without corrections) to predict the diffuse field TL ?
3. *Efficient tool for broadband predictions*. Besides Thomé's, Park's, and Serizawa's works,[10–12] which suffer from the aforementioned drawbacks, there is no simple efficient numerical model for the prediction of the diffuse field sound transmission loss of both circular and rectangular apertures in the frequency range covered in most noise control applications (100–5000 Hz).
4. *Large apertures (openings)*. There are few numerical results regarding the diffuse field sound transmission loss of openings (large apertures).

This paper is an attempt to address the four questions that came out from the aforementioned conclusions by (i) proposing a general, rigorous, and efficient numerical method to predict the diffuse field sound transmission loss of apertures of rectangular and circular cross section, (ii) providing the reader with numerical results regarding the diffuse field TL in relation to the normal incidence case for various geometrical configurations, and (iii) discussing the relevance of using such a sophisticated model compared to more clas-

TABLE I. Summary of the main analytical models of aperture sound transmission loss (plane wave field in aperture).

Model	*Expressions for the sound transmission coefficient*	
Mechel (1986) (Ref. 8)	$\tau(\theta_i,\varphi_i)=\frac{Z_0}{\cos\theta_i}\Re[\hat{Z}_{R2}]p_g^2\lvert\frac{\hat{Z}_a}{\hat{Z}_a(\hat{Z}_1+\hat{Z}_2)\cos(\hat{k}_a d)+j(\hat{Z}_a^2+\hat{Z}_1\hat{Z}_2)\sin(\hat{k}_a d)}\rvert^2$ $\hat{Z}_i=\hat{Z}_{Ri}+j\omega m_i i=1,2$	
	Circular	
	$p_g=\frac{4J_1(k_0 a\sin\theta_i)}{k_0 a\sin\theta_i}$, $\hat{Z}_a=\hat{Z}_f$, $\hat{k}_a=\hat{k}_f$, $\hat{Z}_{Ri}=Z_0\hat{Z}_{circ}$, $\tau_d=\frac{2(1-\cos\theta_{\lim})}{\sin^2\theta_{\lim}}\tau(0)$	
	Slit	
	$p_g=2\,\mathrm{sinc}(k_y a)$, $\hat{Z}_a=\frac{\hat{Z}_f}{\cos\theta_f}$, $\hat{k}_a=\hat{k}_f\cos\theta_f$, $\hat{Z}_{Ri}=Z_0\hat{Z}_{slit,m}$ and $\tau_d\Rightarrow$numerical calculation	
Wilson and Soroka (1965) (Ref. 5) Sauter and Soroka (1970) (Ref. 6)	$\tau(0)=\frac{4R_0}{[A^2(k,d,R_0,X_0)+B^2(k,d,R_0,X_0)]}$	$A(\hat{k}_f,d,R_0,X_0)=2R_0(\cos\hat{k}_f d-X_0\sin\hat{k}_f d)$
	$\tau_d\approx\tau(0)$	$B(\hat{k}_f,d,R_0,X_0)=(R_0^2-X_0^2+1)\sin\hat{k}_f d+2X_0\cos\hat{k}_f d$
	Circular (Wilson and Soroka)	
	$R_0+jX_0=\hat{Z}_{circ}$	
	Rectangular (Sauter and Soroka)	
	$R_0+jX_0=\hat{Z}_{rect}\Rightarrow$numerical calculation	
Gomperts and Kilhman (1964,1967) 4 and 7	Circular	Normal incidence: m=8, n=2
	$\tau=\frac{m(k_0a)^2}{n^2\sin^2 k_0 a\left(\frac{d}{a}+2\varepsilon\right)+4(k_0a)^4}$	Diffuse field: m=16, n=2 (aperture in center of wall), n=1 (aperture at the intersection between to walls), n=1/2 (corner)
		$\varepsilon=\alpha/a$ where α is a tabulated correction length
	Slit	Normal incidence: m=4, n=1
	$\tau=\frac{2mk_0b}{2n^2\sin^2 2k_0 b\left(\frac{d}{2b}+2\varepsilon\right)+2(2k_0b)^2}$	Diffuse field: m=8, n=1 (aperture in middle of wall), n=1/2 (aperture at the intersection between to walls)
		$\varepsilon=\alpha/2b=1/\pi(\ln 8/2k_0b-0.57722)$
Chen (1995) (Ref. 11)	Slit	$\hat{F}=\left(\left[\frac{k_z}{k_0}R_0+(1-k_z b)\right]\cos k_z d+X_0\left(\frac{k_z}{k_0}\right)^2\sin k_z d\right)$
	$\tau(\theta_i,\phi_i)=\frac{4R_0\lvert\hat{K}\rvert^2}{\cos\theta_i\left(1-\frac{1}{2}(2k_z b)\right)^2\lvert\hat{F}\rvert^2}$	$+j\left(\frac{k_z}{k_0}X_0\cos k_z d-X_0\left(\frac{k_z}{k_0}\right)^2 k_z b\sin k_z d\right)$
		$\hat{K}=\left(\frac{k_z}{k_0}\hat{F}_r-\frac{k_z}{k_0}\hat{E}_r\cos k_z d+\hat{E}_i k_z b\sin k_z d\right)+j(\hat{E}_r k_z b\sin k_z d)$
		$\hat{E}=(k_z R_0-1)+jX_0R_0+jX_0=\hat{Z}_{\mathrm{slit},c}$

sical models based on normal incidence calculations. The proposed model is based on the description of the sound field inside the aperture in terms of propagating and evanescent transverse acoustic modes. The radiation of the aperture is accounted for by using a modal radiation impedance matrix. The coupled problem is then solved in terms of modal contribution factors from which the transmission loss is evaluated.

The rest of the paper is organized as follows. First, the theory behind the model is presented. Second, the approach convergence is investigated and a numerical validation based on comparisons with the models discussed earlier is performed. Finally, normal incidence and diffuse field sound transmission losses are compared for various geometrical configurations.

III. THEORY

Consider a rectangular or circular aperture of area S and depth d inserted into a rigid planar baffle and excited acoustically by an oblique plane wave with incidence angles (θ_i,φ_i) as illustrated in Fig. 1.

Let p_i be the incident acoustic pressure (of amplitude A_i), and p_b the blocked acoustic pressure, p_r the acoustic pressure radiated by the front face of the aperture (excitation side), and p_2 the acoustic pressure radiated in medium 2 by the backface of the aperture. The acoustic field inside the aperture is denoted $p_f=p_{f1}+p_{f2}$.

The total acoustic field $\hat{p}_1$ in medium 1 is given by

$$\hat{p}_1=\hat{p}_i+\hat{p}_R+\hat{p}_r=\hat{p}_b+\hat{p}_r, \tag{1}$$

where the convention $p(t)=\hat{p}e^{j\omega t}$ is used.

The blocked pressure satisfies the condition $\partial\hat{p}_b/\partial z|_{z=0}=0$ and is given by

$$\hat{p}_b=2\hat{A}_i e^{-jk_0(\sin\theta_i\cos\varphi_i x+\sin\theta_i\sin\varphi_i y)}\cos(k_0\cos\theta_i z) \tag{2}$$

The pressure radiated in medium 1 can be written as

$$\hat{p}_r(M)=\int_{S_1}G(M,M_0)\frac{\partial\hat{p}_r}{\partial n}(M_0)dS(M_0), \tag{3}$$

where n denotes the outward normal to medium 1 and where G is the baffled Green's function given by $G(M,M_0)$

TABLE II. Summary of the main analytical or semianalytical models of aperture sound transmission loss (three-dimensional wave field in aperture). (MFA: Modal Field in Aperture, LPA: Localized plane wave in Aperture)

Model	*Expressions for the sound transmission coefficient*
Nomura (1960) (Ref. 9) MFA	Circular, normal incidence. See reference.
Thomé (1991) (Ref. 10) LPA	Rectangular $\tau(\theta_i,\phi_i)=\frac{Z_0}{\cos\theta_i}\Re[\hat{Z}_R]\frac{\lvert 2\hat{Z}_f\rvert^2}{\lvert\hat{D}\rvert^2}$ $\hat{Z}_R=\frac{jk_0Z_0}{8\pi ab}\int_{-a}^{a}\int_{-b}^{b}\int_{-a}^{a}\int_{-b}^{b}e^{-jk_0(\sin\theta_i\cos\varphi_i x_0+\sin\theta_i\sin\varphi_i y_0)}\frac{e^{-jk_0R}}{R}e^{jk_0(\sin\theta_i\cos\varphi_i x+\sin\theta_i\sin\varphi_i y)}dx_0dy_0dxdy$ $\hat{D}=[2\hat{Z}_f\hat{Z}_{R,p}\cos\hat{k}_f d+j(\hat{Z}_f^2+\hat{Z}_{R,p}^2)\sin\hat{k}_f d]$ where $\hat{Z}_{R,p}=Z_0\hat{Z}_{rect}$
Park and Eom (1997) (Ref. 12) MFA	$\begin{bmatrix}[\Psi_1] & [\Psi_2]\\ [\Psi_3] & [\Psi_4]\end{bmatrix}\begin{Bmatrix}\{\hat{A}_M\}\\ \{\hat{B}_M\}\end{Bmatrix}=\begin{Bmatrix}\{\hat{F}_M\}\\ \{0\}\end{Bmatrix}$ $\Psi_{1,pqmn}=\varepsilon_m\varepsilon_n ab\cos(k_{mn}d)\delta_{pm}\delta_{qn}+j\frac{(ab)^2}{4\pi^2}k_{pq}\sin(k_{pq}d)\hat{I}_{pqmn}$ $\Psi_{2,pqmn}=-\varepsilon_m\varepsilon_n ab\sin(k_{mn}d)\delta_{pm}\delta_{qn}+j\frac{(ab)^2}{4\pi^2}k_{pq}\cos(k_{pq}d)\hat{I}_{pqmn}$ $\Psi_{3,pqmn}=\varepsilon_m\varepsilon_n ab\,\delta_{pm}\delta_{qn},\ \Psi_{4,pqmn}=-j\frac{(ab)^2}{4\pi^2}k_{pq}\hat{I}_{pqmn}$ $\hat{F}_{mn}=2ab\hat{G}_m(-k_xa)\hat{G}_n(-k_yb),\ \hat{G}_p(u)=-ju\frac{(-1)^p e^{ju}-e^{-ju}}{u^2-\left(p\frac{\pi}{2}\right)^2}$ $k_{pq}=\sqrt{k_0^2-\left(\frac{p\pi}{2a}\right)^2-\left(\frac{q\pi}{2b}\right)^2}$ $\tau(\theta_i,\phi_i)=-\frac{1}{4k_0\cos\theta_i}\Im[\sum\varepsilon_m\varepsilon_n\hat{k}_{mn}^*\hat{A}_{mn}\hat{B}_{mn}^*]$ $\hat{I}_{pqmn}=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\frac{^{(m,n)}}{\sqrt{k_0^2-k_x^2-k_y^2}}\hat{G}_p(k_xa)\hat{G}_q(k_yb)\hat{G}_m(-k_xa)\hat{G}_n(-k_yb)dk_xdk_y$
Serizawa and Hongo (KP) (2002) (Ref. 13) MFA	Rectangular, oblique incidence. Expressions too lengthy to be included in the paper. See reference.

$=e^{-jk_0R}/2\pi R$ with $R=\sqrt{(x-x_0)^2+(y-y_0)^2+z^2}$ and $\partial\hat{p}_r/\partial n=\partial\hat{p}_r/\partial z$.

The aperture can be filled with an acoustic material of characteristic impedance $\hat{Z}_f$ and complex propagation constant $\hat{k}_f$. For simple shapes such as rectangular and circular cross sections, the acoustic field inside the aperture can be expanded analytically in terms of propagating and evanescent modes:

$$\hat{p}_f(x,y,z)=\sum_P(\hat{A}_P e^{-j\hat{k}_P(z-d)}+\hat{B}_P e^{j\hat{k}_P(z-d)})\phi_P(x,y) \qquad (4)$$

where

$$\hat{k}_P=\hat{k}_{pq}=\sqrt{\hat{k}_f^2-\left(\frac{p\pi}{2a}\right)^2-\left(\frac{q\pi}{2b}\right)^2},$$

$$\phi_P(x,y)=\phi_{pq}(x,y)=\cos\left(\frac{p\pi}{2a}(x+a)\right)\cos\left(\frac{q\pi}{2b}(y+b)\right)$$

for rectangular cross section of size $2a\times 2$ and

$$\hat{k}_P=\hat{k}_{pq}=\sqrt{\hat{k}_f^2-\frac{\lambda_{pq}^2}{a^2}},$$

TABLE III. Expressions of normalized radiation impedances of apertures of different cross sections. $J_1(x)$: Bessel function of order 1. $S_0(x)$, $S_1(x)$: Struve function of order 0 and 1 respectively. $H_0^{(2)}(u)$, $H_1^{(2)}(u)$: zeroth and first order Hankel function of the second kind respectively. $k_x=k_0\sin\theta_i\cos\varphi_i$, $k_y=k_0\sin\theta_i\sin\varphi_i$, $k_z=k_0\cos\theta_i$, $Z_0=\rho_0c_0$.

Aperture shape	Normalized radiation impedance $Z_R/Z_0=R_0+jX_0$
Circular piston Z_{circ}	$1-\frac{J_1(2k_0a)}{k_0a}+j\frac{S_1(2k_0a)}{k_0a}$
Rectangular piston Z_{rect}	$\frac{jk_0}{8\pi ab}\int_{-a}^{a}\int_{-a}^{a}\int_{-b}^{b}\int_{-b}^{b}\frac{e^{-jk_0R}}{R}dx_0dy_0dxdy$
Slit (Mechel) $Z_{\text{slit},m}$	$2k_0b\left[H_0^{(2)}(u)+\frac{\pi}{2}[H_1^{(2)}(u)S_0(u)-H_0^{(2)}(u)S_1(u)]-\frac{1}{u}H_1^{(2)}(u)+\frac{2j}{\pi u^2}\right]\quad u=2b\sqrt{k_0^2-k_x^2}$
Slit (Chen) $Z_{\text{slit},c}$	$\frac{2k_z}{\pi}\left[\int_0^{k_z}\frac{(1-\cos(2k_yb))}{k_y^2(k_z^2-k_y^2)^{\frac{1}{2}}}dk_y+j\int_{k_z}^{+\infty}\frac{(1-\cos(2k_yb))}{k_y^2(k_y^2-k_z^2)^{\frac{1}{2}}}dk_y\right]$

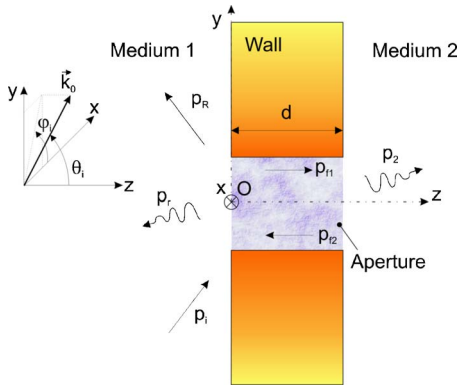


FIG. 1. (Color online) Configuration of the problem.

$$\phi_P(x,y) = \phi_{pqs}(r,\gamma) = J_p\left(\frac{\lambda_{pq} r}{a}\right)\sin\left(p\gamma + s\frac{\pi}{2}\right)$$

for a circular cross section of radius a, λ_{pq} being the zeroes of the derivative of Bessel function of order p ($J'_p(\lambda_{pq})=0$) and s being a symmetry index equal to 0 or 1.

The transmitted pressure field (medium 2) satisfies

$$\hat{p}_2(M) = \int_{S_2} G(M,M_0)\frac{\partial \hat{p}_2}{\partial n}(M_0)dS(M_0) \tag{5}$$

with $\partial\hat{p}_2/\partial n = -\partial\hat{p}_2/\partial z$.

In order to simplify the formulation, it is supposed that mediums 1 and 2 are identical ($\rho_1=\rho_2=\rho_0$ and $c_1=c_2=c_0$). Then, the boundary conditions on both sides of the aperture read

$$\hat{p}_1 = \hat{p}_f \quad \text{over } S_1 \text{ at } z=0,$$

$$\frac{1}{j\omega\rho_0}\frac{\partial \hat{p}_r}{\partial z} = \frac{1}{j\omega\tilde{\rho}_f}\frac{\partial \hat{p}_f}{\partial z} \quad \text{over } S_1 \text{ at } z=0,$$

$$\hat{p}_2 = \hat{p}_f \quad \text{over } S_2 \text{ at } z=d,$$

$$\frac{1}{j\omega\rho_0}\frac{\partial \hat{p}_2}{\partial z} = \frac{1}{j\omega\tilde{\rho}_f}\frac{\partial \hat{p}_f}{\partial z} \quad \text{over } S_2 \text{ at } z=d. \tag{6}$$

Taking the derivative of Eq. (4) with respect to z, using Eq. (6), multiplying all the equations by $\phi_M(x,y)$, and integrating over the surface of the front and backfaces of the aperture leads to

$$\hat{F}_M + \frac{S}{\hat{Z}_f\hat{k}_f}\sum_P \hat{k}_P(-\hat{A}_P e^{j\hat{k}_P d} + \hat{B}_P e^{-j\hat{k}_P d})\hat{Z}_{MP} = (\hat{A}_M e^{j\hat{k}_M d} + \hat{B}_M e^{-j\hat{k}_M d})N_M^2,$$

$$-\frac{S}{\hat{Z}_f\hat{k}_f}\sum_P \hat{k}_P(-\hat{A}_P + \hat{B}_P)\hat{Z}_{MP} = (\hat{A}_M + \hat{B}_M)N_M^2, \tag{7}$$

where one has used the fact that $\tilde{\rho}_f = \hat{Z}_f\hat{k}_f/\omega$ and $S=S_1=S_2$ where

$$\hat{F}_M = \int_{S_1} \hat{p}_b(x,y)\phi_M(x,y)dS,$$

$$\hat{Z}_{MP} = \frac{j\hat{k}_f\hat{Z}_f}{S}\int_S\int_S \phi_M(x,y)G(x,y,x_0,y_0)\phi_P(x_0,y_0) \times dS(M_0)dS(M),$$

$$N_M^2 = \int_S \phi_M^2(x,y)dS(M). \tag{8}$$

For a rectangular cross section, $N_{mn}^2 = \varepsilon_m\varepsilon_n ab$ with $\varepsilon_0=2$ and $\varepsilon_m=1$ otherwise and

$$\hat{F}_{mn} = 2\hat{A}_i\int_{-a}^{a}\int_{-b}^{b} e^{-jk_0(\sin\theta_i\cos\varphi_i x + \sin\theta_i\sin\varphi_i y)} \times\cos\left(\frac{m\pi(x+a)}{2a}\right)\cos\left(\frac{n\pi(y+b)}{2b}\right)dxdy = 2\hat{A}_i ab\hat{G}_m(-\alpha a)\hat{G}_n(-\beta b) \tag{9}$$

with $\alpha = k_0\sin\theta_i\cos\varphi_i$ and $\beta = k_0\sin\theta_i\sin\varphi_i$ and

$$\hat{G}_m(u) = -ju\frac{(-1)^m e^{ju} - e^{-ju}}{u^2 - \left(m\frac{\pi}{2}\right)^2} \quad \text{if } m \neq 0. \tag{10}$$

For a circular cross section,

$$N_{mns}^2 = \frac{a^2\pi\varepsilon_m}{2}\left(1 - \left(\frac{m}{\lambda_{mn}a}\right)^2\right)J_m(\lambda_{mn}a)^2 \quad \text{if } n \neq 1. \tag{11}$$

If $n=1$, $N_{01s}^2 = \pi a^2$. and

$$\hat{F}_{mns} = \frac{4\pi\hat{A}_i j^m a}{\zeta_{mn}^2 - \xi^2}\sin\left(m\varphi_i + s\frac{\pi}{2}\right)[\xi J_m(a\zeta_{mn})J_{m-1}(a\xi) - \zeta_{mn}J_{m-1}(a\zeta_{mn})J_m(a\xi)] \tag{12}$$

where $\zeta_{mn} = \lambda_{mn}/a$ and $\xi = -k_0\sin\theta_i$.

$\hat{Z}_{MP}$ is the aperture cross modal radiation impedance between modes M and P. For a rectangular cross section, it is given by

$$\hat{Z}_{mnpq} = \frac{j\hat{k}_f\hat{Z}_f}{4ab}\int_{-a}^{a}\int_{-b}^{b}\int_{-a}^{a}\int_{-b}^{b}\cos\left(\frac{m\pi(x+a)}{2a}\right) \times\cos\left(\frac{n\pi(y+b)}{2b}\right)G(x,y,x_0,y_0) \times\cos\left(\frac{p\pi(x_0+a)}{2a}\right)\cos\left(\frac{q\pi(y_0+b)}{2b}\right)dx_0dy_0dxdy \tag{13}$$

with

$$G(x,y,x_0,y_0) = \frac{e^{-jk_0\sqrt{(x-x_0)^2+(y-y_0)^2}}}{2\pi\sqrt{(x-x_0)^2+(y-y_0)^2}}.$$

This expression can be efficiently calculated using a change of variables[16] (see also Ref. 17) which transforms the quadruple integral into a double integral that can be resolved with a Gauss integration scheme.

For a circular cross section, the modal radiation impedance terms are given by

$$\hat{Z}_{mnspqt} = \frac{j\hat{k}_f\hat{Z}_f}{\pi a^2}\int_0^{2\pi}\int_0^a\int_0^{2\pi}\int_0^a J_m\left(\frac{\lambda_{mn}r}{a}\right) \times\sin\left(m\gamma + s\frac{\pi}{2}\right)G(r,\gamma,r_0,\gamma_0)J_p\left(\frac{\lambda_{pq}r_0}{a}\right) \times\sin\left(p\gamma_0 + t\frac{\pi}{2}\right)r_0 dr_0 d\gamma_0 r dr d\gamma \tag{14}$$

with

$$G(r,\gamma,r_0,\gamma_0) = \frac{e^{-jk_0\sqrt{r^2+r_0^2-2rr_0\cos(\gamma-\gamma_0)}}}{2\pi\sqrt{r^2+r_0^2-2rr_0\cos(\gamma-\gamma_0)}}.$$

This quadruple integral can be reduced to a simple one if the problem is solved in the wave number domain:

$$\hat{Z}_{mnspqt} = \frac{\hat{Z}_f\hat{k}_f k_0}{S}\pi\varepsilon_m\delta_{mp}\delta_{st}^{(m)}\int_0^{+\infty}\frac{\tau}{\sqrt{1-\tau^2}}D_{mn}(\tau)D_{pq}(\tau)d\tau \tag{15}$$

where $\delta_{st}^{(m)} = \delta_{st}$ *if* $m \neq 0$ and $\delta_{st}^{(0)} = s\delta_{st}$ ($\delta_{st}^{(0)} = 0$ *if* $s = 0$ and $\delta_{st}^{(0)} = 1$ *if* $s = 1$),

$$D_{mn}(\tau) = \int_0^a J_m(\tau k_0 r)J_m\left(\frac{\lambda_{mn}}{a}r\right)rdr = \frac{a}{\tau^2k_0^2 - \left(\frac{\lambda_{mn}}{a}\right)^2}\left[\frac{\lambda_{mn}}{a}J_m(\tau k_0 a)J_{m-1}(\lambda_{mn}) - \tau k_0 J_{m-1}(\tau k_0 a)J_m(\lambda_{mn})\right]. \tag{16}$$

The derivation is provided in Ref. [18]. Alternatively, this result can be obtained using Zorumski's technique[19] based on Sonine's infinite integral and Neumann's addition theorem. Combining all the equations together, the following linear system is finally obtained:

$$\begin{bmatrix}[\Psi_1] & [\Psi_2]\\ [\Psi_3] & [\Psi_4]\end{bmatrix}\begin{Bmatrix}\{\hat{A}_M\}\\ \{\hat{B}_M\}\end{Bmatrix} = \begin{Bmatrix}\{\hat{F}_M\}\\ \{0\}\end{Bmatrix} \tag{17}$$

with

$$\Psi_{1,MP} = N_M^2 e^{j\hat{k}_M d}\delta_{MP} + \frac{S}{\hat{Z}_f}\frac{\hat{k}_P}{\hat{k}_f}e^{j\hat{k}_P d}\hat{Z}_{MP}, \tag{18}$$

$$\Psi_{2,MP} = N_M^2 e^{-j\hat{k}_M d}\delta_{MP} - \frac{S}{\hat{Z}_f}\frac{\hat{k}_P}{\hat{k}_f}e^{-j\hat{k}_P d}\hat{Z}_{MP}, \tag{19}$$

$$\Psi_{3,MP} = N_M^2\delta_{MP} - \frac{S}{\hat{Z}_f}\frac{\hat{k}_P}{\hat{k}_f}\hat{Z}_{MP}, \tag{20}$$

$$\Psi_{4,MP} = N_M^2\delta_{MP} + \frac{S}{\hat{Z}_f}\frac{\hat{k}_P}{\hat{k}_f}\hat{Z}_{MP}. \tag{21}$$

Once the system has been solved, the transmitted acoustic power for oblique incidence excitation can be calculated as

$$\Pi^t(\theta_i,\varphi_i) = \frac{1}{2}\Re\left[\int_{S_2}\hat{p}_2\hat{v}_{2,n}^* dS\right] = \frac{1}{2}\Re\left[-\frac{1}{\tilde{\rho}_f^*\omega}\int_{S_2}\sum_P(\hat{A}_P + \hat{B}_P)\phi_P(x,y) \times\sum_M\hat{k}_M^*(-\hat{A}_M^* + \hat{B}_M^*)\phi_M(x,y)dS\right] = \frac{1}{2}\Re\left[-\frac{1}{\tilde{\rho}_f^*\omega}\sum_M N_M\hat{k}_M^*\hat{C}_M\hat{D}_M^*\right] \tag{22}$$

with $\hat{C}_M = (\hat{A}_M + \hat{B}_M)$ and $\hat{D}_M = (-\hat{A}_M + \hat{B}_M)$

The oblique incidence transmission coefficient $\tau(\theta_i,\varphi_i)$ is defined as

$$\tau(\theta_i,\varphi_i) = \frac{\Pi^t(\theta_i,\varphi_i)}{\Pi^{\text{inc}}(\theta_i,\varphi_i)}, \tag{23}$$

where $\Pi^{\text{inc}}(\theta_i,\varphi_i)$, the incident power, is given by

$$\Pi^{\text{inc}}(\theta_i,\varphi_i) = \frac{|\hat{A}_i|^2\cos\theta_i S}{2\rho_0 c_0}. \tag{24}$$

Combining Eq. (22) and (24), the oblique incidence transmission coefficient reads:

$$\tau(\theta_i,\varphi_i) = -\frac{\rho_0}{k_0\cos\theta_i\rho_f^*|\hat{A}_i|^2 S}\Re\left(\sum_M N_M\hat{k}_M^*\hat{C}_M\hat{D}_M^*\right). \tag{25}$$

The diffuse field sound transmission coefficient can then be calculated numerically using a Gauss integration scheme:

$$\tau_d = \frac{\int_0^{2\pi}\int_0^{\theta_{\text{lim}}}\tau(\theta_i,\varphi_i)\sin\theta_i\cos\theta_i d\theta_i d\varphi_i}{\pi\sin^2\theta_{\text{lim}}}, \tag{26}$$

where the limit angle θ_{lim} is taken equal to 78° in the following numerical examples. The TL is finally calculated using the sound transmission coefficient through the following relation:

$$\text{TL} = -10\log_{10}(\tau). \tag{27}$$

IV. NUMERICAL EXAMPLES

This section presents several examples to validate the proposed approach. In this work, a detailed convergence study has been performed by looking at the variation of several physical quantities such as the sound pressure inside the aperture, the radiation impedance, and the sound transmission loss as the number of modes was increased. In this section, however, for sake of conciseness the presentation is limited to a summary of the main findings. First the convergence of the proposed model is investigated by focusing on the transmission loss and on the truncation frequency which is directly related to the number of modes kept in the modal expansion. Second, a comparison with different existing

models for circular, rectangular, and slit-shaped apertures is proposed and finally, the predicted diffuse field and normal incidence transmission loss data are compared for various configurations of apertures.

A. Convergence of the approach

The convergence of the TL for both rectangular and circular cross sections has been investigated. Due to lack of space, only the main conclusions are provided here. First, this convergence study reveals that as expected, for excitation by oblique incidence plane waves or a diffuse field, the single plane wave is only capable of representing the sound wave field within the aperture below the first transverse mode eigenfrequency. Above this frequency, higher order modes should be accounted for in the modal expansion. Second, it has been found that keeping modes up to the maximum frequency of interest $f_{\max}$ suffices to capture the physics of the problem. The difference between the two solutions obtained with a truncation frequency of $f_{\max}$ and $2f_{\max}$ have been found to be less than 0.1 dB on the whole frequency range. Therefore, in the rest of the paper modes up to $f_{\max}$ will be retained in the modal expansion.

It is interesting to note that even at normal incidence, higher order modes can contribute to the surface radiation impedance below their cut-off frequency because of the modal radiation coupling between mode (0,0) and higher order (even,even) modes. Slight differences can be observed between the converged solution and the approximate solution where only the plane wave mode is kept in the expansion. These discrepancies are small and negligible from a practical standpoint. They are not shown here for sake of conciseness. If the cross modal coupling impedance terms are neglected, the results become identical to the plane mode case.

B. Validation of the model

Figure 2 shows the oblique incidence (θ_i=45°, φ_i=0°) transmission loss of (i) a rectangular aperture of dimensions $2a$=0.4 m, $2b$=0.2 m, d=0.2 m and (ii) a circular aperture with geometrical characteristics a=0.1 m and d=0.15 m as a function of normalized frequency k_0a_{eq}. a_{eq} is the equivalent radius defined as the radius of the circular aperture or as $\sqrt{4ab/\pi}$ for a rectangular aperture. Results obtained by a mixed finite element method-boundry element method (FEM-BEM) numerical model are compared with the present approach. In the FEM-BEM method, the aperture volume is discretized using acoustic Brick (Hexa8) elements and the coupling with the source and receiver semi-infinite fluids is taken into account with an admittance radiation matrix.[20] For the rectangular aperture, results calculated using Park and Eom model's[12] are also shown. An excellent agreement between the proposed approach and the other models is observed for both configurations.

A comparison of Mechel's model[8] which assumes the slit to be *of infinite length* and the presented model is shown in Fig. 3 for the case of a 1-cm-wide and 1-cm-thick slit. Results for a normal incidence plane wave, an oblique incident plane wave (θ_i=45°, φ_i=0°), and a diffuse field (θ_{lim}=78°) are displayed. In the present approach, the slit has been modeled as a rectangular opening of a length to width ratio of 400. Modes up to the maximum calculation frequency were retained in the modal expansion. An excellent agreement is obtained for all types of excitations. Figure 4 compares the normal incidence and diffuse field TL of a slit-shaped aperture as a function of normalized frequency for different aperture lengths. The sound transmission loss of the infinite slit calculated using Mechel's model is also plotted as a reference. As the aperture length increases, the TL converges, as expected, toward the infinite case. In order to be considered as infinite, the slit length must be at least of

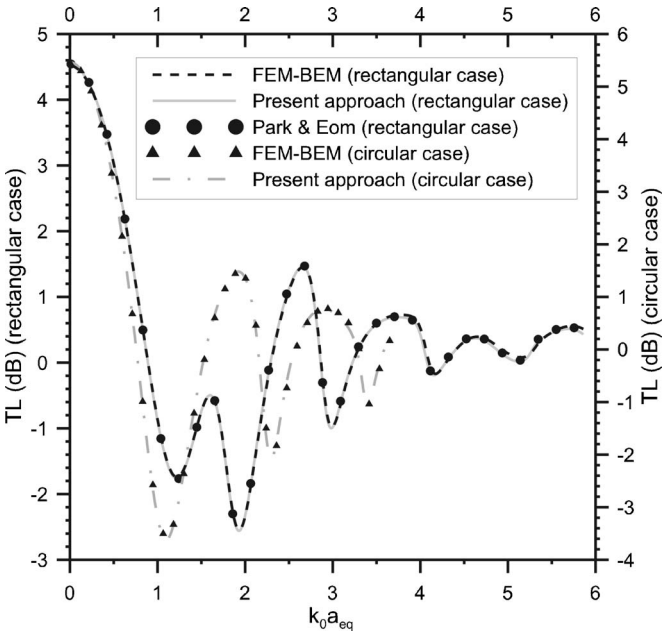


FIG. 2. Oblique incidence (θ_i=45°, φ_i=0°) transmission losses of a rectangular aperture (b/a=1/2, d/a=1) and a circular aperture (d/a=2/3)—comparison of the present approach with other models.

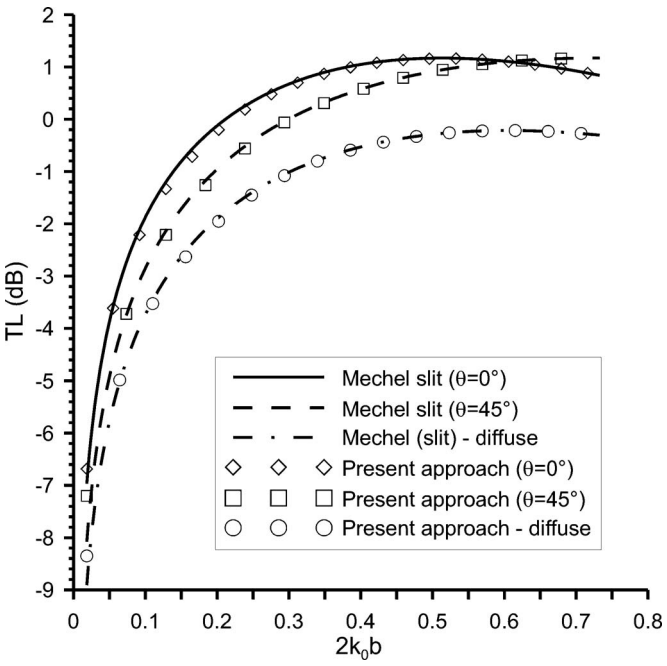


FIG. 3. Transmission loss of a slit—comparison of the present approach with Mechel's model.

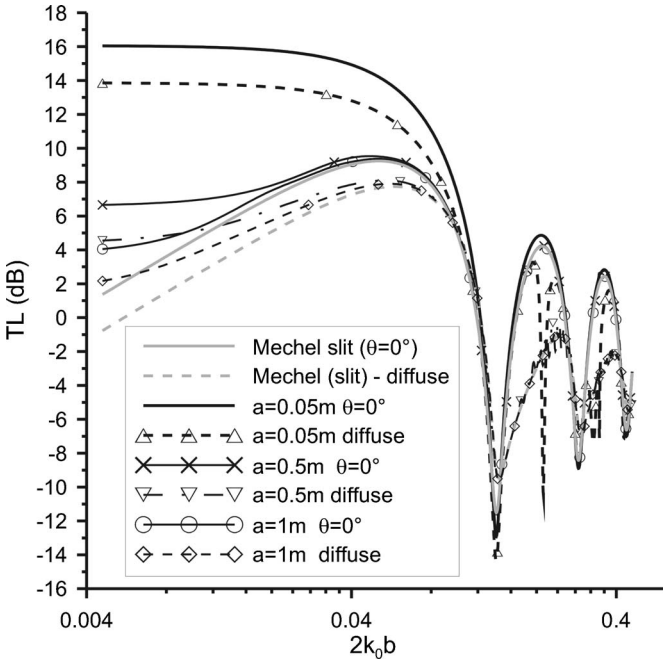


FIG. 4. Sound transmission loss of a finite length slit shaped aperture calculated using Mechel's model and the present approach—influence of the aperture length.

the order of one acoustic wavelength. If this criterion is met, the sound transmission loss becomes independent of the slit length, otherwise discrepancies are observed at low frequencies as seen in Fig. 5. Results in this Fig. 5 also reveal differences between the diffuse field and normal incidence TL data of around 2.2 dB in the low frequency range. This point is discussed later on in the paper.

Comparisons of the normal incidence transmission loss obtained with different models available in the literature are presented for rectangular and circular apertures in Figs. 5 and

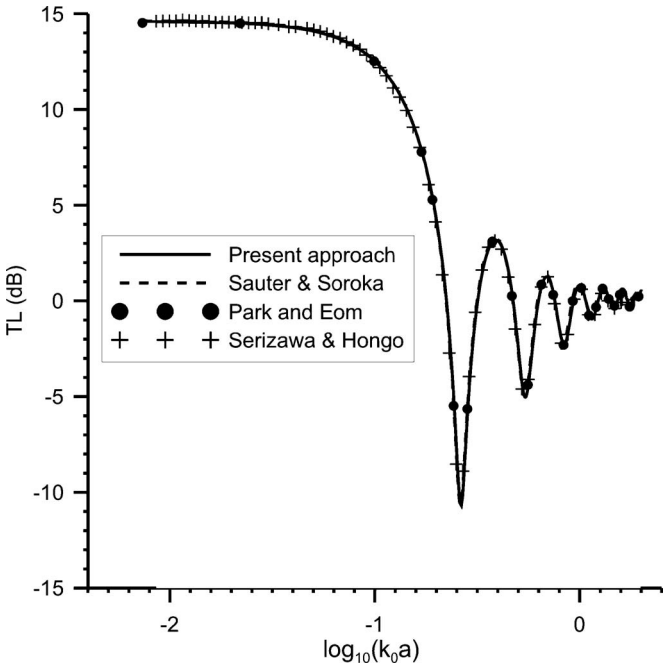


FIG. 5. Normal incidence transmission loss of a rectangular aperture $b/a=2$, $d/a=9.57$—comparison of the present approach with existing models.

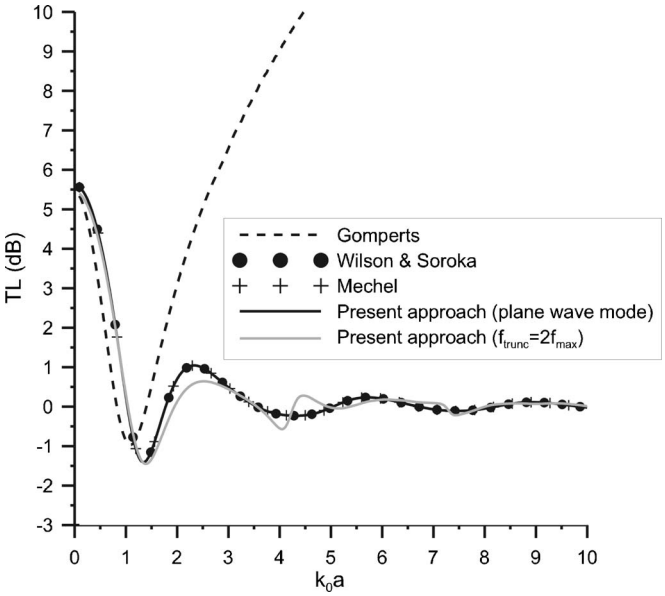


FIG. 6. Normal incidence transmission loss of a circular aperture $d/a=1$—comparison of the present approach with existing models.

6, respectively. Results in Fig. 5 are for a rectangular aperture with the following geometrical characteristics: $b/a=2$ and $d/a=9.57$. Results using Sauter and Soroka's model[6] [one-dimensional (1D) wave field in the aperture] together with Park and Eom's model[12] (3D wave field in the aperture) are displayed. In this work, Park and Eom's model was implemented using an improved integration scheme (see the discussion to follow). It is seen that there is an excellent agreement between the results obtained by the proposed approach and those provided by existing models. In the modal expansions considered in Park and Eom's model and the proposed model, modes up to a truncation frequency equal to the maximum frequency of calculation have been retained. When only mode (0,0) is kept in the expansions, all the curves become identical. Small differences are observed at high frequencies $[\log_{10}(k_0a)>0.2]$ between the 3D models and the 1D approach due to the cross modal coupling between mode (0,0) and higher order even-even modes. This point was discussed earlier. Note that Park and Eom's model requires the calculation of an indefinite integral. This integral must be truncated numerically and convergence problem may appear at lower frequencies especially if the integral upper bound is not chosen appropriately. This probably explains some of the discrepancies between this model and Serizawa's approach observed in the latter author's work.[12] In their paper, Park and Eom proposed a numerical algorithm which is not very efficient computationally. Note in passing that their algorithm can be easily considerably improved by using a Gauss integration scheme accounting for the singularity at hand. Figure 6 shows the normal incidence transmission loss of a circular aperture with the following geometrical characteristics: $d/a=1$ and $d/a=10$. Figure 6 displays the results for three models based on a description of a 1D wave field in the aperture, namely that of Gomperts,[4,7] Wilson and Soroka,[5] and Mechel[8] together with the proposed approach. With the exception of Gompert's model, an excellent agreement is found if one mode is retained in the modal expansion

TABLE IV. Geometrical characteristics of investigated circular and rectangular cross-section apertures.

	a (m)	b (m)	d (m)
Circular	0.00564	⋯	0.1
	0.169	⋯	0.001
	0.1262	⋯	0.1
Rectangular	0.005	0.005	0.1
	0.15	0.15	0.001
	0.25	0.05	0.1

of the proposed approach. It is seen that keeping only the plane wave mode in the modal expansion leads to small numerical differences compared to the converged solution involving all the modes having their eigenfrequencies contained in the frequency band $[0, f_{\max}]$, where $f_{\max}$ is the maximum frequency of the calculation spectrum. Similar to the rectangular case, these discrepancies can be explained by the cross modal coupling between the plane wave mode (0,1) and higher order $(0,n)$ modes. Note that these differences are negligible from a practical point of view. Note, as underlined by Wilson and Soroka,[5] that Gomperts's model[4,7] provides reliable results up to $k_0a=0.5$ but starts to deviate significantly afterwards.

C. Diffuse field versus normal incidence sound transmission loss

Here, the diffuse field and normal incidence transmission losses are compared as a function of normalized frequency for four different geometric configurations and two types of aperture cross sections. The calculations have been performed in the [100–5000] Hz frequency range and subsequently averaged in third-octave bands. Details of the configurations of interest are given in Table IV.

Figure 7 shows that for very small 0.1-m-thick apertures (leaks) there is no difference between the sound transmission losses of rectangular and circular cross sections for both excitations when plotted as a function of k_0a_{eq} (the circular and the rectangular cross sections having the same area). Also, there is an approximately 2 dB difference between the diffuse field and normal incidence TL which is almost constant over the covered frequency range (2.2 dB at low frequencies, 2 dB at high frequencies). Note that for leaks, it is sufficient to keep only the first plane wave mode inside the aperture since the higher order modes eigenfrequencies are well above the upper frequency limit of the calculation spectrum. The dependence of τ on the incidence angle (θ_i, φ_i) can then be shown to be simply inversely proportional to $\cos\theta_i$. It is then straightforward to show that there exists a simple relationship between the normal incidence $\tau(0)$ and diffuse field τ_d transmission coefficients. Indeed Eq. (26) yields

$$\tau_d = \frac{2(1-\cos\theta_{\text{lim}})}{\sin^2\theta_{\text{lim}}}\tau(0). \tag{28}$$

For $\theta_{\text{lim}}=78°$, one obtains $\text{TL}_d=\text{TL}(0)-2.2$ dB, which is in accordance with the previous observations.

Figure 8 shows the sound transmission loss of thin (1-mm-thick) large apertures as a function of normalized frequency. Again, both the circular and the rectangular cross sections have the same area. Similar acoustical behaviors are observed for both types of apertures but there are slight differences between the transmission losses of the two aperture shapes for a given excitation. In addition these frequency behaviors are very different from the previous case (negative diffuse field transmission loss at low frequencies; these nega-

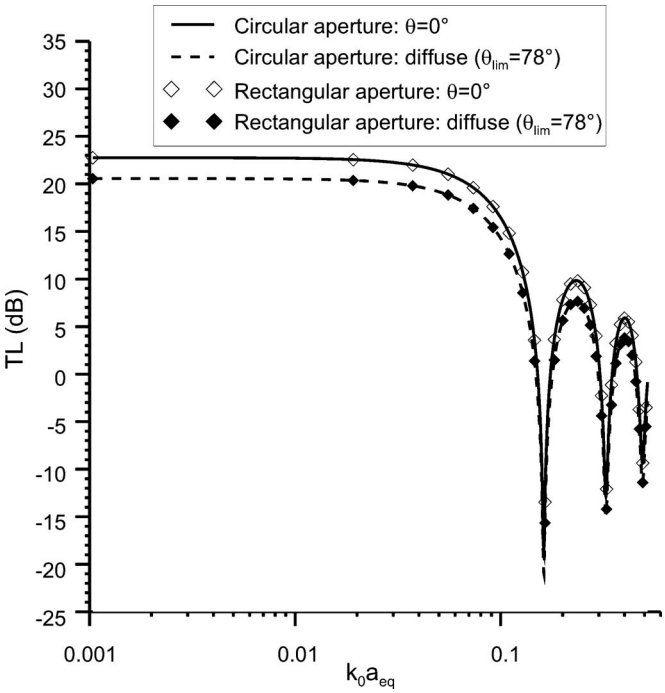


FIG. 7. Normal incidence and diffuse field sound transmission losses of a small circular ($a=0.00564$ m) and a small rectangular aperture ($a=b=0.005$ m) of depth 0.1 m calculated using the present approach as a function of normalized frequency.

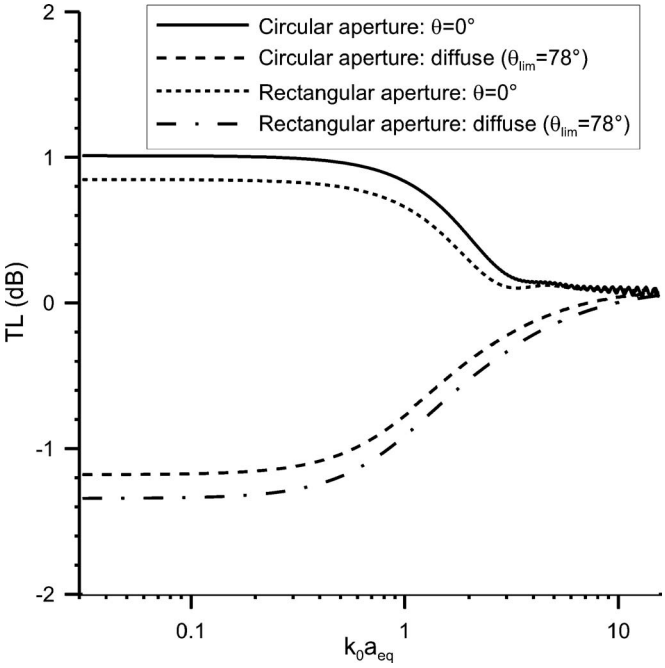


FIG. 8. Normal incidence and diffuse field sound transmission losses of a large circular ($a=0.169$ m) and a large rectangular aperture ($a=b=0.15$ m) of depth 0.001 m calculated using the present approach as a function of normalized frequency.

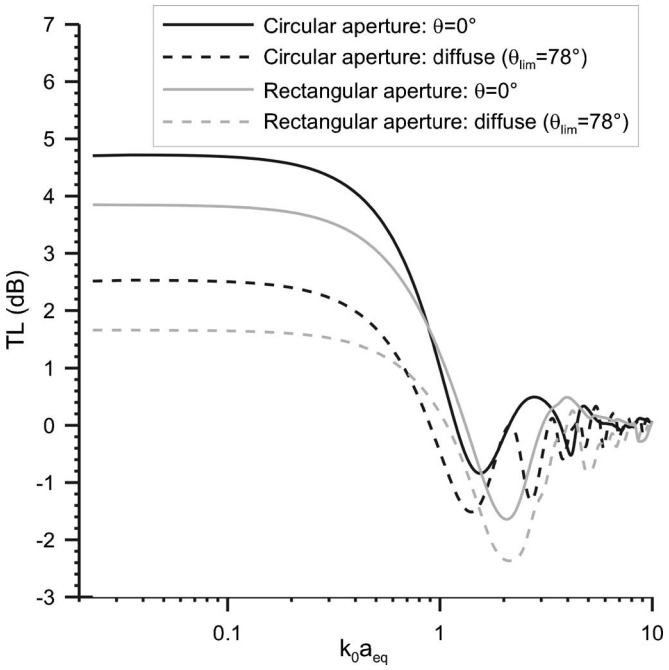


FIG. 9. Normal incidence and diffuse field sound transmission losses of a large circular (a=0.1262 m) and a large rectangular aperture (a=0.25, b =0.05 m) of depth 0.1 m calculated using the present approach as a function of normalized frequency.

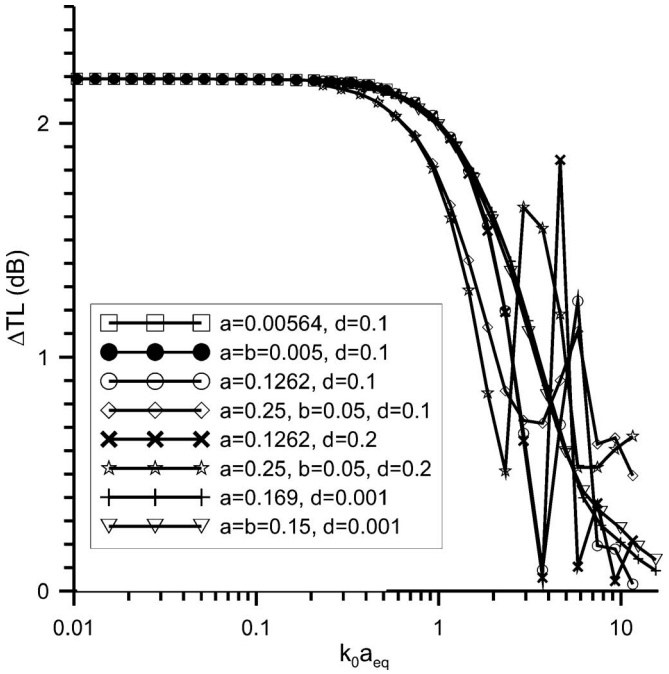


FIG. 10. Differences between normal incidence and diffuse field sound transmission losses for investigated apertures calculated using the present approach as a function of normalized frequency.

tive values are however a consequence of the definition of the TL). At low frequencies where the cross section is small compared to acoustic wavelength, there is a difference of 2.2 dB between the normal incidence and diffuse field sound transmission loss as shown in Eq. (28). For large sized apertures and values of $k_0a_{eq}>0.3$, this difference diminishes until it reaches 0 dB at high frequencies $k_0a_{eq}>6$. At high frequencies, the sound transmission losses are very close to 0 dB for both kinds of excitation. For thin leaks (not shown here for the sake of conciseness), TL_d is positive but very close to 0 dB over the whole frequency range. Note that for equivalent areas, the circular cross section is slightly more efficient than the rectangular one. Although differences are observed between the circular and rectangular shapes, from a practical point of view, the sound transmission loss of a thin large aperture is very low and can be considered equal to 0 dB within a 1.5 dB error on the whole frequency range.

In the two previous cases, the rectangular aperture had an aspect ratio of 1 (square cross section). Figure 9 shows the normal incidence and diffuse field sound transmission losses of a large rectangular aperture having an aspect ratio a/b =5 together with the results for a large circular aperture of equal area as a function of normalized frequency. As the aspect ratio of the rectangular cross section significantly differs from 1, the difference with the circular cross section increases for a given excitation. Also the local maxima and minima are not located at the same normalized frequencies. Similar to the previous configurations, a 2.2 dB difference is observed at low frequencies between TL_d and TL(0). At high frequencies $k_0a_{eq}>6$, the difference becomes less than 1 dB. Again, in this frequency zone, the sound transmission losses are very close to 0 dB for both excitations. In the midfrequency zone, resonant behaviors along the thickness and transverse modes are observed. In this region, the differences between TL_d and TL(0) remain within 2 dB.

Figure 10 summarizes the calculated differences between TL_d and TL(0) (third-octave band averaged) as a function of normalized frequency for all the cases presented in this section in addition to some additional configurations. At low frequencies $k_0a_{eq}<0.3$, the difference between TL_d and TL(0) is equal to 2.2 dB. At high frequencies $k_0a_{eq}>7$, the difference between TL_d and TL(0) is less than 1 dB. In between, the differences oscillate but remain less than 2 dB. If one is rather interested in narrow band transmission loss, differences have been found to be as high as 5 dB in the midfrequency zone for the configurations considered in this paper. These conclusions are in agreement with the observations made by previous authors[5,6] on the differences between experimental results (diffuse field excitation) and their normal incidence model. It is interesting to note that for rectangular cross sections, the asymptotic behavior is valid for aspect ratios which do not differ too much from 1. Indeed, for slits (large or small aspect ratio a/b), the single plane wave mode may not be sufficient anymore to describe the sound field in the aperture at low frequencies. Transverse modes have very low eigenfrequencies and should therefore be retained in the modal expansion. This can also be noticed when examining the analytical expression of the sound transmission loss provided by Mechel, which shows that the acoustic radiation impedance depends on the angle of incidence. Consequently, the asymptotic low frequency expression of the oblique incidence sound transmission loss is not inversely proportional to $\cos(\theta_i)$ and Eq. (28) does not hold anymore. However, for the investigated cases of slits, the difference between TL_d and TL(0) has been found to be less than 2 dB.

Note that the computational time per frequency point for a truncation frequency of $f_{\max}$ in the case of the configurations presented in this paper ranged from a few milliseconds to a few seconds on a typical laptop PC computer (Intel Pentium M, 1.86 GHz, 2 Gbytes memory). This makes the presented method both accurate and computationally efficient.

V. CONCLUSIONS

This paper presented a general and efficient numerical method to predict the diffuse field sound transmission loss of baffled apertures of rectangular and circular cross sections. The approach is based on an expansion of the wave field inside the aperture in terms of propagating and evanescent acoustic modes. The radiation of the aperture is accounted for using a modal radiation impedance matrix which is calculated numerically in an efficient way. The coupled problem is solved in terms of modal contribution factors from which the transmission loss can be evaluated. The developed tool is simple, user-friendly, efficient, does not require meshing tools such as those employed in BEM/FEM techniques, and can easily be utilized to provide input data to SEA codes for instance. Various numerical examples have been presented to validate the proposed approach (slit-shaped, rectangular, and circular cross section apertures). A parameters study has indicated that it is sufficient to keep modes up to the maximum frequency of interest to ensure the convergence of the solution convergence within a 0.1 dB error. Results regarding the diffuse field sound transmission loss have been provided and compared with the classical normal incidence sound transmission loss models. The following conclusions have been drawn.

1. At low frequencies ($k_0 a_{\text{eq}} < 0.3$), the differences between TL_d (diffuse field TL) and TL(0) (normal incidence TL) is 2.2 dB if a value of the limit angle for the diffuse field integration is 78° and 3 dB if 90°.
2. At high frequencies ($k_0 a_{\text{eq}} > 6-7$), the difference between TL_d and TL(0) tends toward 0 dB whatever the aperture shape and depth.
3. Between these two limits, for the cases of aperture investigated here (dimensions between approximately 0.3 m and 1 cm of equivalent radius), the differences between the calculated third-octave averaged TL_d and TL(0) have been found to be less than 3 dB in the frequency range [100–5000 Hz]. If one is rather interested in narrow band transmission loss, differences may reach locally up to 5 dB in the midfrequency zone.
4. The slit transmission loss is independent of its length provided that it is larger than one acoustic wavelength.
5. General conclusion 1 is not strictly valid for slits and a complete calculation should be carried out. However, the differences between TL_d and TL(0) have been found to be less than 2 dB.
6. Conclusion 1 can be considered as valid for a wide frequency range [100–5000 Hz]. For thin apertures, TL_d is positive but is very close to 0 dB on the whole frequency range.
7. As expected, for large, *thin* apertures TL_d has negative values but can be taken equal to 0 dB on the whole frequency range within a 1.5 dB error.

From this rigorous study, it can be concluded that a simple normal incidence transmission loss model for apertures can be used with a correction factor of 2 or 3 dB at low frequencies. The maximum error will be expected to be less than 3 dB at medium frequencies, and 1 dB at high frequencies.

Future works will involve (i) validating the approach for various apertures using experiments based on sound intensity measurements for the transmitted acoustic field to obtain more up-to-date experimental results and (ii) modeling apertures with acoustic linings (e.g., modeling silencers in enclosures).

[1] C. J. Bouwkamp, "Diffraction theory," Rep. Prog. Phys. **17**, 35–100 (1954).
[2] R. D. Spence, "The diffraction of sound by circular disks and apertures," J. Acoust. Soc. Am. **20**, 380–386 (1948).
[3] K. A. Mulholland and H. D. Parbrook, "Transmission of sound through apertures of negligible thickness," J. Sound Vib. **5**, 499–508 (1967).
[4] M. C. Gomperts and T. Kihlman, "The sound transmission loss of circular and slit-shaped aperture in walls," Acustica **14**, 1–16 (1964).
[5] G. P. Wilson and W. W. Soroka, "Approximation to the diffraction of sound by a circular aperture in a rigid wall of finite thickness," J. Acoust. Soc. Am. **37**, 286–297 (1965).
[6] A. Sauter and W. W. Soroka, "Sound transmission through rectangular slots of finite depth between reverberant rooms," J. Acoust. Soc. Am. **47**, 5–11 (1970).
[7] M. C. Gomperts and T. Kihlman, "The sound transmission loss of circular and slit-shaped aperture in walls," Acustica **18**, 144–150 (1967).
[8] F. P. Mechel, "The acoustic sealing of holes and slits in walls," J. Sound Vib. **111**, 297–336 (1986).
[9] Y. Nomura and S. Inawashiro, "On the transmission of acoustic waves through a circular channel of a thick wall," Res. Inst. Elec. Commun. **2**, 57–71 (1960).
[10] J. P. Thomé, "Contribution to the study of the sound transmission through a finite length slit—Application to the study of acoustic leaks in industrial machinery enclosures" (in French), Ph.D. thesis, Industrial Acoustic and Vibrations Department, University of Technology of Compiègne, France.
[11] K. T. Chen, "Study of acoustic transmission through apertures in a wall," Appl. Acoust. **46**, 131–151 (1995).
[12] H. H. Park and H. J. Eom, "Acoustic scattering from a rectangular aperture in a thick hard screen," J. Acoust. Soc. Am. **101**, 595–598 (1997).
[13] H. Serizawa and K. Hongo, "Evaluation of an acoustic plane wave transmitted through a rectangular hole in a thick hard screen," Wave Motion **36**, 103–117 (2002).
[14] C.-J. Huang and C. Y. Chen, "Diffracted acoustics fields about circular apertures in soft and hard baffles," Acust. Acta Acust. **85**, 301–311 (1990).
[15] Y. Furue, "Sound propagation from the inside to the outside of a room through an aperture," Appl. Acoust. **31**, 133–146 (1990).
[16] H. Nelisse, O. Beslin, and J. Nicolas, "A generalized approach for the acoustic radiation from a baffled or unbaffled plate with arbitrary boundary conditions," J. Sound Vib. **211**, 207–225 (1998).
[17] L. Lewin, *Advanced Theory of Waveguides*, Iliffe, London, 1951.
[18] F. Sgard, Modeling of the acoustic transmission through rectangular and circular apertures, France, internal report #2006–FS-1, IRSST, Canada, 70 pp.
[19] W. E. Zorumski, "Generalized radiation impedances and reflection coefficients of circular and annular ducts," J. Acoust. Soc. Am. **54**, 1667–1673 (1973).
[20] N. Atalla, F. Sgard, and C. K. Amedin, "On the modeling of sound radiated from poroelastic materials," J. Acoust. Soc. Am. **120**, 1990–1995 (2006).

Sound quality characteristics of refrigerator noise in real living environments with relation to psychoacoustical and autocorrelation function parameters

Shin-ichi Sato,[a] Jin You, and Jin Yong Jeon
School of Architectural Engineering, Hanyang University, Seoul 133-791, Korea

(Received 21 November 2006; revised 17 April 2007; accepted 22 April 2007)

Psychoacoustical and autocorrelation function (ACF) parameters were employed to describe the temporal fluctuations of refrigerator noise during starting, transition into/from the stationary phase and termination of operation. The temporal fluctuations of refrigerator noise include a click at start-up, followed by a rapid increase in volume, a change of pitch, and termination of the operation. Subjective evaluations of the noise of 24 different refrigerators were conducted in a real living environment. The relationship between objective measures and perceived noisiness was examined by multiple regression analysis. Sound quality indices were developed based on psychoacoustical and ACF parameters. The psychoacoustical parameters found to be important for evaluating noisiness in the stationary phase were *loudness* and *roughness*. The relationship between noisiness and ACF parameters shows that sound energy and its fluctuations are important for evaluating noisiness. Also, refrigerator sounds that had a fluctuation of pitch were rated as more annoying. The tolerance level for the starting phase of refrigerator noise was found to be 33 dBA, which is the level where 65% of the participants in the subjective tests were satisfied. © *2007 Acoustical Society of America.* [DOI: 10.1121/1.2739440]

PACS number(s): 43.50.Qp, 43.50.Jh [ADP] Pages: 314–325

I. INTRODUCTION

The number of complaints regarding noise in living environments, including those from household electric appliances, is rapidly increasing. Refrigerators, unlike other home appliances, operate all day. Refrigerator owners are sensitive to the noise refrigerators generate, and noise level is among the most important consumer criteria when buying household appliances. To meet the increasing demand for quietness in living environments, it is necessary to reduce the noise levels of refrigerators. Compressors and fans are the primary sources of refrigerator noise. Structural improvements can reduce the level of noise produced by these components. For example, the application of a flexible joint on the back cover of a refrigerator has been shown to reduce its sound pressure level by 2 dB (Paris and Tabuenca, 2002).

The standard industry method for measuring refrigerator noise in an anechoic chamber, according to ISO 3745 (2003), has been shown to give different sound characteristics than those produced in living environments. Jeon *et al.* (2007) found that the sound pressure level of refrigerator noise in an actual living room was about 10 dB higher than the level measured in an anechoic chamber 1 m from the front of the refrigerator. They also investigate the tolerable level of stationary refrigerator noise in the real living environment. It was found that 70% of the participants in the subjective tests were satisfied with a sound level of 30 dBA (L_{eq}).

Despite the continuous reduction of refrigerator noise levels, complaints and indications of noise discomfort persist. Therefore, sound quality must be considered in addition to noise level. In terms of psychoacoustics, noise from home appliances is not only influenced by sound pressure level SPL, but also by time of day, duration of noise, background noise, and frequency characteristics. Also, studies of the relationships among objective measurements of subjective responses to refrigerator noise should be conducted (Ebbitt and Davies, 1998). Cardozo and van der Veen (1979) showed that subjective annoyance results primarily from the low-frequency component of refrigerator noise. Tamesue *et al.* (2005) proposed a method for predicting a psychological impression of noise by using the instantaneous spectral distance. Similar evaluations of sound quality have been conducted in several mechanical engineering applications, such as automobiles, heating, ventilating and air conditioning (HVAC) systems, and wheel loaders (Lee *et al.*, 2005; Khan and Hogstrom, 2001; Khan and Dickson, 2002).

The psychoacoustical parameters, originally proposed by Zwicker and Fastl (1999), were designed to quantify the listener's perception and evaluation of sound quality. These parameters are used as a sound quality index for noises, mainly stationary ones. *Loudness* involves both frequency and temporal masking effects. The high-frequency components of a sound determine its perceived sharpness, and the *sharpness* of sound increases annoyance. *Roughness* and *fluctuation strength* describe the fluctuation of the signal. The target modulation frequencies of *roughness* and *fluctuation strength* are around 70 and 4 Hz, respectively. Since it was initially proposed, the procedure for calculating *Zwicker loudness* for stationary sound has been standardized (ISO 532, 1975). In our previous study, which also investigated the Sound Quality Index of the stationary part of refrigerator

[a]Author to whom correspondence should be addressed. Electronic mail: s_sato@mac.com

 0001-4966/2007/122(1)/314/12/$23.00

noise (Jeon *et al.*, 2007), it was shown that *loudness* and *roughness* are the psychoacoustical parameters that most significantly influenced subjective evaluations of refrigerator noise.

Refrigerator noise, especially at the initiation of operation, contains temporal fluctuations in sound, such as a click, a rapid increase in volume, and a change in pitch. The autocorrelation function (ACF) parameters may also describe noises related to sensory perception such as loudness, pitch, timbre, and duration. It has been found that the perceived pitch and strength (i.e., tonality) of complex sounds are extracted from the maximum peak in the ACF (Ando *et al.*, 1999; Ando, 2001). ACF parameters are typically applied to discussions of missing fundamental signals. When a signal contains harmonics but no fundamental frequency, the fundamental frequency is heard as a pitch. The frequency perceived from such harmonic components is called a residue pitch, a periodical pitch, a subjective pitch, or a virtual pitch. This phenomenon cannot be explained from the spectrum of source signals.

Some ACF-based models for predicting residue pitch have been proposed. The autocorrelation model of pitch perception was originally known as a "duplex" model (Licklinder, 1951). Three models for predicting residue pitch, dubbed "pattern recognition" models, have been widely accepted since the 1970s (Wightman, 1973a; Goldstein, 1973; Terhardt, 1974). A pattern-transformation model was proposed because the phase does not influence pitch perception (Wightman, 1973a, b). A pitch transformer based on ACF detects the locations of peaks from the output wave form in each frequency band. Consequently, pitch strength can be estimated from the height of the maximum peak extracted from the ACF form.

The effectiveness of the pattern-transformation model was examined to evaluate the validity of the peripheral weighting model using ripple noises (Yost, 1978; 1996a, b). The time delay of the maximum peaks of an ACF for a given sound source was used as a significant parameter for predicting pitch in the pitch-matching study. The loudness and annoyance of narrowband noise are related to the decay rate of the normalized ACF envelope, which describes the degree of persistence/randomness of a sound signal (Sato *et al.*, 2001; Sato *et al.*, 2002; Soeta *et al.*, 2004). When a real living environment is considered, information regarding the source direction, source size, and directional fluctuation are also important subjective attributes. These can be evaluated by the IACF (interaural cross-correlation function) parameters. Binaural measurements, which are actual human listening conditions, better reflect the spatial attributes of noise. For example, the peak value of IACF represents the degree of similarity of sound waves arriving at both ears, which is a significant factor in determining the degree of subjective diffuseness in a sound field (Ando, 1999).

ACF/IACF parameters have been used to describe the acoustic properties of stationary and impulsive sounds, that is, aircraft, trains, drainage, and floor impact sounds (Sakai *et al.*, 2001; Fujii *et al.*, 2001; Sakai *et al.*, 2002; Kitamura *et al.*, 2002; Jeon, 2001). Continuous measurement of fluctuating ACF/IACF parameters allows their impact on listener perception to be evaluated. The effect of the fluctuation of these factors on annoyance has previously been investigated (Fujii *et al.*, 2002; Sato *et al.*, 2004). Jeon (2001) used the ACF/IACF parameters of floor impact sounds to investigate the similarity between human-made impact sounds and standard heavy- and lightweight impact sounds. They then examined the relationship between loudness and the ACF/IACF or psychoacoustical parameters. Later, the effect of different sound insulation treatments on perceived loudness and annoyance from floor impact sounds were investigated in terms of impact sound pressure levels (Jeon *et al.*, 2002). They found that sound insulation in the floor and walls was effective in reducing loudness and annoyance.

The purpose of the present study is to investigate the effects of psychoacoustical and ACF/IACF parameters on the subjective evaluation of noise generated by the starting and stationary phases of refrigerator operation. This study seeks to determine which objective aspects of refrigerator noise correlate best with the subjective perception of noise. Acoustical characteristics of refrigerator sounds in real living environments were clarified using psychoacoustical and ACF/IACF parameters. Subjective tests were conducted to obtain a subjective rating score. The relationships between noisiness and objective measures were examined by multiple regression analyses and sound quality (SQ) Indices. Finally, the level of human tolerance of refrigerator noise was determined.

II. REFRIGERATOR NOISE MEASUREMENTS

The distribution of the sound level and psychoacoustical parameters in a unit of an apartment was already investigated by Jeon *et al.* (2007). In this study, the distribution of the IACF parameters (IACC and τ_{IACC}) in the room was investigated. The IACF between two sound signals at both ears $f_\ell(t)$ and $f_r(t)$ is defined by

$$\Phi_{\ell r}(\tau) = \frac{1}{2T}\int_{-T}^{+T} f'_\ell(t) f'_r(t+\tau)dt, \quad |\tau| \leqslant 1.0 \text{ ms}, \tag{1}$$

where $f'_\ell(t)$ and $f'_r(t)$ are obtained by signals $f_{\ell,r}(t)$ after passing through the A-weighted network, which approximately correspond to the ear sensitivity, $s(t)$, so that $f'_{\ell,r}(t) = f_{\ell,r}(t)^* s(t)$. The normalized IACF is defined by

$$\phi_{\ell r}(\tau) = \frac{\Phi_{\ell r}(\tau)}{\sqrt{\Phi_{\ell\ell}(0)\Phi_{rr}(0)}}, \tag{2}$$

where $\Phi_{\ell\ell}(0)$ and $\Phi_{rr}(0)$ are the ACFs at $\tau = 0$ for the left and right ear, respectively. IACC is the maximum value of the normalized interaural cross-correlation function for the possible maximum interaural time delay, say, within ±1 ms; τ_{IACC} is the interaural time delay at which the IACC is defined. W_{IACC}, used in Secs. III and V, is defined by the interval of the delay time at a value of δ (=0.1) below the IACC (Ando *et al.*, 1999). The definitions of these factors are illustrated in Fig. 1.

The two refrigerators were installed in the kitchen of a 100 m^2 apartment, which is the most common sized residential unit in Korea. The kitchen and living room area floor plans are shown in Fig. 1. Noise was measured at 22 posi-

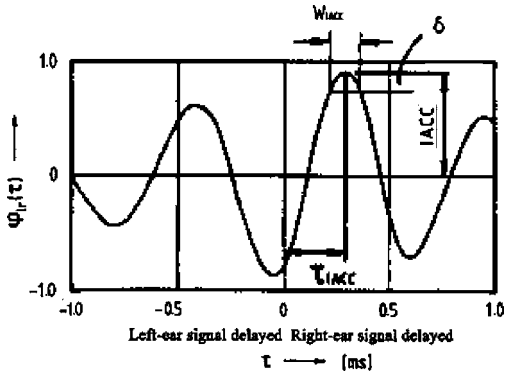


FIG. 1. Definition of IACC, τ_{IACC}, and W_{IACC}.

tions to investigate the distribution of refrigerator noise. The refrigerator sounds were recorded binaurally through a dummy head (B&K 4100). The distances from the ceiling and floor to the dummy head were 2.1, and 1.2 m, respectively. At every position, the dummy head faced the refrigerator. The duration of the refrigerator noise at each position was 5 s. The running IACC was calculated with an integration interval 3.0 s and the running step was 0.1 s. The average measures obtained for the signals in the range of 0–5 s were investigated.

IACC has been utilized to evaluate the spatial impression inside an auditorium. The large IACC value is observed in the area near the stage because the effect of the direct sound is dominant near the sound source. The large IACC is also observed in the area on the symmetrical axis due to the symmetrical reflections. Small IACC is realized by appropriate lateral reflections from the sidewalls. IACC in the apartment unit indicated the similar distribution.

Figure 2(a) shows the IACC distribution in the apartment unit. The positions which have relatively large IACC values (>0.4) were observed near the sound source and on the diagonal line between the kitchen and the living room. Figure 2(b) shows the distribution of τ_{IACC}. Except for the areas in front of corners, all other τ_{IACC} values are negative. The positive and negative τ_{IACC} indicate left and right side sound localization, respectively. The motor and compressor are located in the back of the refrigerator. The sound radiated from behind of the refrigerator is reflected by the sidewalls. The result of τ_{IACC} suggests that the refrigerator sound radiated from the behind and reflected by the walls near the refrigerator is the main sound path inside the apartment unit.

III. ANNOYANCE OF REFRIGERATOR SOUNDS IN TERMS OF PSYCHOACOUSTICAL AND ACF/IACF PARAMETERS

A. Procedure

The sounds from 24 refrigerators from six manufacturers from two countries (Korea and USA) were used in this test. The refrigerator sounds were convolved with the binaural impulse response measured by a dummy head (B&K 4100) at 7 m from their source in a 100 m^2 apartment unit. The reverberation time of the unit was around 0.7 s at 500 Hz. These sounds were reproduced in a test chamber using a stereo dipole technique including cross-talk canceling filters.

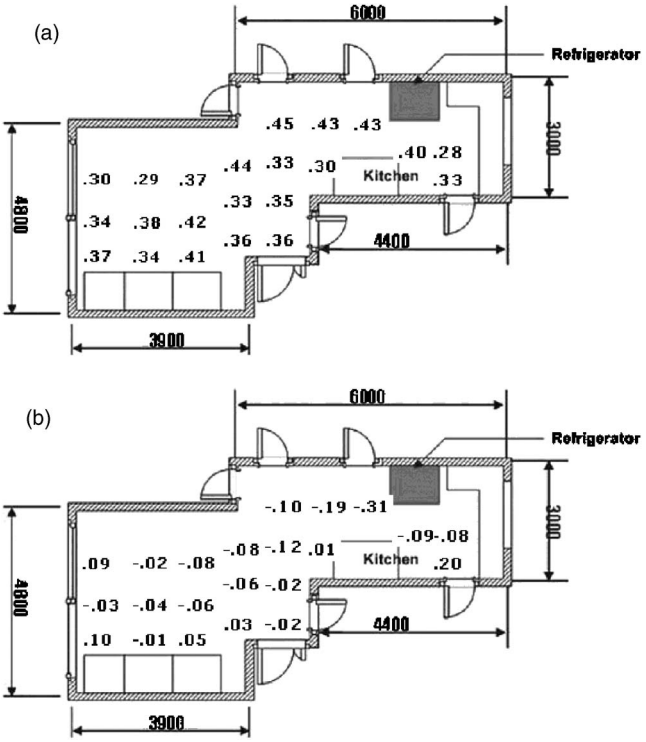


FIG. 2. IACC (a) and τ_{IACC} (b) distribution in the apartment unit.

They were played for 30 s in the range of 49–62 dBA. The original sound pressure levels were increased to ease subjective judgment. In order to confirm the objective parameters under experimental conditions, the auralized signals presented to listeners were recorded through the dummy head and the ACF parameters and SQ metrics of the signals were calculated. After listening to each signal, subjects were asked to evaluate their annoyance to each refrigerator sound using a two-category annoyance scale by selecting a magnitude ranging from 1 to 9 on the answer sheet. As shown in Table I, the two annoyance categories were; "noisiness," which specifies noise intelligibility in relation to the sound property like sound level, pitch and their fluctuations, and "amenity" as an evaluation of suitability of living including the activity disturbance in the real living environment. These categories, originally proposed by the Architectural Institute of Japan, were based on the residents' responses to noise corresponding to the classification of sound insulation. The validity of the category scales for subjective evaluation was previously

TABLE I. Subjective evaluation scale.

Annoyance group	Subjective magnitude	Noisiness	Amenity
Not annoying	1	Nothing perceivable	Excellent
	2	Hardly perceivable	Very fine
	3	Far-off noise	Good
Annoying	4	Slightly heard	Controllable
	5	Heard	Endurable
	6	Clearly heard	Yielding
Very annoying	7	Noisy	Unbearable
	8	Very noisy	Intolerable
	9	Extremely noisy	Let's move out

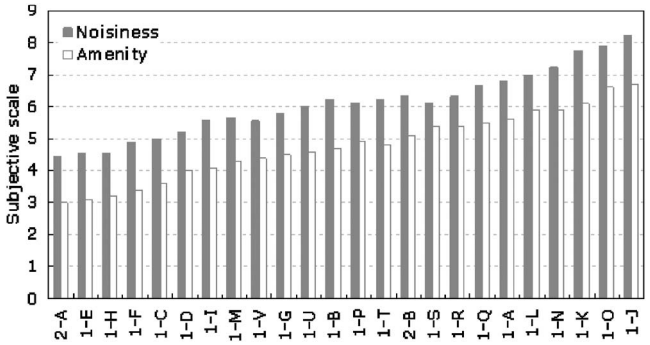


FIG. 3. Subjective rating scale of noisiness and amenity for individual refrigerator noises.

examined in the investigation of annoyance to floor impact sound (Jeon *et al.* 2004). They confirmed a linear relationship between the subjective magnitude scales and the sound level. Nine subjects with normal hearing took part in the subjective test. Each participant was given 13 min to evaluate the sounds of all 24 refrigerators.

B. Results

The averaged subjective noisiness and amenity results for the 24 refrigerator models are presented in Fig. 3. While noisiness relates to the evaluation of the refrigerator sound itself, amenity is influenced by the individual's background and experience. The average of the subjective scales of noisiness and amenity for the nine subjects was highly correlated ($r=0.98$; $p<0.01$). The correlation between the average of the subjective scales of noisiness for the nine subjects and the noisiness of each of the nine subjects was more than 0.70. The correlation between the average of the subjective scales of amenity for the nine subjects and the amenity of each of the nine subjects was greater than 0.52. Due to this high degree of correlation between individual respondent and average group results, the average of the subjective scales of noisiness for the nine subjects is used hereafter. The amenity scale is used to discuss the results of the tolerable level of refrigerator noise in Sec. VI. Higher values of noisiness indicate that the refrigerator was evaluated as being more annoying. To evaluate the relationship between the subjective rating scale of noisiness and the objective parameters, multiple regression analyses were conducted. The individual noisiness scale is also discussed later.

TABLE II. Correlation between the noisiness scale and psychoacoustical parameters.

	Loudness	Sharpness	Roughness	Fluctuation strength
Starting phase	0.75[a]	0.08	0.30	0.67[a]
Stationary phase	0.83[a]	0.10	0.45[b]	0.44[b]

[a] $p<0.01$.
[b] $p<0.05$.

C. Noisiness and psychoacoustical parameters

Psychoacoustical parameters (*loudness*, *sharpness*, *roughness*, and *fluctuation strength*) were used to clarify the acoustical characteristics of refrigerator noise. Because the procedure for calculating psychoacoustical parameters has not yet been standardized, the results calculated by one software package may differ from those of another. In the present study, the parameters were calculated by Pulse Software (Brüel & Kjær). The definition of the psychoacoustical parameters is described in Appendix A. The concept of psychoacoustical parameter calculation is illustrated in Fig. 4. The time interval between the spectra was set at 10 ms for the calculation of *roughness* and *fluctuation strength*. Mean values of *loudness* and *sharpness* were used to describe the acoustical property of refrigerator sound.

Each refrigerator sound was divided into two parts: the starting (0–5 s) and stationary (5–30 s) phases. The starting phase of refrigerator noise includes a click, a rapid increase of volume, and a change in pitch. The correlations between noisiness and psychoacoustical parameters for the starting and the stationary phases are listed in Table II. *Loudness* and *fluctuation strength* were significantly correlated with noisiness during the starting phase, with noisiness increasing with increases in *loudness* and *fluctuation strength*. In the stationary phase, *loudness* was shown to be significantly correlated

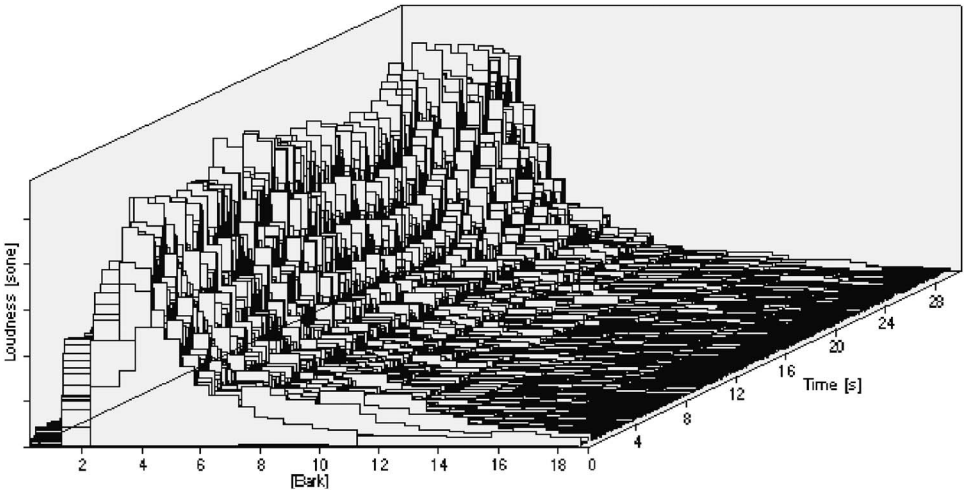


FIG. 4. Calculation of psychoacoustical parameters.

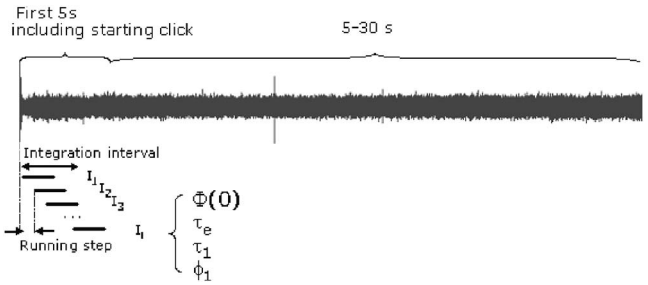


FIG. 5. Concept of short-time moving ACF in measurements.

with noisiness. *Roughness* and *fluctuation strength* were also correlated with noisiness, though to a lesser extent than *loudness*. Compared to the starting and stationary phases, *roughness* contributes more to the noisiness of the initial phase than to the noisiness of the stationary phase.

D. Noisiness and ACF/IACF parameters

ACF/IACF parameters (τ_1, ϕ_1, τ_e, $\Phi(0)$, IACC, τ_{IACC}, and W_{IACC}) were also used to clarify the acoustic characteristics of refrigerator sound. The definitions of the ACF and its parameters are described in Appendix B. The concept of a short-time moving ACF is illustrated in Fig. 5. The ACF/IACF are calculated at every given integration interval $2T$. The start of each integration interval was delayed for a short time (the moving step). The moving analysis aids in the effective description of temporal properties such as the transition, propagation, decay, and fluctuation of a sound. The physical parameters are extracted as fine structures of ACF/IACF. In this study, the integration interval was 3.0 s and the running step was 0.1 s. Average measures obtained for the signals in the left and right ear and those in the range of the starting (0–5 s) and stationary phases (5–30 s) were investigated. The standard deviation of each parameter was also calculated to express the temporal variation of each parameter.

The correlations between the noisiness and the ACF/IACF parameters for the starting and the stationary phase are listed in Table III. $\Phi(0)$ and the standard deviation of τ_1 of the starting phase were significantly correlated with noisiness. Noisiness increases with increases in $\Phi(0)$ and/or the deviation of τ_1. Correlations for the stationary phase indicate that $\Phi(0)$ was also significantly correlated with noisiness. No significant correlation was observed with the standard deviation of the parameters of the stationary phase.

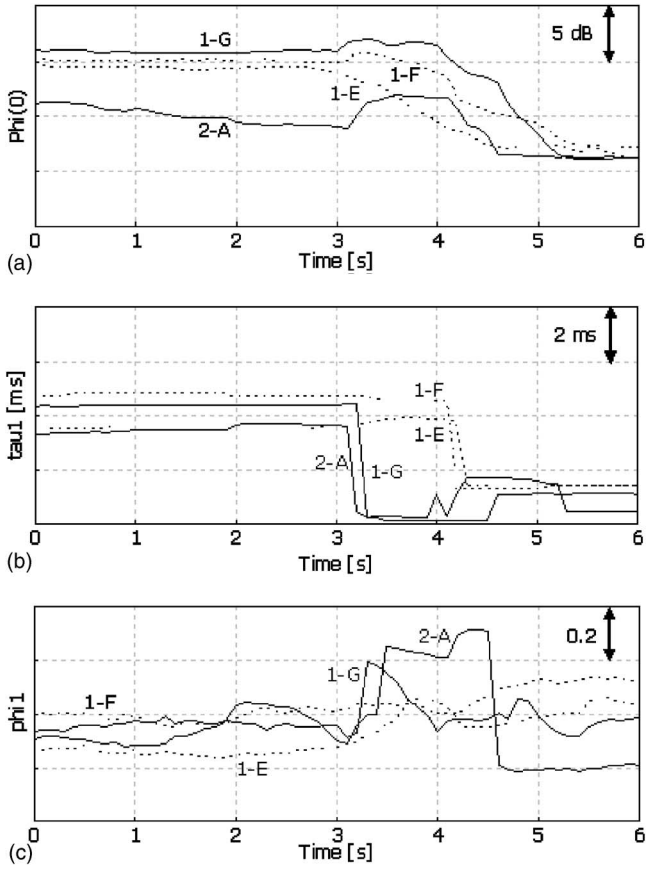


FIG. 6. FIG. 6. ACF parameters for an operational phase ending with a rattle sound (——; 1-G and 2-A) and a fade-out sound (- - - -; 1-E and 1-F). (a) $\Phi(0)$, (b) τ_1, and (c) ϕ_1.

IV. ENDING PHASE OF REFRIGERATOR OPERATION IN TERMS OF PSYCHOACOUSTICAL AND ACF PARAMETERS

Only the ending phase of refrigerator operation was evaluated objectively because the termination of the operation does not seem to affect people's annoyance. The psychoacoustical and ACF parameters of the ending phase were calculated in a manner similar to how the starting and stationary phases of operation were calculated. Ten of the 24 refrigerator sounds were available for the analysis of the ending phase. The ending phase can be distinguished from the stationary part by a decrease in sound level and a rattle caused by motor cessation. At most, each ending phase of operation lasted 3 s. Figure 6 shows the typical values of $\Phi(0)$, τ_1, and ϕ_1 for the ending phase of operation. Values greater than 3 s correspond to the ending phase, the sounds

TABLE III. Correlation between the noisiness scale and ACF parameters.

	τ_e	τ_1	ϕ_1	$\Phi(0)$	IACC	τ_{IACC}	W_{IACC}
Starting phase (mean)	0.13	0.17	−0.16	0.76[a]	−0.02	−0.29	−0.18
Starting phase (S.D.)	0.14	0.62[a]	0.23	0.46[b]	0.03	0.48[b]	−0.29
Stationary phase (mean)	0.06	0.17	−0.03	0.71[a]	−0.02	−0.19	−0.27
Stationary phase (S.D.)	0.15	0.54	−0.23	0.01	−0.17	0.03	−0.35

[a] $p<0.01$.
[b] $p<0.05$.

TABLE IV. Standardized regression coefficients for the psychoacoustical parameters.

	a_1	a_2	a_3	a_4	r
Starting phase	0.50[a]	−0.20	0.07	0.32	0.77[a]
Stationary phase	0.73[a]	−0.20	0.41[a]	0.02	0.88[a]

[a] $p<0.01$.

of which may be classified into "fade-out sound" and "rattle sound" groups. Rattle sounds are characterized by small τ_1 values and large ϕ_1 values. The high-frequency component of the rattle sound is easily heard; however, the fade-out sound decreases and with no significant changes in the pitch and strength. In terms of the psychoacoustical parameters, no significant characteristics classify the sounds into the two groups.

V. SQ (SOUND QUALITY) INDICES FOR NOISINESS

A. SQ Index for psychoacoustical parameters

In the previous section, the relationship between noisiness and each parameter was investigated. Here, the contribution of each parameter to noisiness is investigated using multiple regression analysis. All four psychoacoustical parameters were examined.

$$\mathrm{SQ} = a_1\mathrm{Loud} + a_2\mathrm{Sharp} + a_3\mathrm{Rough} + a_4\mathrm{Fluct} + c. \quad (3)$$

The standardized partial regression coefficients in Eq. (3) for the starting and stationary phases are listed in Table IV. Regarding the starting phase, *loudness* was a significant factor, while the regression analysis indicated that *loudness* and *roughness* were significant factors during the stationary phase. Perceived noisiness increased with increases in *loudness*, *roughness*, and/or *fluctuation strength* and with decreases in *sharpness*. Regression coefficients for the starting and stationary phases were 0.77 and 0.88, respectively. This suggests that the psychoacoustical parameters of the stationary phase effectively describe refrigerator sound noisiness. Then, SQ indices were examined using factors that were determined to be significant by regression analysis. Stepwise multiple regression was conducted to obtain an optimal equation. For the starting phase, the best combination of variables in terms of the correlation strength between noisiness and the psychoacoustical parameters was found to be *loudness* ($r=0.75$, $p<0.01$).

$$\mathrm{SQ}_{\mathrm{starting}} = a_1\ \mathrm{Loud} + c. \quad (4)$$

For the stationary phase, the best combination of variables, in terms of the correlation strength between noisiness and the psychoacoustical parameters, was found to be *loudness* and *roughness* ($r=0.87$, $p<0.01$). The standardized partial regression coefficients of *loudness* and *roughness* in Eq. (5) were 0.77 and 0.28, respectively, and these values were statistically significant ($p<0.05$).

$$\mathrm{SQ}_{\mathrm{stationary}} = a_1\ \mathrm{Loud} + a_2\ \mathrm{Rough} + c. \quad (5)$$

The standardized partial regression coefficients of each individual are listed in Table V. The relationships between the measured and calculated SQ indices for psychoacoustical parameters are shown in Fig. 7(a). SQ indices were calculated from Eqs. (4) and (5). Our study investigated *loudness* in terms of the Just Noticeable Difference (JND) (Jeon and You, 2006). The perceivable difference of the SQ index for

TABLE V. Standardized regression coefficient of each individual for each psychoacoustical parameter.

	Starting phase		Stationary phase		
	a_1	r	a_1	a_2	r
Global	0.33[a]	0.75[a]	0.77[b]	0.28[b]	0.87[a]
A	0.27[a]	0.60[a]	0.42[a]	0.72	0.72[a]
B	0.47[a]	0.75[a]	0.68[a]	1.46[a]	0.81[a]
C	0.25[b]	0.48[a]	0.45[a]	0.80	0.68[a]
D	0.31[a]	0.61[a]	0.38[a]	1.32[b]	0.71[a]
E	0.36[a]	0.64[a]	0.55[a]	1.02	0.75[a]
F	0.26[a]	0.65[a]	0.40[a]	0.12	0.69[a]
G	0.41[a]	0.79[a]	0.64[a]	0.40	0.86[a]
H	0.32[a]	0.54[a]	0.61[a]	0.73	0.74[a]
I	0.27[a]	0.72[a]	0.36[a]	0.24	0.66[a]

[a] $p<0.01$.
[b] $p<0.05$.

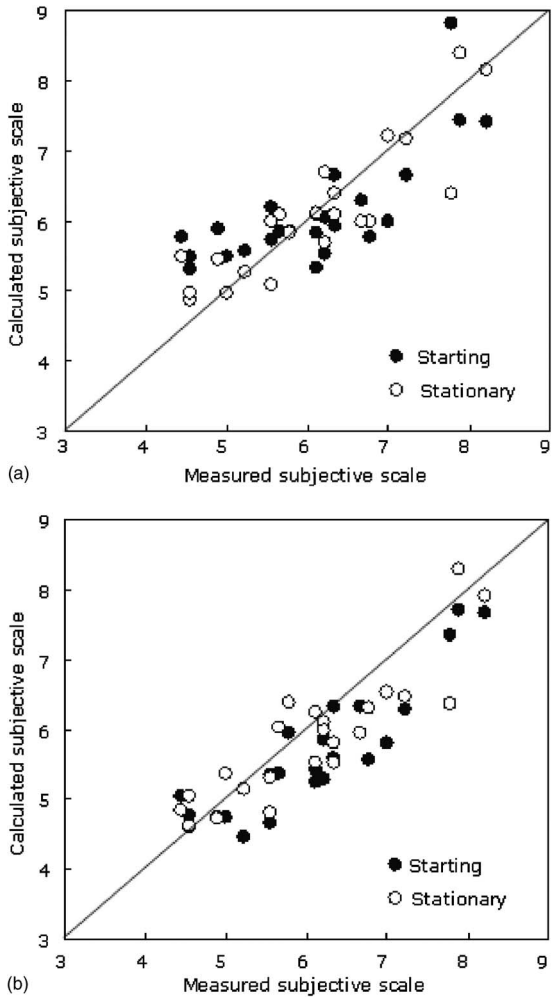


FIG. 7. Relationships between subjective scale obtained by subjective judgments and subjective scale: (a) values calculated by Eqs. (4) and (5) for psychoacoustical parameters; (b) values calculated by Eqs. (7) and (8) for ACF parameters.

TABLE VI. Standardized regression coefficients for ACF parameters.

	a_1	a_2	a_3	a_4	a_5	a_6	a_7	r
Starting phase	0.39	−0.12	−0.62[a]	0.83[a]	0.08	−0.02	0.05	0.86[a]
Stationary phase	0.16	0.01	−0.80[a]	1.10[a]	0.19	−0.05	0.12	0.88[a]

[a]$p<0.01$.

the psychoacoustical parameters was determined to be around 0.16.

B. SQ index for ACF/IACF parameters

The ACF/IACF parameters were examined in a similar fashion.

$$\mathrm{SQ} = a_1\tau_e + a_2\tau_1 + a_3\phi_1 + a_4\Phi(0) + a_5\mathrm{IACC} + a_6\tau_{\mathrm{IACC}} + a_7 W_{\mathrm{IACC}} + c. \quad (6)$$

The standardized partial regression coefficients in Eq. (6) for the starting and stationary phases are listed in Table VI. Both starting and stationary phases, $\Phi(0)$ and ϕ_1, were significant factors in the prediction of noisiness. Perceived noisiness increased with increases in $\Phi(0)$ as well as with decreases in ϕ_1. Regression coefficients for the starting and stationary phases were both 0.88.

As with the psychoacoustical parameters, SQ indices were examined using only the factors determined to be significant. Here, the standard deviations of ACF/IACF parameters were also included in the analyses. Stepwise multiple regression was conducted to obtain an optimal equation. For the starting phase, the best combination of variables, in terms of the correlation strength between noisiness and ACF/IACF parameters, was found to be $\Phi(0)$, ϕ_1, $sd\_\Phi(0)$ and sd_τ_1 ($r=0.93$, $p<0.01$). The standardized partial regression coefficients of $\Phi(0)$, ϕ_1, $sd\_\Phi(0)$ and sd_τ_1 in Eq. (7) were 0.78, −0.28, 0.22, and 0.23, respectively, and these coefficients were statistically significant ($p<0.05$).

$$\mathrm{SQ}_{\mathrm{starting}} = a_1\Phi(0) + a_2\phi_1 + a_3\mathrm{sd}\_\Phi(0) + a_4\mathrm{sd}\_\tau_1 + c. \quad (7)$$

For the stationary phase, the best combination of variables, in terms of the correlation strength between noisiness and ACF/IACF parameters, was found to be $\Phi(0)$ and ϕ_1 ($r=0.87$, $p<0.01$). The standardized partial regression coefficients of $\Phi(0)$ and ϕ_1 in Eq. (8) were 1.01 and −0.50, respectively, and these coefficients were statistically significant ($p<0.05$).

$$\mathrm{SQ}_{\mathrm{stationary}} = a_1\Phi(0) + a_2\phi_1 + c. \quad (8)$$

Equations (7) and (8) compare the starting and stationary phases of refrigerator sounds. $sd\_\Phi(0)$ and sd_τ_1 appear only in Eq. (7) for the starting part, while $\Phi(0)$ and ϕ_1 appear in both equations. IACF parameters were not found to be significant for both the starting and the stationary phases. The standardized partial regression coefficients of each individual are listed in Table VII. The relationships between the measured and calculated SQ indices for ACF parameters are shown in Fig. 7(b). SQ indices were calculated from Eqs. (7) and (8). In both equations, $\Phi(0)$ contributes most significantly to noisiness. The JND for $\Phi(0)$ can be calculated from the experimental condition by Jeon and You (2006). The perceivable difference of the SQ index for ACF parameters was determined to be around 0.5.

VI. TOLERABLE LEVEL OF REFRIGERATOR NOISE

A. Procedure

The acceptable level of perceived noise for the starting phase of refrigerator noise was investigated. Ten of the 24 refrigerator sounds were used for the test, the sounds of which were recorded in an anechoic chamber. The microphone was located 1 m from the back of the refrigerator at a height of 1 m. The signals recorded were convolved with the same binaural impulse response as discussed in the previous section. These sounds were reproduced in a test chamber using a stereo dipole technique including cross-talk cancel-

TABLE VII. Standardized regression coefficient of each individual for each ACF parameter.

	Starting phase					Stationary phase		
	a_1	a_2	a_3	a_4	r	a_1	a_2	r
Global	0.78[b]	−0.28[b]	0.22[b]	0.23[b]	0.93[a]	1.01[b]	−0.50[b]	0.87[a]
A	0.65[a]	−0.15	0.12	0.12	0.73[a]	0.80[a]	−0.50[b]	0.66[a]
B	0.54[a]	−0.25	0.32[a]	0.26	0.85[a]	0.91[a]	−0.78[a]	0.84[a]
C	0.36	−0.12	0.36[b]	0.30	0.73[b]	0.94[b]	−0.51[a]	0.78[a]
D	0.36	−0.15	0.26	0.35	0.73[a]	0.78[a]	−0.51[b]	0.65[a]
E	0.51[b]	−0.09	0.29	0.31	0.81[a]	0.94[a]	−0.47[b]	0.78[a]
F	0.71[a]	−0.21	0.33[a]	0.05	0.84[a]	0.88[a]	−0.42[b]	0.73[a]
G	0.74[a]	−0.13	0.22	0.19	0.89[a]	1.02[a]	−0.70[a]	0.85[a]
H	0.61[b]	−0.11	0.22	0.09	0.71[a]	0.90[a]	−0.45[b]	0.74[a]
I	0.18[a]	−0.17	0.38[a]	0.27	0.85[a]	0.82[a]	−0.30	0.69[a]

[a]$p<0.01$.
[b]$p<0.05$.

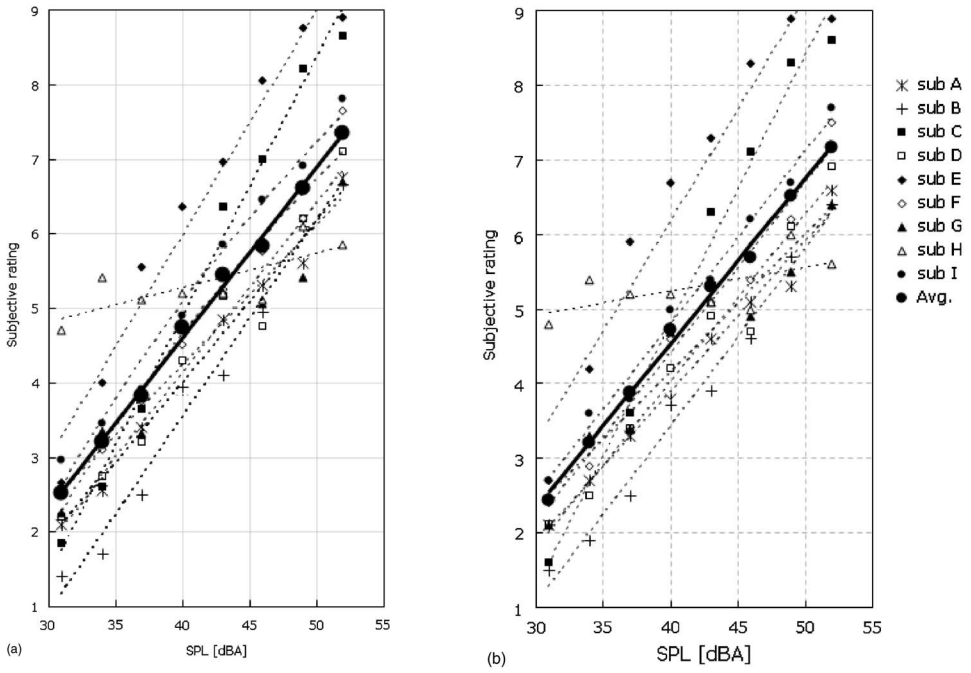


FIG. 8. Subjective rating scale of the sound pressure levels: (a) Noisiness and (b) amenity.

ing filters. Because the starting phase of refrigerator noise includes a large fluctuation click and an increase in volume, the maximum level L_{Amax} was considered. Eight noise sources, ranging from 31 to 52 dBA (L_{Amax}) in 3 dB intervals, were prepared from each refrigerator. The background noise of the test chamber was around 26 dBA. A total of 80 noise sources were randomly presented for a duration of 5 s. Each test of ten refrigerators was completed in about 15 min.

Ten subjects with normal hearing ability took part in the test. After listening to each signal, subjects were asked to evaluate their annoyance to each refrigerator sound using a nine-point, two-category annoyance scale by selecting a magnitude from 1 to 9 on an answer sheet. The two annoyance categories were noisiness and amenity, the same as the annoyance test described in Sec. III. The nine-point scales of both categories are shown in Table I. Subjects conducted the test on the assumption that they were relaxing in their living room at night. As shown in Eq. (9), percentage of satisfaction for noisiness and amenity was determined by dividing the number of responses below a given allowable limit by the total number of subjective responses.

$$\text{Percentage of satisfaction} = \frac{\text{number of responses within allowable limit}}{\text{total number of responses}} \times 100(\%). \tag{9}$$

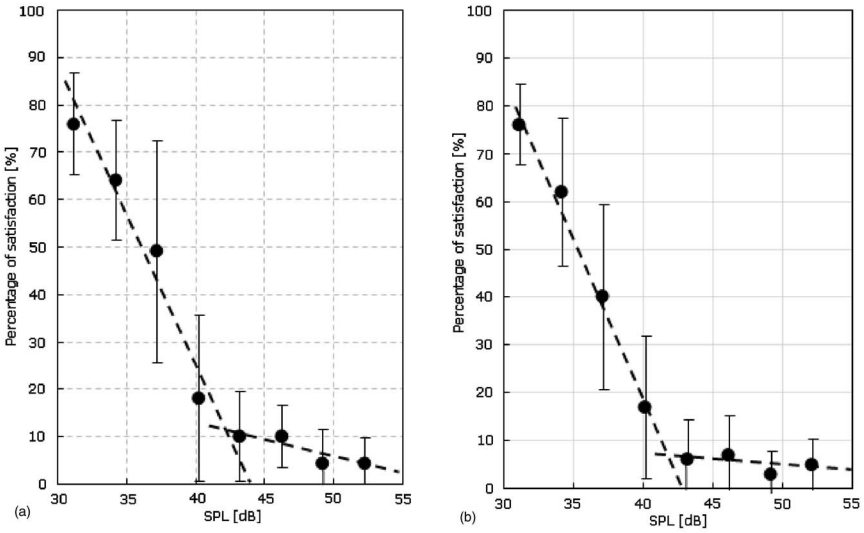


FIG. 9. Percent satisfaction as a function of the sound pressure level: (a) Noisiness and (b) amenity.

Based on an earlier study on the stationary part of refrigerator noise (Jeon *et al.*, 2007), level 3 on the subjective rating scale was selected as the allowable limit.

B. Results

The relationship between the maximum sound pressure level of the refrigerator and the individual subjective rating score is shown in Fig. 8. At level 3, the individual acceptable levels for noisiness and amenity were almost the same (30–38 dB) except for subject H. Level 3 is regarded as a kind of threshold for perceiving noise annoyance which includes noisiness and amenity. In contrast, at level 6, which can be regarded as the tolerable level, the range of the individual acceptable levels for amenity (39–51 dB) was slightly larger than that for noisiness (40–50 dB). The regression line for noisiness and amenity confirmed the linear relationship between the subjective rating and the sound pressure level. The regression line obtained by the average scale for the nine subjects indicated that a noise level of about 33 dBA (L_{Amax}) was determined to correspond to level 3 of the subjective scale.

Figure 9 shows percentage of satisfaction calculated by Eq. (9) as a function of the sound pressure level. Both noisiness and amenity indicate that the percent satisfaction for the sound pressure level below 40 dB decreases greatly. The regression curves (in the range less than 40 dB) obtained by the average scale for the ten refrigerators indicate that 65% of the respondents were satisfied with a sound level of 33 dBA.

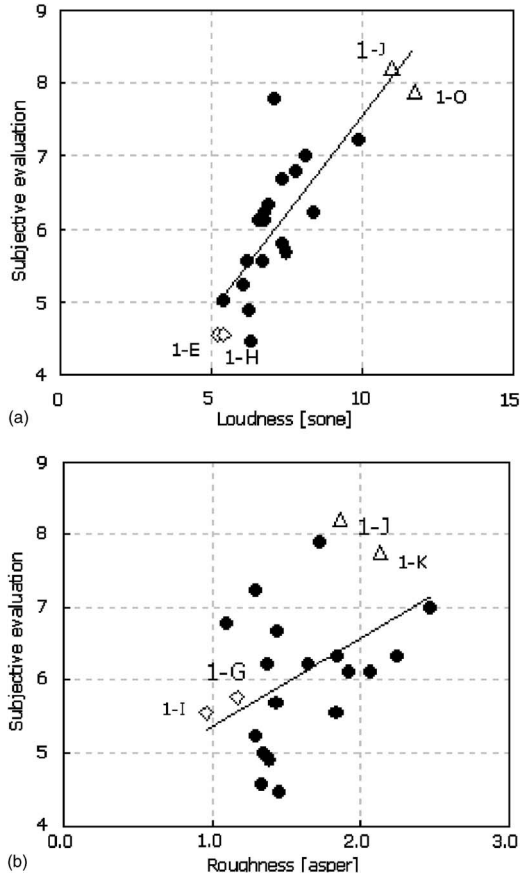


FIG. 10. Example of the refrigerator noise with higher (△; 1-J, 1-K, and 1-O) and lower noisiness (◇; 1-E, 1-G, 1-H, and 1-I) in terms of (a) loudness and (b) roughness.

VII. DISCUSSION

Figure 10 shows examples of the noisiness of refrigerator noise in terms of psychoacoustical parameters. Figure 11 shows how refrigerator sounds fluctuate, in terms of ACF parameters, through time. Refrigerators with higher noisiness have a greater $\Phi(0)$ and $sd\_\Phi(0)$ (Fig. 11(a)). Higher levels of noisiness indicate greater fluctuation in τ_1. On the other hand, refrigerators with lower noisiness have smaller $\Phi(0)$ and $sd\_\Phi(0)$. Lower levels of noisiness indicate smaller fluctuation in τ_1. These results show that a given refrigerator sound will have a higher level of noisiness when it has greater loudness, and greater temporal fluctuations of loudness and pitch. In terms of ϕ_1, refrigerator noises with higher noisiness have relatively smaller ϕ_1 values, and refrigerator noises with lower noisiness have relatively larger ϕ_1 values. This means that refrigerator noise containing more noise components (larger temporal fluctuations) will be perceived as having a higher level of noisiness.

Correlations between the ACF and the psychoacoustical parameters are listed in Table VIII. *Loudness* is well correlated with $\Phi(0)$. The starting and stationary phases of *roughness* also were well correlated with ϕ_1 and the fluctuation of $\Phi(0)$. *Roughness* may be related to low-frequency scattering. The correlation between the fluctuation of $\Phi(0)$ and *fluctuation strength* for the starting and ending phases, which fluctuate more drastically than the stationary phase, was greater than that for the stationary phase. *Loudness* and *fluctuation strength* can be calculated by using ACF parameters as follows:

$$\textit{Loudness} = 0.26\Phi(0) - 7.09. \tag{10}$$

$$\textit{Fluctuation strength} = 0.40sd\_\Phi(0) + 0.48. \tag{11}$$

The target modulation frequency of *fluctuation strength* is lower than that of *roughness*. When the frequency range of $sd\_\Phi(0)$ is limited up to the 500 Hz octave band, the regression coefficient for *fluctuation strength*, in terms of $sd\_\Phi(0)$, increases from 0.67 (A-weighted $sd\_\Phi(0)$) to 0.74. *Fluctuation strength* reflects the level of fluctuation in the time domain. *Roughness* and *fluctuation strength* result from the vibration of the internal mechanisms of a refrigerator. If damping materials were installed in the appropriate position, *roughness* and *fluctuation strength* could be reduced. Then, the SQ Index for the psychoacoustical parameter could be evaluated mainly by *loudness*.

In a previous study, we investigated the stationary part of refrigerator sounds and investigated the SQ Index for psychoacoustical parameters (Jeon *et al.*, 2007). As in the present study, *loudness* and *roughness* were significant factors. In addition, the SQ Index for ACF parameters can be calculated from the refrigerator sounds used in the study. Also, as in the present study, $\Phi(0)$ was found to be a significant factor in SQ index. The standardized partial regression coefficients of a_1, a_2, a_3 and a_4 in Eq. (6) were 0.34,

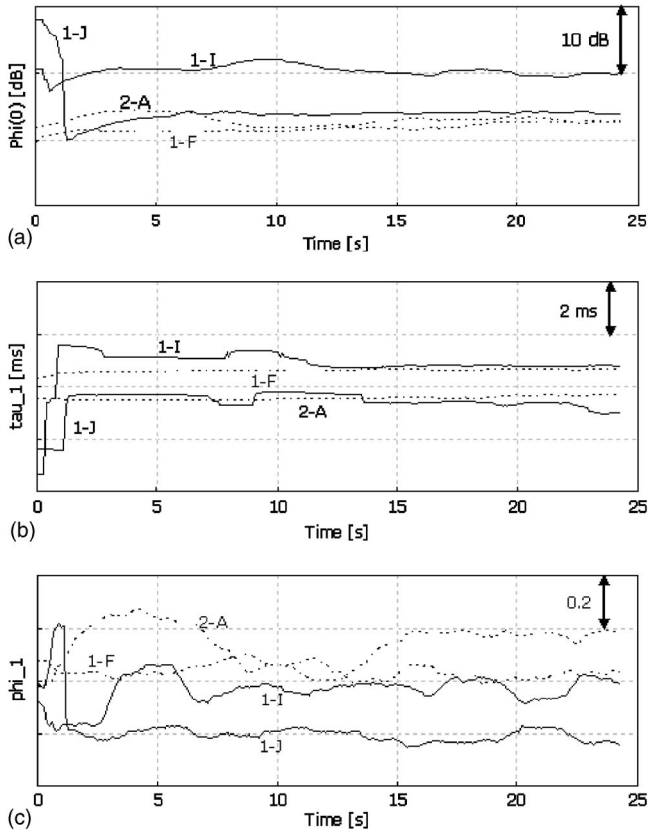


FIG. 11. ACF parameters of refrigerator noise with more (——; 1-J and 1-I) and less noisiness (- - - -; 1-F and 2-A). (a) $\Phi(0)$, (b) τ_1, and (c) ϕ_1.

−0.0001, 0.02, and 0.49, respectively. In the current study, IACF parameters were not determined to be significant in the SQ indices. This is because the refrigerator sound at only one typical position in the living room was reproduced. IACF parameters mainly change according to the listening position. The effects of the loudness, pitch, and their fluctuation of the sound source itself, which can be described by the psychoacoustical and ACF parameters on the subjective evaluation, were dominant.

The sound pressure level corresponding to 65% satisfaction was 33 dBA (L_{Amax}). This result is close to that of an earlier study on the stationary part of refrigerator noises, which indicated 30 dBA (L_{eq}) (Jeon *et al.*, 2007). The noise level obtained in this study is equivalent to 38 dBA at 1 m from the source in a living room and 26 dBA at 1 m from the source in an anechoic chamber. These results suggest that the reduction of the sound pressure level of a refrigerator should be the first priority, followed by improvement of sound quality.

In the first experiment, which is described in Sec. III, the refrigerator sounds were presented to the subjects in the range of 49–62 dBA. Even though these sound pressure levels are higher than actual refrigerator sound levels in a real living environment, the second experiment, which is described in Sec. VI, investigated the sound pressure levels ranging from 31 to 52 dB to reflect the noise levels in a real living environment. The experimental conditions of two experiments therefore cover the variation of sound levels in a real living environment.

TABLE VIII. Correlations between the psychoacoustical and ACF parameters. Top, middle, and bottom figures indicate the correlations for the starting, stationary, and ending phases, respectively.

	Loudness	Sharpness	Roughness	Fluctuation strength
$\Phi(0)$	0.68[a]	−0.17	−0.04	0.47
	0.83[a]	−0.35	−0.06	−0.09
	0.84[a]	0.41	0.86[a]	0.89[a]
τ_e	−0.11	−0.24	−0.17	−0.17
	0.01	−0.31	−0.30	−0.32
	0.17	−0.26	0.00	−0.26
τ_1	−0.05	−0.53[a]	−0.32	−0.25
	0.24	−0.21	−0.13	0.10
	0.08	−0.51	−0.02	−0.45
ϕ_1	−0.14	−0.24	−0.50[b]	−0.30
	0.15	−0.50[b]	−0.50[b]	−0.54[a]
	0.31	−0.49	0.06	0.4
sd_$\Phi(0)$	0.60[a]	0.11	0.28	0.50[b]
	−0.01	0.11	0.04	−0.18
	0.81[a]	−0.02	0.89[a]	0.71[b]
sd_τ_e	−0.10	−0.20	−0.14	−0.15
	0.01	−0.21	−0.20	−0.21
	0.44	−0.21	0.39	0.06
sd_τ_1	0.49	0.11	0.35	0.50[b]
	0.12	−0.10	−0.15	−0.23
	0.72[b]	−0.47	0.57	0.36
sd_ϕ_1	0.17	0.47	0.22	0.29
	−0.22	0.61[a]	0.08	0.11
	−0.12	0.00	−0.12	0.11

[a] $p < 0.01$.
[b] $p < 0.05$.

VIII. CONCLUSIONS

The results of our study lead us to the following conclusions:

- Important factors for evaluating refrigerator noisiness in the stationary phase are *loudness* and *roughness*. Regarding the amplitude fluctuation of refrigerator sounds, *fluctuation strength* is more important in the starting phase, and *roughness* is more important in the stationary phase.
- Among the ACF parameters, $\Phi(0)$, ϕ_1, and the standard deviations of $\Phi(0)$ and τ_1 of the initial phase are important.
- In terms of τ_1 and ϕ_1, the sounds associated with the ending phase of refrigerator operation can be divided into two types; the fade-out sound and the rattle sound.
- The acceptable starting phase noise level was found to be 33 dBA (L_{Amax}) when observers were situated 7 m from the source in a living room. This noise level is equivalent to 38 dBA at 1 m from the source in a living room and 26 dBA at 1 m from the source in an anechoic chamber. These levels correspond to a satisfaction rate of 65%.

APPENDIX A: PSYCHOACOUSTICAL PARAMETERS

The Zwicker *loudness* calculation is the basic parameter for measurements of both stationary and nonstationary signals. In this study, nonstationary *loudness* was considered. Nonstationary analysis results in a multi-spectrum with spectra equally spaced in time. Nonstationary analysis accounts

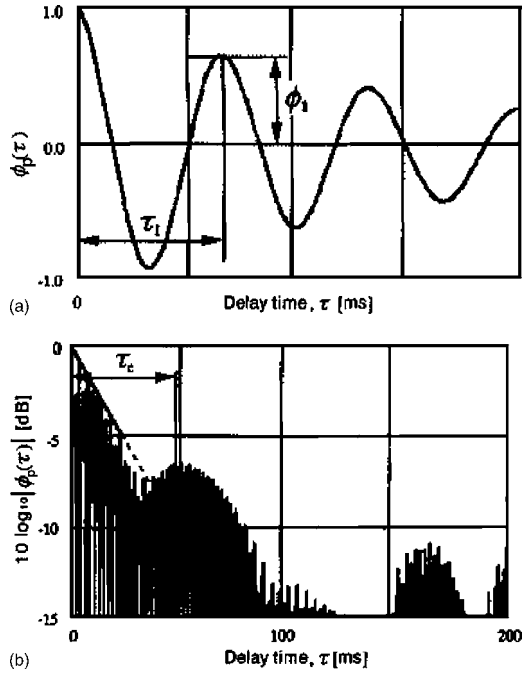


FIG. 12. Definition of ACF parameters. (a) τ_1 and ϕ_1; and (b) τ_e.

for time masking by using a set of exponentially decaying filters. From the multi-spectrum it is possible to calculate the total *loudness N*, and the total *loudness* as a function of time $N(t)$. Each specific *loudness* band, $N'(z)$, can be shown as a function of time, or each spectrum can be shown individually. The detailed procedure for the loudness calculation is described by Zwicker and Fastl (1999).

Sharpness is a parameter that measures the annoyance of high-frequency components in a signal. This is done by adding a weighting function to the specific *loudness* spectrum. The weighting function $g(z)$ is from 1 up to 16 Bark and grows exponentially above this band. The definition of Zwicker's *sharpness* is

$$sharpness = 0.11 \frac{\int^{24} N'(z)g(z)zdz}{\int^{24} N'(z)dz} \text{acum}. \tag{A1}$$

The denominator in Eq. (A1) gives the total loudness *N*. The upper integral is the first moment of specific loudness over the critical-band rate, but uses an additional factor $g(z)$ that is the critical-band rate

$$g(z) - \begin{cases} 1 & \text{for } z \leqslant 16 \\ 0.066e^{0.171z} & \text{for } z > 16 \end{cases}. \tag{A2}$$

Roughness is calculated on the basis of the time-varying (nonstationary) *loudness* multispectrum. The result can either be a single number or a function of frequency (Bark scale), which gives the specific roughness $R'(z)$. The formula for *roughness* was first developed by Zwicker. The formula has been slightly modified, using envelope analysis to find the modulation frequency, and the percentile *loudness* values, $N(1)$ and $N(99)$, to find the variation of the amplitude. The formula for $R'(z)$ is

$$R'(z) = 0.0003 f_{\text{mod}}(z)\Delta L(z)\Delta z dz, \tag{A3}$$

where

$$\Delta L = 20 \log(N'(1)/N'(99)) \tag{A4}$$

and f_{mod} is the modulation frequency. *Roughness* is a sum of all specific *roughnesses*:

$$roughness = \sum^{24} R'(z)\text{asper}. \tag{A5}$$

Fluctuation strength is also calculated on the basis of the time-varying (nonstationary) *loudness* multi-spectrum. The formula for specific *fluctuation strength* $F'(z)$ is:

$$F'(z) = \frac{0.032\Delta L(z)\Delta z}{f_{\text{mod}}(z)/4 + 4/f_{\text{mod}}(z)}, \tag{A6}$$

where

$$\Delta L = 20 \log(N'(1)/N'(99)). \tag{A7}$$

The fluctuation strength is a sum of all specific *fluctuation strengths*:

$$fluctuation\ strength = \sum^{24} F'(z)\text{vacil}. \tag{A8}$$

For both *roughness* and *fluctuation strength* calculations, envelope analysis is carried out for each critical band, giving modulation frequency as a function of the critical band. The peak frequencies closest to the maximum impact frequency for roughness and fluctuation strength are 70 and 4 Hz, respectively.

APPENDIX B: AUTOCORRELATION FUNCTION AND ITS PARAMETERS

The ACF of a given source signal $p(t)$ is defined by

$$\Phi(\tau) = \frac{1}{2T}\int_{-T}^{+T} p'(t)p'(t+\tau)\mathrm{d}t, \tag{B1}$$

where $p'(t) = p(t) * s(t)$, $s(t)$ being the ear sensitivity, which consists of the head-related, the ear canal, and the middle ear transfer functions (Onchi, 1961; Rubinstein *et al.*, 1966; Mehrgardt and Mellert, 1977). For convenience, $s(t)$ may be chosen as the impulse response of an *A*-weighted network. The normalized ACF is defined by

$$\phi(\tau) = \Phi(\tau)/\Phi(0). \tag{B2}$$

The first ACF parameter is energy represented at the origin of the delay, $\Phi(0)$. The second and third ACF parameters are the delay time and amplitude, τ_1, and ϕ_1, of the first dominant peak of the ACF. The fourth parameter is the effective duration of the envelope of the ACF, τ_e. This factor is defined by the 10-percentile delay representing a kind of repetitive feature or reverberation within the source signal itself. The definitions of these factors are illustrated in Fig. 12.

Ando, Y. (**2001**). "A theory of primary sensations and spatial sensations measuring environmental noise," J. Sound Vib. **241**, 3–18.

Ando, Y., Sato, S., and Sakai, H. (**1999**). "Fundamental subjective attributes of sound fields based on the model of auditory–brain system," in *Computational Acoustics in Architecture*, edited by J. J. Sendra (WIT, Southampton), Chap. 4, pp. 63–99.

Cardozo, B. L., and van der Veen, K. G. (**1979**). "Estimation of annoyance due to low level sound," Appl. Acoust. **12**, 389–396.
Ebbitt, G., and Davies, P. (**1998**). "The role of sound quality assessment in noise control," *Proceedings of the NOISE-CON 98*, Ypsilanti, MI.
Fujii, K., Atagi, J., and Ando, Y. (**2002**). "Temporal and spatial factors of traffic noise and its annoyance," J. Temporal Des. Arch. Environ. **2**, 33–41, www.jtdweb.org(last viewed on 4/18/2007).
Fujii, K., Soeta, Y., and Ando, Y. (**2001**). "Acoustical properties of aircraft noise measured by temporal and spatial factors," J. Sound Vib. **241**, 69–78.
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.
International Organization for Standardization (**1975**). "Acoustics-Method for Calculating Loudness Level: ISO 532," Geneva.
International Organization for Standardization (**2003**). "Acoustics–Determination of sound power levels of noise sources using sound pressure–Precision methods for anechoic and hemi-anechoic rooms: ISO 3745," Geneva.
Jeon, J. Y. (**2001**). "Subjective evaluation of floor impact noise based on the model of ACF/IACF," J. Sound Vib. **241**, 147–155.
Jeon, J. Y., Jeong, J. H., and Ando, Y. (**2002**). "Objective and subjective evaluation of floor impact noise," J. Temporal Des. Arch. Environ. **2**, 20–28, www.jtdweb.org (last viewed on 4/18/2007).
Jeon, J. Y., Jeong, J. H., Vorländer, M., and Thaden, R. (**2004**). "Evaluation of floor impact sound insulation in reinforced concrete buildings," Acta Acust. Acust., **90**, 313–318.
Jeon, J. Y., and You, J. (**2006**). "Just noticeable difference of sound quality metrics of refrigerator noise," *Proceedings of the 35th INTER-NOISE*, Honolulu.
Jeon, J. Y., You, J., and Chang, H. Y. (**2007**). "Sound radiation and sound quality characteristics of refrigerator noise in real living environments," Appl. Acoust. (accepted for publication)
Khan, M. S., and Dickson, C. (**2002**). "Evaluation of sound quality of wheel loaders using a human subject for binaural recording," Noise Control Eng. J. **50**, 117–126.
Khan, M. S., and Hogstrom, C. (**2001**). "Determination of sound quality of HVAC systems on trains using multivariate analysis," Noise Control Eng. J. **49**, 276–283.
Kitamura, T., Sato, S., Shimokura, R., and Ando, Y. (**2002**). "Measurement of temporal and spatial factors of a flushing toilet noise in a downstairs bedroom," J. Temporal Des. Arch. Environ. **2**, 13–19, www.jtdweb.org(last viewed on 4/18/2007).
Lee, J. K., Park, Y. W., and Chai, J. B. (**2005**). "Development of a sound quality index for the evaluation of an intake noise of a passenger car," J. Kor. Soc. Noise Vib. Eng. **15**, 939–944.
Licklinder, J. C. R. (**1951**). "A duplex theory of pitch perception," Experimentia **7**, 128–134.
Mehrgardt, S., and Mellert, V. (**1977**). "Transformation characteristics of the external human ear," J. Acoust. Soc. Am. **61**, 1567–1576.
Onchi, Y. (**1961**). "Mechanism of the middle ear," J. Acoust. Soc. Am. **33**, 794–805.
Paris, J. L., and Tabuenca, P. S. (**2002**). "Evaluation of the noise emitted by the condenser of a household refrigerator," *Proceedings of the 1st Pan-American/Iberian Meeting on Acoustics*, Cancun, Mexico.
Rubinstein, M., Feldman, B., Fischler, H., and Frei, E. H. (**1966**). "Measurement of stapedial-footplate displacements during transmission of sound through the middle ear," J. Acoust. Soc. Am. **40**, 1420–1426.
Sakai, H., Hotehama, T., Ando, Y., Prodi, N., and Pompoli, R. (**2002**). "Diagnostic system based on the human auditory-brain model for measuring environmental noise: An application to railway noise," J. Sound Vib. **250**, 9–21.
Sakai, H., Sato, S., Prodi, N., and Pompoli, R. (**2001**). "Measurement of regional environmental noise by use of a PC-based system: An application to the noise near the airport "G. Marconi" in Bologna," J. Sound Vib. **241**, 57–68.
Sato, S., Kitamura, T., and Ando, Y. (**2002**). "Loudness of sharply (2068 dB/oct.) filtered noises in relation to the factors extracted from the autocorrelation function," J. Sound Vib. **250**, 47–52.
Sato, S., Kitamura, T., and Ando, Y. (**2004**). "Annoyance of noise stimuli in relation to the spatial factors extracted from the interaural cross-correlation function," J. Sound Vib. **277**, 511–521.
Sato, S., Kitamura, T., Sakai, H., and Ando, Y. (**2001**). "The loudness of "complex noise" in relation to the factors extracted from the autocorrelation function," J. Sound Vib. **241**, 97–103.
Soeta, Y., Maruo, T., and Ando, Y. (**2004**). "Annoyance of bandpass filtered noises in relation to the factor extracted from autocorrelation function," J. Acoust. Soc. Am. **116**, 3275–3278.
Tamesue, T., Yamaguchi, S., Saeki, T., and Kato, Y. (**2005**). "A method for prediction of listening score and psychological impression in an actual noise environment," J. Sound Vib. **287**, 625–636.
Terhardt, E. (**1974**). "Pitch, consonance, and harmony," J. Acoust. Soc. Am. **55**, 1061–1069.
Wightman, F. L. (**1973a**). "The pattern-transformation model," J. Acoust. Soc. Am. **54**, 407–416.
Wightman, F. L. (**1973b**). "Pitch and stimulus fine structure," J. Acoust. Soc. Am. **54**, 397–406.
Yost, W. A., Hill, R., and Perez-Falcon, T. (**1978**). "Pitch and pitch discrimination of broadband signals with rippled power spectra," J. Acoust. Soc. Am. **63**, 1166–1173.
Yost, W. A. (**1996a**). "Pitch of iterated rippled noise," J. Acoust. Soc. Am. **100**, 511–518.
Yost, W. A. (**1996b**). "A time domain description for the pitch strength of iterated rippled noise," J. Acoust. Soc. Am. **99**, 1066–1078.
Zwicker, E., and Fastl, H. (**1999**). *Psychoacoustics: Facts and Models* (Springer-Verlag, Berlin).

A new definition of boundary point between early reflections and late reverberation in room impulse responses

Takayuki Hidaka,[a] Yoshinari Yamada, and Takehiko Nakagawa
Takenaka R&D Institute, 1-5-1, Otsuka, Inzai, Chiba 270-1395, Japan

(Received 8 May 2006; revised 2 May 2007; accepted 2 May 2007)

The early reflections in the room impulse response are usually defined as those observed within the initial 80 ms after the arrival of the direct sound, after which time the sound field is called reverberant. This number was chosen from measurements of other functions in a limited number of halls. In order to give an objective foundation to this time separation and to establish a physical indicator for it, a new method is proposed that defines a "transition time t_L," which is the time at which the energy correlation between the direct plus initial sound and the subsequent decaying sound first achieves a specified low value. For various halls this number is shown and its relevance as a new parameter is discussed. © *2007 Acoustical Society of America.* [DOI: 10.1121/1.2743161]

PACS number(s): 43.55.Br, 43.55.Gx, 43.55.Ka [NX] Pages: 326–332

I. INTRODUCTION

In room acoustics a wave-related parameter (early reflections) and a statistical parameter (late reverberation) are utilized in the characterization of the sound fields in halls for the performance of music. It seems important that there should be a single physical measure for determining the point at which the early reflections have been overtaken by the late reverberant sound, i.e., the point at which the correlation between the two is at a low value.

Previously, several studies discussed the range of early reflections from a subjective point of view. Reichardt *et al.* (1974) showed that 80 ms gives the boundary in the room impulse response (RIR) between usable and useless transparency. Beranek (1992, Figs. 10 and 33) infers from two measured parameters that 80 ms is an approximate time separation between early reflections and late reverberation. Barron (1993) writes that individual reflections (in large halls) arriving after 100 ms are no longer distinguishable. Kuttruff (1993) indicates that the characteristic time that separates the early and late parts of the impulse response should be 100–150 ms. Hidaka *et al.* (1995) state "…the time that separates the early reflections from the late reflections exists in the range from 50 to 200 ms." This range in values indicates that a more meaningful separation time is needed, particularly for auralization (Heinz, 1993; Kuttruff, 2000), electronic reverberators (Schroeder, 1962; Gardner, 1998), and other sound field simulations (Krokstad *et al.*, 1983; Vorländer, 1989).

This paper investigates the possibility of a separation time t_L where the correlation between the direct plus initial sound and the subsequent sound first achieves a specified low value. This time t_L is a physical quantity that can be measured in actual halls and is not based on psychoacoustics or on geometrical acoustics. It will be shown that the value of t_L is related to the type of hall (symphony, chamber, opera), shape of the hall (shoebox versus non-shoebox), and reverberation time.

II. DEFINITION AND ANALYSIS

A. Early reflections

The early reflections (ER) are defined as the early part of the room impulse response (RIR) and are related to several of the subjective attributes perceived by audiences attending a concert (Beranek, 2004). The first reflection comes from a sidewall or ceiling and its arrival time is about 20–60 ms after the direct sound in typical halls (Beranek, 2004, Fig. 4.14; Hidaka and Beranek, 2000; Hidaka and Nishihara, 2004). Since the mean free path (MFP) of sound rays in any room is given by $4V/S$, the mean arrival time of each subsequent reflection MFP/c is about 35–50 ms in large concert halls. (Here, V, S, and c are respectively room volume, total surface area, and sound speed.) When it is assumed that ER is no greater than 80 ms; those potential reflections included in ER are limited to at most those that have bounced once or twice at surfaces (Vorländer, 1995). In saying this, we are assuming that the reflecting surfaces are nearly rigid with dimensions large compared to a wave length so that the phases of those reflections are determined only by the total path length from a sound source to a receiving point. This means that that ER is a deterministic signal. Because the audience areas in halls are highly absorptive, the sound energy that impinges on these areas will not result in significant reflections.

B. RIR with increases in time

When a sound wave from a source reflects at interior surfaces repeatedly, the density of the number of reflections per unit time of the RIR increases rapidly as time increases. Concurrently, both the amplitude and the phase of the RIR are changed by the running phenomena so that the wave form of the RIR becomes complicated to a sizable degree. Basic assumptions to what follows in this chapter, including references to the literature, are the following:

[a]Electronic mail: hidaka.takayuki@takenaka.co.jp

 0001-4966/2007/122(1)/326/7/$23.00

- Each surface in a hall is basically reflective, except for audience areas, and its size is finite so that the reflecting points that are geometrically determined on it are distributed randomly. Waveform deformation in the reflected sound occurs after undergoing several reflections because of the phase randomization effect (Schroeder, 1987), that is, the phase of the reflected sound becomes a sum of random variables.
- There are lots of edges, corners, and similar obstacles in a hall. When a sound wave encounters one of these objects, the diffracted wave that is generated propagates in every direction as a cylindrical wave (Bowman *et al.*, 1987). Accordingly, abundant multiple diffracted waves build up as time increases and nonspecular (incoherent) reflected sound is generated (Beckmann and Spizzichino, 1987). With the latter phenomenon, the random component in the RIR increases with time and, as near sound diffusion evolves, late reverberation LR is achieved. In other words, LR can be considered a stochastic signal.

C. Late reverberation

Next, one can say that RIR is equal to the sum of ER, plus intermediate transition reflections, plus LR. When does LR become a diffuse sound field? Schroeder (1959) defines diffusion at a point as the angular distribution of sound energy flux in the plane wave expansion of the sound field and, if the distribution over the solid angle is uniform, he calls the sound field at this point "completely diffuse." Junius (1959) experimentally infers that diffuse reverberant sound begins after 100–150 ms. Even if the reverberant sound is completely diffuse, it is not clear how the measurement of RIR is affected when measured by a single omnidirectional microphone, i.e., with the results plotted on a time axis.

Kuttruff (1993) writes in his research on the digital synthesis of the LR in the artificial RIR that the "gross" temporal and spectrum characteristics are important rather than the fine structure and that the phase spectrum is subjectively insignificant. He states that it is only necessary to adjust the gross spectrum and temporal structure, neglecting its microscopic structures in order to achieve sufficient subjective impression. Kuttruff (1991) reports that phase randomization in the RIR occurs when the distance from the sound source exceeds twice the "reverberation distance" so that phase angles may safely be neglected. Thus, when evaluating LR, one may choose any phase function that is convenient for the processing.

Summarizing the above, one can infer that the LR is the posterior component of the RIR, which has gone through more than second or third order reflections; that the phase or microstructure of the RIR has negligible meaning; that the gross spectrum and temporal characteristics are meaningful; and that stochastic manipulation is applicable for analysis (Ebeling, 1984; Schroeder, 1987).

We must note also that the ensemble averages (defined as the intensities over many repeated measurements under identical physical conditions) of the rise in intensity of sound in a large room, starting with the instant at which the sound is turned on, is equal to the steady state intensity minus the intensity during the decay of sound, starting with the instant the sound is turned off (Schroeder, 1966). In other words, the longer the reverberation time, the longer the rise time.

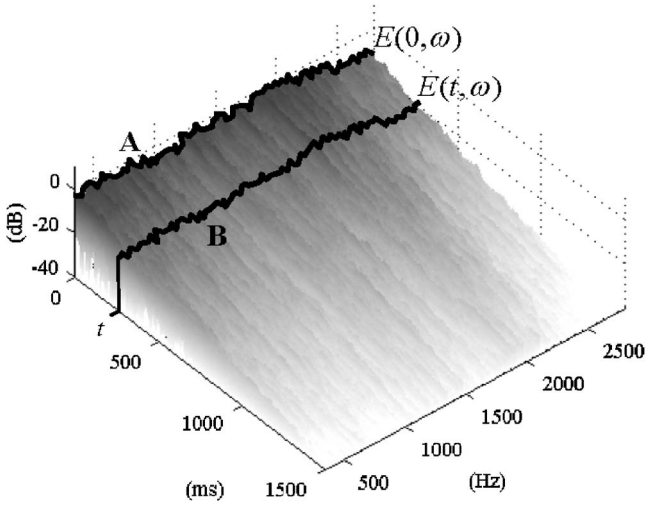


FIG. 1. Example of the time-frequency analysis of acoustic energy $E(t,\omega)$ by Eq. (1), calculated from the room impulse response, RIR, measured at Boston Symphony Hall (2625 seats). It shows the reverberant decay curves at each sampling frequency as well as the temporal variation of the amplitude of RIR. The measurement procedure of the RIR is given in a later section.

D. Definition of transition time, t_L

By applying Levin distribution to evaluate a transient signal, the time-frequency distribution of the acoustic energy $E(t,\omega)$ is given by (Yamada and Hidaka, 2001)

$$E(t,\omega) = \left| \int_t^\infty p(\tau)\exp(i\omega\tau)d\tau \right|^2, \tag{1}$$

where $p(t)$ is the RIR between a sound source and a receiving position in the hall.

By taking the frequency average of Eq. (1), the next relation holds:

$$\begin{aligned}\langle E(t,\omega)\rangle &= \frac{1}{\omega_2-\omega_1}\int_{\omega_1}^{\omega_2} E(t,\omega)d\omega \\ &= \frac{2\pi}{\omega_2-\omega_1}\int_t^\infty \int_t^\infty p(\tau)p(\tau')\,\delta(\tau-\tau')d\tau d\tau' \\ &= \frac{1}{f_2-f_1}\int_t^\infty p^2(\tau)d\tau. \end{aligned} \tag{2}$$

Here, the integral in the last expression is identical to the Schroeder integration (Schroeder, 1965), and $t=0$ means the time when the direct sound arrives. Hereafter, the notation $\langle\ \rangle$ means the frequency averaging.

Figure 1 is a calculated example of Eq. (1) that shows the temporal variation of the amplitude of the RIR for a large concert hall, that is, the reverberant decay curves at each sampling frequency between frequency range (ω_1,ω_2). Because the response at $t=0$ (curve A) includes the direct sound and all reflections, it is equal to the integration of the squared

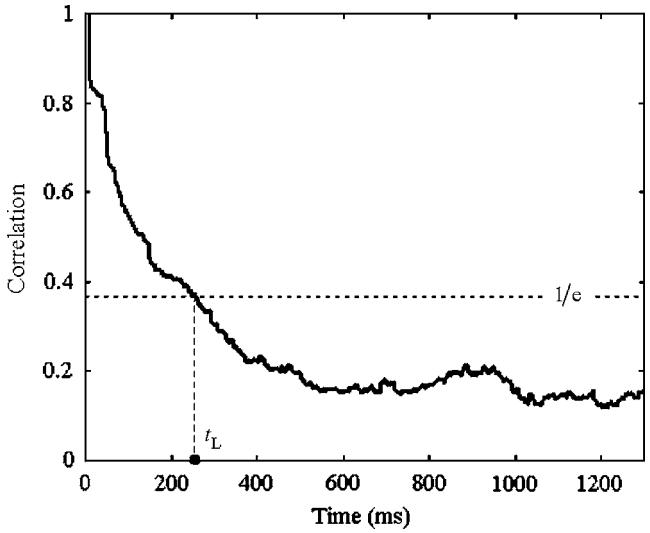


FIG. 2. Definition of the transition time t_L, the boundary point between the early reflections and the late reverberation on the RIR when the correlation function first becomes sufficiently low, i.e., $r(t)=1/e=0.367$. The vertical axis means normalized correlation between acoustic energies at $t=0$ and t. The frequency range of the RIR is from 354 to 2828 Hz that corresponds to 3 octave bands; 500 Hz, 1 kHz, and 2 kHz. Same data of Boston Symphony Hall in Fig. 1 were used.

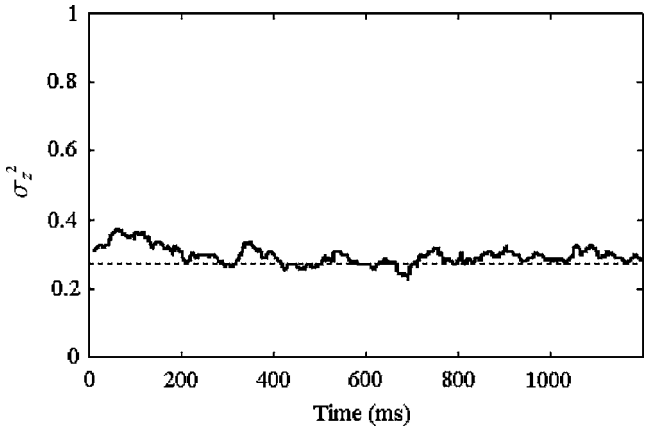


FIG. 3. Plot of the variance of normalized amplitude of the frequency response, $z=\rho/\langle\rho\rangle$ as a function of time. When z obeys the Rayleigh distribution, the theoretical value of σ_z^2 is $4/\pi-1=0.27$. The measurement result approaches the theoretical value early on. The frequency range is 500 Hz octave band. Same data as in Fig. 1 were used.

impulse response from $t=\infty$ to 0. Curve A contains all reflections that arrive after the time $t=0$, usually called the frequency response curve.

For t small, the frequency curve contains, even if the direct sound does not exist, lower order reflections with no phase randomization. Stated differently, the RIR is a sum of deterministic and stochastic components. Accordingly, Eq. (2) decreases stepwise with time as shown in Fig. 2.

If we consider the case where ER includes the direct sound plus the reflections within the initial 80 ms after the arrival of the direct sound, only second or third order reflections are involved. Measurements of C_{80} values in actual halls (Beranek, 2004) show that the sound energies before and after the 80 ms of the RIR are nearly the same order. If we start with this observation and decide on a time separation that includes more than second or third order reflections, then the deterministic component ER and stochastic component LR are, in curve A, about equal, i.e., we can also say that the ER and LR are stochastically independent from each other.

Now let us introduce a time dependant correlation function to evaluate the similarity between curves A and B by utilizing Pearson's correlation coefficient,

$$r(t)=\frac{\langle(E(0,\omega)-\mu(0))(E(t,\omega)-\mu(t))\rangle}{\sqrt{\langle(E(0,\omega)-\mu(0))^2\rangle\langle(E(t,\omega)-\mu(t))^2\rangle}}, \tag{3}$$

where $\mu(t)=\langle E(t,\omega)\rangle$ means a frequency average. This gives the correlation between the energy of the initial state ($t=0$ s) and the energy of state at any time t later for a particular frequency range.

At what point in time does Eq. (3) becomes sufficiently small for practical purposes? Certainly, ER should include more than one or two early reflections, and must not include sound that is diffused. In this paper we define a "transition time t_L" that is based on the historically low correlations commonly used when treating stochastic process in acoustics and other sciences (fluid dynamics, meteorology, and so on), namely $r(t)=1/e=0.367$ (Tatarskii, 1961; Skudrzyk, 1971; Ando, 1977; Bass and Fuks, 1979; Ishimaru, 1997). This definition is also an extension of the correlation length[1] in correlation analysis (Papoulis, 1984, Chap. 1). It seems definite that for this transition time the sound field has become incoherent compared to its initial state at $t=0$.

Considering that the definition of a reverberation decay curve assumes a diffuse sound field, t_L is the time when the total sound field (ER plus LR) of a hall changes to a sound field that is dominated by the energy of the reverberant sound alone.

E. Discussion

The time t for curve B in Fig. 1 is taken to be large enough that all the reflections involved have undergone multiple reflections in the hall. This frequency curve is equal to the steady state response of the RIR including only the LR. The amplitude $\rho=\sqrt{E(t,\omega)}$ obeys approximately the Rayleigh distribution because of the phase randomization (Schroeder, 1987; Kuttruff, 1991). Then, introducing the normalized variable $z=\rho/\langle\rho\rangle$, one obtains the variance $\sigma_z^2=\langle z^2\rangle-\langle z\rangle^2$ equal to $4/\pi-1=0.27$ (Papoulis, 1984, Chap. 5). A measurement example of σ_z^2 (Fig. 3) shows that the assumption of a Rayleigh distribution is likely true. This result was obtained for all the halls in this study. Thus, one can show that the squared amplitude $E(t,\omega)=\rho^2$ obeys the exponential distribution, where the probability density function is given by following equation:

$$p(E)=\frac{1}{2\sigma_E^2}\exp\left(-\frac{1}{2\sigma_E^2}E\right). \tag{4}$$

As discussed above, the initial part of the RIR decreases with increasing time. When t is larger than several times MFP/c, multiple reflections occur and the randomization in RIR is emphasized gradually. As a result, one can expect that

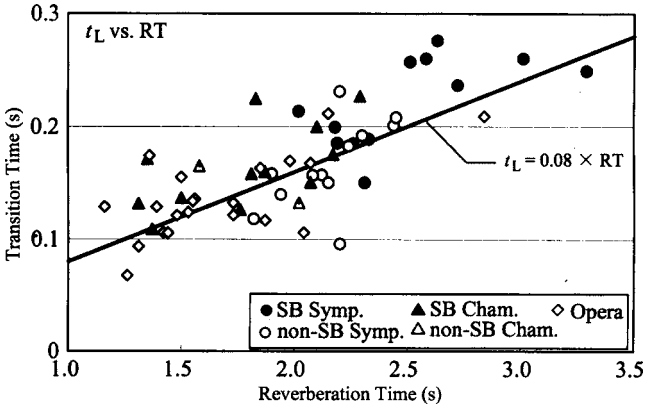


FIG. 4. Plot of the transition time versus the reverberation time. These values are the average of two receivers near the center positions at the main floor measured under the unoccupied state. The measured halls are divided into five categories: shoebox (SB) symphony hall, non-SB symphony hall, SB chamber hall, non-SB chamber hall, and opera house. The regression line (r=0.76) is shown. The frequency range is 500 Hz octave band.

$E(0,\omega)$ and $E(t,\omega)$ are stochastically independent of each other and the next approximation holds (Papoulis, 1984, Chap. 6),

$$\langle E(0,\omega)E(t,\omega)\rangle \cong \langle E(0,\omega)\rangle\langle E(t,\omega)\rangle. \tag{5}$$

One obtains from Eq. (4) calculating Nth moment $\langle E^N\rangle = \int_0^\infty E^N p(E)\,dE$,

$$\langle E\rangle = 2\sigma_E^2, \quad \langle E^2\rangle = 8\sigma_E^4, \tag{6}$$

so that the denominator of Eq. (3) converges to the finite value $4\sigma_{E(0)}^2\sigma_{E(t)}^2$. Then, from Eq. (2) it is concluded $r(t) \to 0$ for $t \to \infty$.

III. RESULTS OF ANALYSIS

The RIR was measured in each of a number of halls at receiving positions near the center of a hall, but off the centerline, with an omnidirectional source located on the centerline of the stage, usually located 3 m from the lip. The sampling frequency of the RIR was 44.1 kHz and the energy distribution of Eq. (1) was calculated for each 1/1 octave band.

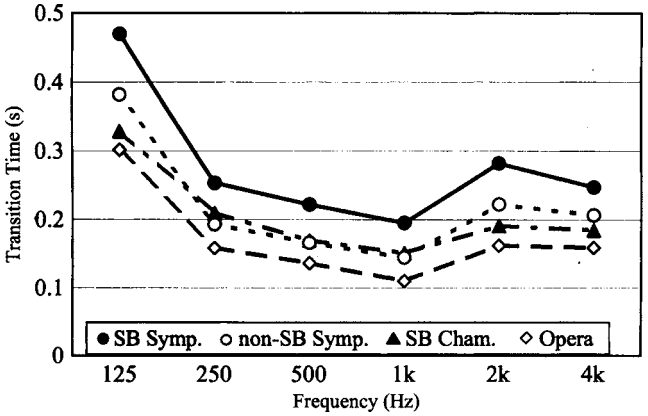


FIG. 5. Averaged values of the transition time for four category halls: 12 shoebox symphony halls, 12 non-shoebox symphony halls, 11 shoebox-chamber halls, and 22 opera houses. The measurement positions are the same as Fig. 4.

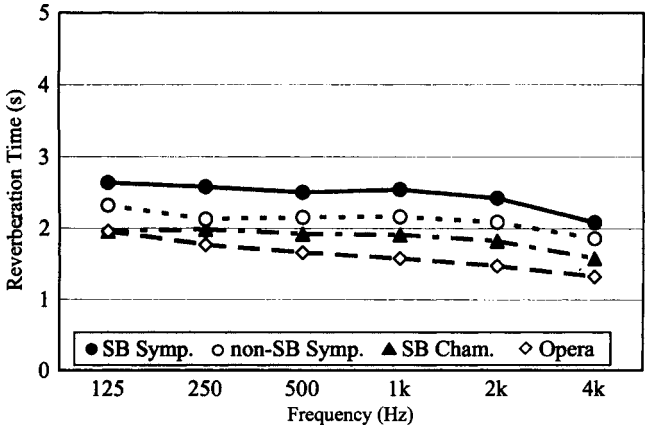


FIG. 6. Averaged values of the reverberation time for four category halls. Key is the same as Fig. 6.

Transitions times t_L are plotted for the 500-Hz band in Fig. 4 against reverberation time RT for 24 symphony halls (Beranek, 2004), 13 chamber music halls (Hidaka and Nishihara, 2004), and 22 opera houses (Hidaka and Beranek, 2000). The first two types of halls are further divided into shoebox (rectangular) and non-shoebox (other) shapes. The correlation coefficients between t_L and RT for each octave band are 0.66 (125 Hz), 0.75 (250 Hz), 0.76 (500 Hz), 0.82 (1k Hz), 0.79 (2k Hz), and 0.66 (4k Hz). When RT value is known, it is seen that the transition time for the 500-Hz octave band is roughly estimated by the regression equation.

$$t_L = 0.08 \times \mathrm{RT}\ (\text{in s}) \tag{7}$$

As shown in Fig. 4, there is a definite increase of t_L with regards to RT, which is consistent with Schroeder (1966), i.e., the longer the reverberation time, the later the LR will be achieved. Because the RT is determined from the simplified reverberation theory and the t_L is an actual measurement, the scatter of data from the regression line (±80 ms at maximum) is partially caused by that difference. In other words, t_L is not only determined by RT but also by that fact that the early part of RIR is influenced by the wave component.

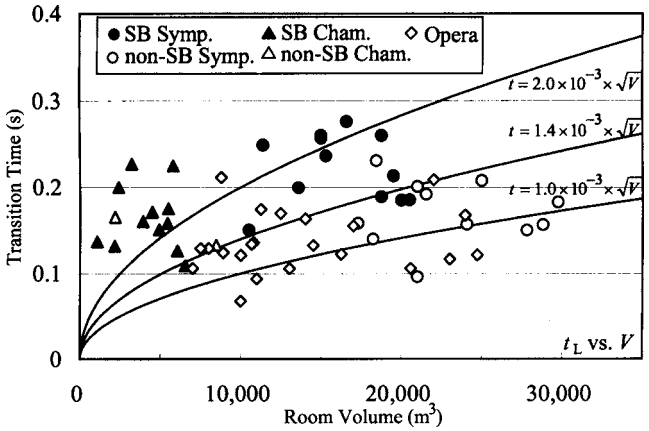


FIG. 7. The transition times for concert halls and opera houses are plotted against the room volume V. Key is the same as Fig. 4. Three solid curves are theoretical ones assuming Sabine's theory. The upper curve coincides with the *Nachhalleinsatzeit*.

TABLE I. Correlation coefficients between t_L and objective parameters. "ALL" means all the halls studied. The frequency range is 500 Hz octave band, but, for BQI and No. of reflections, 3 octave bands (500 Hz, and 1 kHz, and 2 kHz) are taken.

	V	S_A	RT	EDT	C_{80}	G	ITDG	BQI[a]	No. of reflections[a]
t_L (ALL)	0.14	−0.12	0.76	0.78	−0.77	0.21	−0.15	0.17	0.34
t_L (SB)	0.54	0.49	0.79	0.80	−0.76	−0.27	−0.07	−0.37	−0.64
t_L (non-SB)	0.15	0.05	0.48	0.51	−0.60	−0.28	0.08	−0.17	−0.31
t_L (Opera)	0.17	0.03	0.66	0.68	−0.54	−0.16	−0.05	−0.11	−0.12

3-octave band (500 Hz to 2 kHz)

Let us look at the ranges of t_L near an RT of 2.1 s. From Fig. 4 at 500 Hz, we see that the average values, given by the regression equation for each hall shape, plus/minus the residual standard deviation, are 190±29 for shoebox, 163±32 for non-shoebox, and 163±28 for opera. This indicates that t_L is dependent on the hall shape and is about 17% longer for a shoebox hall than for the other two shapes.

Figure 5 shows frequency dependence of t_L. The four plots are averaged values for each category: 12 symphony halls (SB), 12 symphony halls (non-SB), 11 chamber halls (SB), and 22 opera houses. Except for the lowest frequency band, the transition times are nearly constant with frequency. But at 125 Hz, t_L is double its values at higher frequencies. The mean reverberation times versus frequency for the four types of halls are shown in Fig. 6. This is another indication of why t_L is influenced by both RT and the hall shape.

The number of reflections per unit time in a rectangular room, with uniform but not large absorption, is $4\pi c^3 t^2/V$. Hence, if the density of reflections relates directly to the transition time t_L, it should be proportional to $V^{1/2}$. In Fig. 7, t_L is plotted against room volume V. A general increase is seen if one considers the two categories: (1) shoebox symphony and shoebox chamber halls, 110 to 280 ms, and (2) non-shoebox halls and opera houses, 70 to 210 ms. The upper curve $2\times10^{-3}V^{1/2}$ (in s) corresponds to the well-known *Nachhalleinsatzeit* (literal translation, "time of onset of reverberation") (Cremer and Müller, 1982), but big scatter is seen.

Let us explore the range of t_L for the different hall types, considering that in each type the mean ratio of the room volume to the acoustical seating area, V/S_A, is 13.5 m. Calculation for this ratio shows that t_L is dependent on the hall shape and that the ranges are 193±42 ms for SB, 158±30 ms for non-SB, and 143±34 ms for opera house, so that t_L for SB is 22% greater than that of non-SB. Beranek (2006) reports that RT for SB halls is 5% longer on average than that for non-SB halls when both V/S_A and the mean absorption coefficient α are the same, that is, similar types of seats are installed in each, which indicates that in a rectangular (SB) room with only one surface highly absorbent, i.e., the audience area, the early sound can persist longer in the upper space than in nonrectangular (non-SB) rooms where there is no upper space. From the Sabine equation, taking into account of the relation $V/S_A=\mathrm{RT}\alpha/0.161$, this result (22%) shows that t_L is more sensitive to the room shape than RT.

The correlations of t_L with other architectural and acoustical parameters are shown in Table I. High correlations with EDT and C_{80} are found because RT correlates highly with these two parameters (Hidaka, 2005). Additionally, the number of ER (with amplitudes greater than −20 dB relative to the direct sound) was counted for each hall (Hidaka and Nishihara, 2002) and this number shows little correlation with t_L. Also, the other orthogonal parameters (Beranek, 2004), i.e., audience area, binaural quality index BQI(=1 $-\mathrm{IACC}_{\mathrm{E3}}$), ITDG, and G, do not contribute to t_L.

Finally, Fig. 8 is an example of spatial distribution of t_L. The transition time is virtually constant in the greater part of the seating area, i.e., $240<t_L<320$ (ms) and has smaller values in the front and rear seats, because the sound field at the front and rear areas is physically simpler. At front seats, direct sound weakens the influence of the successive reflections, and at rear seats, the times intervals between early reflections are short so that the onset time of the reverberation begins sooner.

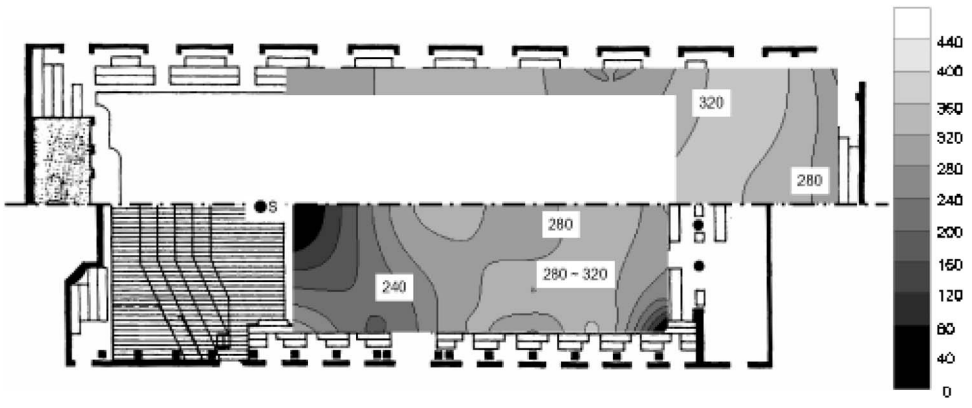


FIG. 8. Contour plots of the transition time measured in Musikvereinssaal Vienna (1680 seats). This quantity virtually takes constant value, except the front and rear seating areas in the hall. Each gray color means 40-ms step. The frequency range is 500 Hz octave band. S is a sound source on the centerline and at 3 m from the lip. Twenty-nine receiving points are used for the interpolation.

IV. CONCLUSION

The objective parameter t_L, proposed in this paper, is a definite means for separating the room impulse response RIR into two parts, that part which correlates meaningfully with the initial sound and that part which has low correlation ($r(t) \leqslant 1/e = 0.37$). If the classical separation of about 80 ms were used, $r(t)$ would equal 0.6 on average for large concert halls, indicating a substantial presence of ER in the later physical state. The time t_L is a scale value based on the correlation analysis in its own terms, which is in a similar situation as other room acoustical parameters. However, t_L is the only physical quantity that is derived from RIR measured in actual halls, from which one may judge that the sound field makes the transition from ER to LR.

For 57 actual halls for music performance t_L varies from 70 to 280 ms as shown in Fig. 4. Although there is wide scatter, the median for a shoebox symphony hall is 225 ms and the median for a shoebox chamber is 160 ms, while opera houses take about one-half of the former, 130 ms, as a median. The *Nachhalleinsatzeit* gives an intermediate value for shoebox halls but with big scatter (see the upper solid curve in Fig. 7). Transition time depends on RT, but does not show explicit dependence on simple architectural parameters like V or S_A, that is, 57% ($r^2 = 0.76^2$) of t_L is determined by RT but the remaining 43% is relevant to other factors. Transition time is related to room shape. In other words, one might advocate that t_L depends on not a simple number (or a density) of reflections but the intrinsic combination of the early reflections in RIR.

Transition time t_L is an independent parameter from basic architectural quantities, volume and audience area, and conventional objective parameters such as G, BQI, and so on. It is found that mean t_L takes constant value, except at the 125-Hz band, and shows small variation at most seating areas in the hall. It is interesting that the magnitude of t_L, 70 to 300 ms, is nearly equal to or slightly greater than the conventional values of 80 to 200 ms that are based on subjective or hybrid premises.

Transition time has no direct relation to subjective attribute in a hall, but gives one of the physical aspects of the sound field in it. By introducing the transition time, an objective premise can be established for numerical simulation or auralization of sound field in concert halls. Further research on the relationship between the transition time and subjective impressions of halls might demonstrate other aspects in room acoustics.

ACKNOWLEDGMENT

The authors wish to thank Dr. Leo Beranek with his fruitful discussions and kind editorial assistance. They also appreciate the careful reading of the manuscript by the reviewers and their helpful comments in revising this paper.

[1]For irregular processes of natural phenomenon such as atmospheric fluctuations, the correlation function has its maximum at a time lag equal to 0 and decreases with time. Correlation length (sometimes called correlation time) is defined as a measure of average distance from a point beyond which there is no further correlation of a physical property associated with that point. Values for a given property at distances beyond the correlation length can be considered purely random. This length is only a characteristic temporal scale value, because the correlation analysis is a tool to cope with not deterministic but stochastic phenomenon.

Ando, Y. (**1977**). "Subjective preference in relation to objective parameters of music sound fields with a single echo," J. Acoust. Soc. Am. **62**, 1436–1441.

Barron, M. (**1993**). *Auditorium Acoustics and Architectural Design* (E. & FN Spon, London), p. 17.

Bass, F. G., and Fuks, I. M. (**1979**). *Wave Scattering from Statistically Rough Surfaces* (Pergamon, Oxford), Chap. 1.

Beckmann, P., and Spizzichino, A. (**1987**). *The Scattering of Electromagnetic Waves from Rough Surfaces* (Artech House, Boston), Chap. 5.

Beranek, L. L. (**1992**). "Concert hall acoustics—1992," J. Acoust. Soc. Am. **82**, 1–39.

Beranek, L. L. (**2004**). *Concert Halls and Opera Houses—Music, Acoustics, and Architecture* (Springer, New York).

Beranek, L. L. (**2006**). "Analysis of Sabine and Eyring equations and their application to concert hall audience and chair absorption," J. Acoust. Soc. Am. **120**, 1399–1410.

Bowman, J. J., Senior, T. B. A, and Uslenghi, P. L. E. (**1987**). *Electromagnetic and Acoustic Scattering by Simple Shapes* (Hemisphere, New York), Chaps. 6 and 8.

Cremer, L., and Müller, H. A. (**1982**). *Principles and Applications of Room Acoustics* (Applied Science, London), Chap. III.2.1.

Ebeling, K. J. (**1984**). "Statistical properties of random wave fields," in *Physical Acoustics* (Academic, New York), vol. **17** pp. 233–311.

Gardner, W. G. (**1998**). *Applications of Digital Signal Processing to Audio and Acoustics*, edited by M. Kahrs (Kluwer Academic, Boston), Chap. 3.

Heinz, R. (**1993**). "Binaural room simulation based on an image source model with addition of statistical methods to include the diffuse sound scattering of walls and to predict the reverberation tail," Appl. Acoust. **38**, 145–159.

Hidaka, T., Beranek, L. L., and Okano, T. (**1995**). "Interaural cross-correlation, lateral fraction, and low- and high-frequency sound levels as measures of acoustical quality in concert halls," J. Acoust. Soc. Am. **98**, 988–1007.

Hidaka, T., and Beranek, L. L. (**2000**). "Objective and subjective evaluations of 23 opera houses in Europe, Japan and the Americas," J. Acoust. Soc. Am. **107**, 368–383.

Hidaka, T., and Nishihara, N. (**2002**). "On the objective parameter of texture," Proc. of Forum Acusticum Seville, September 16–20.

Hidaka, T., and Nishihara, N. (**2004**). "Objective evaluation of chamber-music halls in Europe and Japan," J. Acoust. Soc. Am. **116**, 357–372.

Hidaka, T. (**2005**). "Supplemental data of dependence of objective room acoustical parameters on source and receiver positions at field measurement," Acoust. Sci. & Tech. **26**, 128–135.

Ishimaru, A. (**1997**). *Wave Propagation and Scattering in Random Media* (IEEE, New York), Chap. 16.3.

Junius, von W. (**1959**). "Raumakustische Untersuchungen mit neuren Messverfahren in der Liederhalle Stuttgart (Room acoustical investigations with new measuring procedures at the Liederhalle Stuttgart)," Acustica **9**, 289–303.

Krokstad, A., Strøm S., and Sørsdal, S. (**1983**) "Fifteen years' experience with computerized ray tracing," Appl. Acoust. **16**, 291–312.

Kuttruff, H. (**1991**). "On the audibility of phase distortion in rooms and its significance for sound reproduction and digital simulation in room acoustics," Acustica **74**, 3–7.

Kuttruff, H. (**1993**). "Auralisation of impulse responses modeled on the basis of ray-tracing results," J. Audio Eng. Soc. **41**, 876–880.

Kuttruff, H. (**2000**). *Room Acoustics* (E. & FN Spon, London), Chap. 3.

Papoulis, A. (**1984**). *Probability, Random Variables, and Stochastic Processes*, 2nd ed. (McGraw-Hill, New York), Chap. 1.

Reichardt, W., Abdel Alim, O., and Schmidt, W. (**1974**). "Abhängigkeit der Grenzen zwischen brauchbarer und unbrauchbarer Durchsichtigkeit von der Art des Musikmotives, der Nachhallzeit und der Nachhalleinsatzzeit (Dependence of borders between usable and useless transparency on the type of music motives, reverberation time and beginning time of reverberation)," Appl. Acoust. **7**, 243–264.

Schroeder, M. J. (**1959**). "Measurement of sound diffusion in reverberant chamber," J. Acoust. Soc. Am. **31**, 1407–1414.

Schroeder, M. J. (**1962**). "Natural sounding artificial reverberation," J. Audio Eng. Soc. **10**, 219–223.
Schroeder, M. J. (**1965**). "New method of measuring reverberation time," J. Acoust. Soc. Am. **37**, 409–412.
Schroeder, M. J. (**1966**). "Complementarity of sound buildup and decay," J. Acoust. Soc. Am. **40**, 549–551.
Schroeder, M. J. (**1987**). "Statistical parameters of the frequency response curves of large rooms," J. Audio Eng. Soc. **35**, 299–305.
Skudrzyk, E. (**1971**). *The Foundations of Acoustics* (Springer, Vienna), Chap. VII.
Tatarskii, V. I. (**1961**), *Wave Propagation in a Turbulent Medium* (McGraw-Hill, New York). Chap. 1.
Vorländer, M. (**1989**). "Simulation of the transient and steady-state sound propagation in rooms using a new combined ray-tracing/image-source algorithm," J. Acoust. Soc. Am. **86**, 172–178.
Vorländer, M. (**1995**). "Revised relation between the sound power and the average sound pressure level in rooms and for acoustics measurements," Acustica **81**, 332–343.
Yamada, Y., and Hidaka, T. (**2001**). "Application of Page-Levin distribution for the evaluation of transient responses in rooms," Proceedings of 17th ICA, Rome, September 2–7.

Characteristics of surface sound pressure and absorption of a finite impedance strip for a grazing incident plane wave

K. S. Sum[a] and J. Pan[b]
School of Mechanical Engineering, The University of Western Australia, 35 Stirling Highway, Crawley, Western Australia 6009, Australia

(Received 22 August 2006; revised 4 February 2007; accepted 10 February 2007)

Distributions of sound pressure and intensity on the surface of a flat impedance strip flush-mounted on a rigid baffle are studied for a grazing incident plane wave. The distributions are obtained by superimposing the unperturbed wave (the specularly reflected wave as if the strip is rigid plus the incident wave) with the radiated wave from the surface vibration of the strip excited by the unperturbed pressure. The radiated pressure interferes with the unperturbed pressure and distorts the propagating plane wave. When the plane wave propagates in the baffle-strip-baffle direction, it encounters discontinuities in acoustical impedance at the baffle-strip and strip-baffle interfaces. The radiated pressure is highest around the baffle-strip interface, but decreases toward the strip-baffle interface where the plane wave distortion reduces accordingly. As the unperturbed and radiated waves have different magnitudes and superimpose out of phase, the surface pressure and intensity increase across the strip in the plane wave propagation direction. Therefore, the surface absorption of the strip is nonzero and nonuniform. This paper provides an understanding of the surface pressure and intensity behaviors of a finite impedance strip for a grazing incident plane wave, and of how the distributed intensity determines the sound absorption coefficient of the strip. © *2007 Acoustical Society of America.* [DOI: 10.1121/1.2713710]

PACS number(s): 43.55.Ev, 43.55.Dt, 43.20.El [NX] Pages: 333–344

I. INTRODUCTION

The relationship between the plane wave reflection coefficient, R_L, and the specific normal acoustical impedance, ζ_L, of an infinitely flat boundary of local reaction (also known as an impedance boundary) is well known and given by

$$R_L = \frac{\zeta_L \cos\theta - 1}{\zeta_L \cos\theta + 1}, \tag{1}$$

where θ is the incident angle of the plane wave relative to the normal of the boundary. Equation (1) relates the impedance to the sound absorption coefficient of the boundary for plane wave excitations at oblique incidence and diffuse incidence.[1–3] It also allows the impedance to be related to the decay rates of acoustic modes[4,5] and diffuse sound fields in rooms.[1,6] In addition, the equation has been extensively used in the development of many experimental techniques for measuring surface impedances of porous and nonporous materials.[7–9] It is also a basis for the image method[10] in architectural acoustics, and numerous studies of sound fields above boundaries of local or extended reaction in outdoor sound propagation.[11–14]

However, at $\theta=90°$ when the plane wave grazes the impedance boundary, Eq. (1) becomes problematic for any nonzero value of ζ_L. In this case, $R_L=-1$ where the incident and reflected waves have an equal magnitude but are 180° out of phase when they completely cancel each other. Therefore, there is no sound pressure above the boundary and this is practically not possible. Also, the sound absorption coefficient of the boundary ($=1-|R_L|^2$) is zero, which implies that the boundary does not absorb sound at all. This condition can still exist even though monopole or dipole sources are used if the emitted spherical waves are decomposed into plane wave components.[8] It can also be present in the absorption of periodically arranged absorptive materials as long as Eq. (1) is invoked.[15] Traditionally, workers have avoided the problem with grazing plane waves by using a point source and considering the incident wave to be spherical, where the sound field above the boundary is described in terms of the Sommerfeld integrals as well as an integral with a Bessel function of zero order.[11–14,16,17] Although these integrals do not lead to a zero pressure for a grazing incident wave, many relevant past studies were mainly mathematical, and much effort was focused on investigating the characteristics of the integral with the Bessel function and the pole of this integral.[13,16,17] In addition, this integral is too complicated for an exact analytical solution of the sound pressure to be obtained so asymptotic analytical solutions have been proposed for any incident angle of the spherical wave. But, most reported works that employed these solutions only dealt with either the application/development of methods for predicting the sound pressure above the boundary,[11,12,14] or the formulation and validation of techniques for predicting the impedance and absorption coefficient of the boundary.[18–22]

In rooms, boundaries are finite and acoustical mode shapes can be observed when the sound fields are not diffuse. For rectangular rooms, it is known that these modes are identifiable and composed of propagating plane waves in regions which are sufficiently far from the sound source, even when the source emits spherical waves. Thus, the planar nature of

[a]Electronic mail: ksum@mech.uwa.edu.au
[b]Electronic mail: pan@mech.uwa.edu.au

the incident wave that impinges upon a boundary cannot be simply neglected. However, based on experimental observations that absorption coefficients of impedance surfaces in rectangular rooms are nonzero (the surfaces absorb sound) for grazing plane waves,[23] reflected wave fronts were believed to have curvatures when incident angles of plane waves are far from the normals of the surfaces. From a recent study, one may also expect these curvatures to be significant for highly absorptive surfaces and at very low frequencies as under such conditions, the sound pressures in rectangular rooms could not be accurately predicted using the plane wave reflection coefficient [i.e., Eq. (1)].[24] These past works focused on the measurement of surface absorption coefficients at resonances of rooms[23] and accuracy of predictive models for room acoustics.[24] Also, many published studies on absorptions of impedance surfaces in rooms only either reported experimental data of absorption coefficients to compare the size/edge effects of the surfaces,[25–27] or applied/developed methods for comparing these effects at normal, oblique, and random incidences of plane waves.[28–32] As a result, the physical picture of the reflected wave front curvature of an impedance boundary and its effect on the boundary absorption is still obscure. It follows that the reason for the curvature, its relationship to the boundary impedance, and the resulting boundary absorption are not clearly understood. Therefore, it is necessary to investigate how both the finiteness and impedance of the boundary determine the shape of the reflected wave front even though the incident wave is planar, and the way that the boundary absorption is related to the behavior of the reflected wave.

In this paper, characteristics of sound pressure and intensity on the surface of a flat and finite impedance strip flush-mounted on a rigid baffle are studied. The specific case of plane wave excitation at grazing incidence is considered. Previous studies for grazing incident plane waves have been concerned only with mathematical analyses of scattering due to rough surfaces.[33,34] The boundary integral method is used, where the surface pressure of the strip is expressed in terms of the superposition between the unperturbed wave and the radiated wave. This method is well established, and has been extensively validated and used for the prediction of pressures above impedance surfaces for spherical[35–37] and plane[3,28,32,38] wave excitations. The unperturbed wave consists of the specularly reflected wave as if the strip is rigid and the incident wave, while the radiated wave arises from the surface vibration of the strip excited by the unperturbed pressure. The total of the reflected and radiated waves is the scattered wave, while the total of the unperturbed and radiated waves is called the perturbed wave. Hence, the "surface pressure" and "surface intensity" terms refer to the perturbed pressure and intensity on the surfaces of both the baffle and the strip, unless one of the surfaces is otherwise specified. A numerical procedure called the quadrature technique[35,36,39] is employed to solve the boundary integral problem. As the present work is concerned with surface acoustical properties, singularities exist in the numerical solutions of surface pressure of the strip. Thus, the quadrature technique is modified to deal with the ill condition and the convergence of the solutions is examined. Examples are used to show the variation of radiated pressure on the surface of the strip with the impedance of the strip as well as with the excitation frequency. The distortion of the propagating plane wave along the strip by this pressure is then illustrated. Characteristics of the resulting surface pressure and intensity across the strip, and the sound absorption coefficient of the strip determined by the intensity distribution, are also studied. This paper provides a visualization of the way that the radiated pressure of a finite impedance strip distorts a plane wave that grazes the strip. It also provides an understanding of the radiated wave effects on the surface pressure and intensity of the strip, how the distributed intensity determines the absorption coefficient of the strip, and the reason why the absorption of the strip is not zero even though the incident wave is a grazing plane wave.

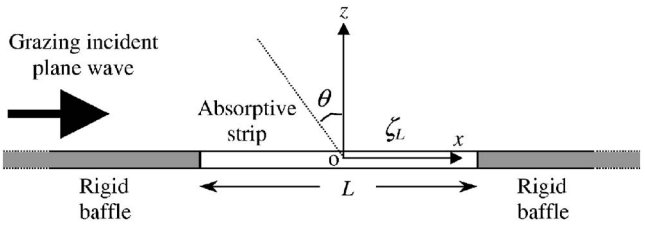


FIG. 1. Schematic illustration of the impedance strip flush-mounted on a rigid baffle and the grazing incident plane wave.

II. SURFACE PRESSURE AND INTENSITY OF THE STRIP

A. Formulation of the boundary integral problem

Figure 1 depicts an impedance strip which is flush-mounted on a rigid baffle. L and ζ_L are the length and the specific acoustical impedance of the strip. The coordinate system is also shown and its origin is located at the center of the strip. First, consider the steady-state excitation by a plane wave at an oblique incident angle of θ from the normal of the strip. By using the Green's function technique in the two-dimensional acoustic wave equation for the sound field on and above the baffle and the strip, the sound pressure, p, at any position (x,z) for $-\infty < x < \infty$ and $z \geqslant 0$, can be derived as[3,28,32,38]

$$p(x,z) = p_{\text{unpt}}(x,z) + p_{\text{rad}}(x,z), \tag{2a}$$

$$p_{\text{unpt}}(x,z) = 2P_{\text{inc}} e^{j\omega t - jkx \sin\theta} \cos(kz\cos\theta), \tag{2b}$$

$$p_{\text{rad}}(x,z) = \int_L G(x,z|x^{(s)},0) \left.\frac{\partial p}{\partial z}\right|_{z=0} dx^{(s)}. \tag{2c}$$

In Eq. (2b), P_{inc} is the complex pressure amplitude of the plane wave, t is time, ω is the radian frequency ($=2\pi f$; f is the frequency in hertz), and k is the wave number ($=\omega/c_0$; c_0 is the speed of sound in air). In Eq. (2c), G is the Green's function of the sound field for the plane wave excitation, and $x^{(s)}$ is the x coordinate of a point on the surface of the strip (i.e., $z=0$). As the strip is of local reaction, the boundary condition of the sound field on the surface of the strip is

$$\left.\frac{\partial p}{\partial z}\right|_{z=0} = j\rho_0\omega v(x^{(s)}) = \frac{jkp(x^{(s)},0)}{\zeta_L}, \tag{3}$$

where ρ_0 is air density, and v is the normal velocity of the surface [note: $v(x^{(s)})=p(x^{(s)},0)/\rho_0 c_0\zeta_L$]. p_{unpt} given by Eq. (2b) is the unperturbed pressure that consists of the combination of the specularly reflected pressure as if the strip is rigid and the incident pressure. On the surface of the strip, the unperturbed pressure is also widely known as the blocked pressure. p_{rad} given by Eq. (2c) depends on $\partial p/\partial z|_{z=0}$ that is a function of v [see Eq. (3)] and therefore, it represents the radiated pressure due to the surface vibration of the strip that is excited by the unperturbed pressure. p evaluated by Eq. (2a) is the perturbed pressure and specifically at $z=0$, it is called the surface pressure.

Upon splitting the incident pressure, p_{in}, from p_{unpt}, Eq. (2a) can be reexpressed as

$$p(x,z) = p_{\text{in}}(x,z) + p_{\text{scat}}(x,z), \tag{4a}$$

where

$$p_{\text{in}}(x,z) = P_{\text{inc}}e^{j\omega t - jkx\sin\theta + jkz\cos\theta}, \tag{4b}$$

$$p_{\text{scat}}(x,z) = P_{\text{inc}}e^{j\omega t - jkx\sin\theta - jkz\cos\theta} + p_{\text{rad}}(x,z). \tag{4c}$$

Equations (4a)–(4c) are similar to those for scattering by elastic objects,[40] where Eq. (4c) suggests that the interference of the reflected and radiated waves generates scattering. p_{scat} is the scattered pressure which implicitly consists of the specularly reflected pressure of the baffle and the strip, and the diffracted pressure of the two ends of the strip due to discontinuities in acoustical impedance.

From the above description, it can be seen that Eqs. (2a) and (4a) provide two different physical interpretations of the perturbed wave. Equation (2a) describes the difference between the cases with and without the vibration of the strip, and the perturbed wave results when the radiated wave from the vibrating strip interferes with the propagating plane wave as if the strip is rigid. On the other hand, Eq. (4a) describes the difference between the cases with and without the boundary at $z=0$, and the perturbed wave results when the scattered wave from the baffle and the vibrating strip interferes with the incident wave. Although both equations yield the same p, the behaviors of p_{rad} and p_{scat} are different. Since the behavior of the radiated wave provides a direct indication of how the vibrating strip affects the unperturbed wave that is of interest in the present work, only Eqs. 2(a)–2(c), p_{unpt}, and p_{rad} will be concerned in the following.

As the radiated wave is generated by the whole surface of the strip, G can be defined as the Green's function at (x,z) due to the radiation at $(x^{(s)},0)$, and it is given by[28,36,38]

$$G(x,z|x^{(s)},0) = jH^{(1,0)}[k\sqrt{(x-x^{(s)})^2+z^2}]/2, \tag{5}$$

where $H^{(1,0)}$ is the Hankel function of the first kind and zero order. By substituting Eqs. (3) and (5) into Eq. (2c), considering $\theta=90°$ in Eq. (2b) for a grazing incident plane wave that propagates from the left (see Fig. 1), and ignoring the alternating term, $e^{j\omega t}$, Eq. (2a) becomes

$$p(x,z) = 2P_{\text{inc}}e^{-jkx} - \frac{k}{2\zeta_L}\int_{-L/2}^{L/2} p(x^{(s)},0) \times H^{(1,0)}[k\sqrt{(x-x^{(s)})^2+z^2}]dx^{(s)}. \tag{6}$$

Given P_{inc}, ζ_L, k, and L, Eq. (6) indicates that the surface pressure distribution of the strip is required if the perturbed pressure at any position on or above the baffle or the strip is to be calculated. As $H^{(1,0)}$ is within the integral and depends on $x^{(s)}$, it is obvious from the equation that an exact analytical solution of $p(x,z)$ cannot be obtained even when $p(x^{(s)},0)$ is known. So, other ways of solving $p(x,z)$ have been employed, such as the variational procedure and numerical procedures. However, in the variational procedure, it was still too difficult to derive a full solution, and only approximated solutions for high frequencies (large k) and lightly absorptive surfaces (large ζ_L) have been obtained.[3,28,38] Thus, a numerical procedure called the quadrature technique[35,36,39] is used here.

B. Solutions by the quadrature technique

In this technique, the strip is discretized into M small elements of length h (i.e., $L=Mh$). The total radiated pressure on the surface of the strip is then the sum of radiated pressures at all the discrete elements (i.e., at each element, the radiated pressure is the surface pressure weighted by a $kh/2\zeta_L$ factor and $H^{(1,0)}$). Mathematically, the integral in Eq. (6) is replaced by a summation over the M elements and the equation is written as

$$\tilde{p}(x,z) = 2P_{\text{inc}}e^{-jkx} - \frac{kh}{2\zeta_L}\sum_{m=1}^{M}\tilde{p}(x_m^{(s)},0) \times H^{(1,0)}[k\sqrt{(x-x_m^{(s)})^2+z^2}], \tag{7}$$

where the tildes refer to approximated pressures because of the discretization. $x_m^{(s)}$ is taken as the x coordinate at the midpoint of the mth element such that $x_m^{(s)}=-L/2+(m-1/2)h$ [i.e., $x_1^{(s)}=(-L+h)/2$ and $x_M^{(s)}=(L-h)/2$].

Since $\tilde{p}(x_m^{(s)},0)$ is still not known but required before $\tilde{p}(x,z)$ can be evaluated, Eq. (7) must first be used to solve the perturbed pressure at each element on the surface of the strip. Therefore, at the nth element, $x=x_n^{(s)}$ and $z=0$, and the equation becomes

$$\tilde{p}(x_n^{(s)},0) + \frac{kh}{2\zeta_L}\sum_{m=1}^{M}\tilde{p}(x_m^{(s)},0)H^{(1,0)}(k|x_n^{(s)}-x_m^{(s)}|) = 2P_{\text{inc}}e^{-jkx_n^{(s)}}. \tag{8}$$

It is apparent from Eq. (8) that for $n=m$ in which $x_n^{(s)}=x_m^{(s)}$, the argument of $H^{(1,0)}$ is zero [i.e., $H^{(1,0)}(0)$], where the Hankel function is undefined and the equation is singular. Hence, there are M singularities for $n=1,\ldots,M$. Although this ill condition was mentioned in previous works,[36,39] exact ways of solving the problem have not been available for plane wave excitations. Even for spherical and cylindrical wave excitations, only an asymptotic way for large k and ζ_L has been given.[36]

In order to avoid those singularities in the present work, a mean value of the Hankel function, $\bar{H}_{n,m}^{(1,0)}$, is chosen for use in the evaluation of the surface pressure at the nth element due to the radiation at the mth element. It is defined as the average of the two values of the Hankel function when both end points of the mth element are treated as the radiation points instead of its midpoint:

$$\bar{H}_{n,m}^{(1,0)} = [H^{(1,0)}(k|x_n^{(s)} - x_m^{(s1)}|) + H^{(1,0)}(k|x_n^{(s)} - x_{m+1}^{(s1)}|)]/2. \tag{9}$$

In Eq. (9), $x_m^{(s1)} = -L/2 + (m-1)h$ so that $x_1^{(s1)} = -L/2$, $x_{M+1}^{(s1)} = L/2$, and $x_m^{(s)} = (x_m^{(s1)} + x_{m+1}^{(s1)})/2$ (note: $x_m^{(s)}$ corresponds to the midpoint, and $x_m^{(s1)}$ and $x_{m+1}^{(s1)}$ correspond to the end points of the mth element). As $x_m^{(s)} \neq x_m^{(s1)}$ and $x_m^{(s)} \neq x_{m+1}^{(s1)}$, there are no singularities. Thus, Eq. (8) can be rewritten as

$$\tilde{p}(x_n^{(s)},0) + \frac{kh}{2\zeta_L}\sum_{m=1}^{M} \tilde{p}(x_m^{(s)},0)\bar{H}_{n,m}^{(1,0)} = p_{\text{unpt}}(x_n^{(s)},0), \tag{10}$$

where $p_{\text{unpt}}(x_n^{(s)},0) = 2P_{\text{inc}}e^{-jkx_n^{(s)}}$ is the unperturbed pressure at the nth element. If $M \to \infty$, then $h \to 0$, $x_m^{(s)} \approx x_m^{(s1)} \approx x_{m+1}^{(s1)}$, and $\tilde{p}(x_m^{(s)},0)$ for $m = 1, \ldots, M$ will approach the continuous distribution of the surface pressure of the strip, $p(x^{(s)},0)$.

By considering M equations of the form of Eq. (10), a matrix equation can be constructed and the discrete surface pressure of the strip can be numerically solved as

$$\begin{bmatrix} \tilde{p}(x_1^{(s)},0) \\ \vdots \\ \tilde{p}(x_M^{(s)},0) \end{bmatrix} = \begin{bmatrix} 1 + kh\bar{H}_{1,1}^{(1,0)}/2\zeta_L & \cdots & kh\bar{H}_{1,M}^{(1,0)}/2\zeta_L \\ \vdots & \ddots & \vdots \\ kh\bar{H}_{M,1}^{(1,0)}/2\zeta_L & \cdots & 1 + kh\bar{H}_{M,M}^{(1,0)}/2\zeta_L \end{bmatrix}^{-1} \times \begin{bmatrix} p_{\text{unpt}}(x_1^{(s)},0) \\ \vdots \\ p_{\text{unpt}}(x_M^{(s)},0) \end{bmatrix}, \tag{11}$$

where $[\ ]^{-1}$ denotes the inverse of the matrix. Once the surface pressure distribution of the strip has been obtained, the radiated pressure at the nth element can be evaluated from the second term on the left-hand side of Eq. (10) as

$$\tilde{p}_{\text{rad}}(x_n^{(s)},0) = -\frac{kh}{2\zeta_L}\sum_{m=1}^{M} \tilde{p}(x_m^{(s)},0)\bar{H}_{n,m}^{(1,0)}. \tag{12}$$

In Eq. (12), a negative sign has been inserted because the radiated pressure was originally defined on the right-hand side of Eq. (2a). On the other hand, the radiated and perturbed pressures at elsewhere, $\tilde{p}_{\text{rad}}(x,z)$ and $\tilde{p}(x,z)$ for $z=0$ and $-\infty < x < -L/2$ or $L/2 < x < \infty$ (i.e., on the baffle), or $z > 0$ and $-\infty < x < \infty$ (i.e., above the baffle and the strip), can be calculated from Eq. (7). The radiated pressure is given by the last term in the equation as

$$\tilde{p}_{\text{rad}}(x,z) = -\frac{kh}{2\zeta_L}\sum_{m=1}^{M} \tilde{p}(x_m^{(s)},0)H^{(1,0)}[k\sqrt{(x-x_m^{(s)})^2 + z^2}]. \tag{13}$$

For any value of θ, the complex sound intensity of the strip at $(x_m^{(s)},0)$ in the z direction is written as

$$I_z(x_m^{(s)}) = -\tilde{p}(x_m^{(s)},0)\tilde{v}^*(x_m^{(s)})/2 = -|\tilde{p}(x_m^{(s)},0)|^2/2\rho_0 c_0 \zeta_L^*, \tag{14}$$

where $\tilde{v}(x_m^{(s)})$ $[=\tilde{p}(x_m^{(s)},0)/\rho_0 c_0 \zeta_L]$ is the normal velocity of the surface of the strip and the asterisk (*) denotes the complex conjugate. The negative sign in Eq. (14) indicates that the normal intensity flow into the strip is opposite to the z direction. From this equation, the surface intensity can be evaluated when the surface pressure and impedance of the strip are known. The sound absorption coefficient of the strip can then be obtained as

$$\alpha_L = \frac{1}{M}\sum_{m=1}^{M} \text{Re}[I_z(x_m^{(s)})]/-|P_{\text{inc}}|^2 \cos\theta/2\rho_0 c_0, \tag{15}$$

where Re[ ] denotes the real part of the complex quantity. The numerator on the right-hand side of Eq. (15) describes the spatial average real intensity (or absorbed sound power per unit length) of the strip. The denominator with a negative sign represents the intensity of the incident plane wave in the direction which is normal to the strip and opposite to the z direction.

When $\theta = 90°$ as is of interest here, it is obvious from Eq. (15) that the denominator becomes zero (due to $\cos\theta = 0$) and for any nonzero finite value of the numerator, α_L is infinite. In other words, when the definition of α_L in Eq. (15) is employed, α_L cannot describe the sound absorptivity of the strip for the excitation at grazing incidence. However, since θ is fixed throughout the present work because only $\theta = 90°$ is considered, the full intensity of the incident wave in the propagation direction (i.e., $-|P_{\text{inc}}|^2/2\rho_0 c_0$) can be reasonably used as a comparative measure for the extent of absorbed (real part) intensity of the strip. The absorption coefficient can then be redefined to provide a description of the absorptivity of the strip, and it is given by

$$\alpha_L = \frac{1}{M}\sum_{m=1}^{M} \text{Re}[I_z(x_m^{(s)})]/-|P_{\text{inc}}|^2/2\rho_0 c_0. \tag{16}$$

As will be shown later in Sec. III B, Eq. (16) gives $\alpha_L = 0$ (the strip does not absorb sound) when the length of the strip is infinite. Therefore, this definition of α_L is also consistent with Eq. (1) that an infinite impedance surface does not absorb sound when $\theta = 90°$.

III. RESULTS AND DISCUSSION

In the following, examples are presented to illustrate the effects of the radiated wave on the distributions of surface pressure and absorbed intensity of the strip for a grazing incident plane wave. The influence of the surface intensity distribution on the absorption coefficient of the strip is also revealed. From Eq. (10), as $\tilde{p}(x_n^{(s)},0)$ is dependent on $p_{\text{unpt}}(x_n^{(s)},0)$ that has the $e^{-jkx_n^{(s)}}$ term, the surface pressure solved by Eq. (11) will alternate in terms of kx, and the variations with k and L are the same for the surface pressure and intensity. In addition, for a given P_{inc} and ζ_L, the surface pressure and the intensity do not change with k and L as long as kL is the same. So, in the examples, $P_{\text{inc}} = 1.0$ Pa and $L = 3$ m are used, ζ_L and k are altered, but kL is quoted.

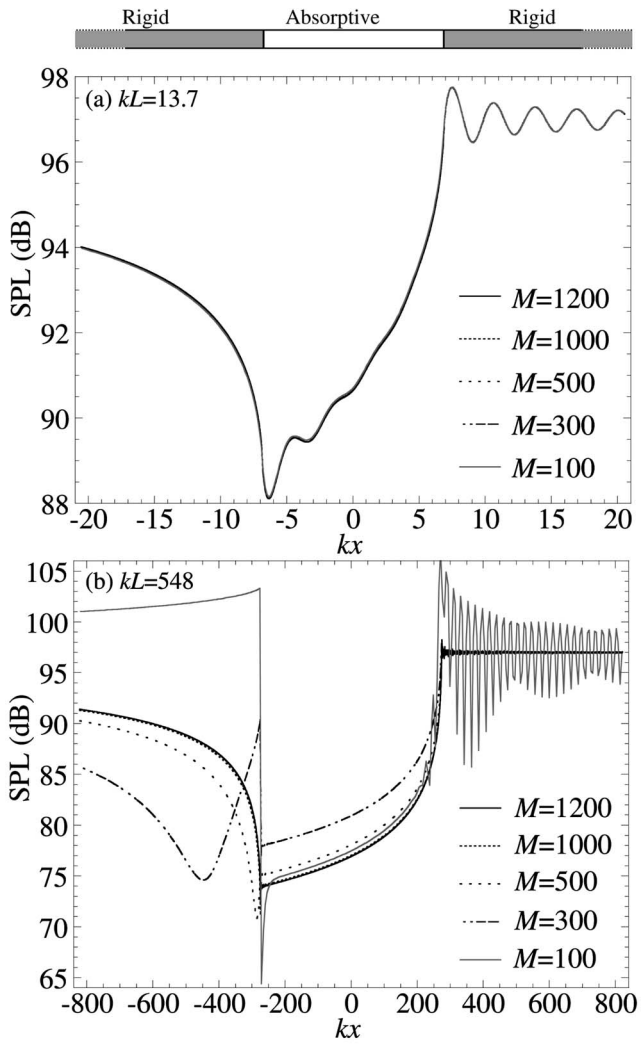


FIG. 2. Distributions of perturbed sound pressure level at $z=0$ for $P_{inc}=1.0$ Pa, $\zeta_L=0.2-j2$, and five different values of M. (a) $kL=13.7$ and (b) $kL=548$.

A. Convergence of surface pressure of the strip

As far as discretization is concerned, it is known that there must be at least four discrete elements within a wavelength, λ, for the frequency of a sinusoidal wave to be accurately identified (i.e., discretize the surface of the strip at quarter wavelengths). Mathematically, the corresponding length of each element, h_{conv}, and the number of elements, M_{conv}, are such that $h_{conv}=\lambda/4$ (or $L/M_{conv}=\pi/2k$), which implies $M_{conv}=2kL/\pi$. In this case, M_{conv} only provides a rough indication of the onset of convergence of the surface pressure of the strip [thus, the convergences of the radiated pressure and intensity on the surface of the strip from Eqs. (12) and (14), and the perturbed and radiated pressures at elsewhere from Eqs. (7) and (13)]. Hence, $M \gg 2kL/\pi$ if a perturbed or radiated wave with a fully converged pressure is to be obtained. As a result, a greater M is required for a larger kL.

Figure 2 depicts the distributions of perturbed sound pressure level (SPL) at $z=0$ for five different values of M, $\zeta_L=0.2-j2$, and $f=250$ Hz and 10 kHz (i.e., $kL=13.7$ and 548, and correspondingly, $M_{conv}\approx 9$ and 349). When kL is small, λ is large compared to L, and the surface pressure of the strip varies slowly with distance such that a low resolution of discretization (i.e., small M or large h since $h=L/M$) is adequate. As an example, Fig. 2(a) shows that for the small kL of 13.7, the perturbed SPL has already converged even when $M=100$ ($\gg M_{conv}$ of 9), and a further increase in M does not produce any distinguishable change in the pressure distribution. When kL is large, λ is small compared to L, and the surface pressure of the strip varies rapidly with distance. A high resolution of discretization (i.e., large M or small h) is thus required to avoid losing details of the pressure within small intervals on the surface. For example, as shown in Fig. 2(b) for the large kL of 548, there are still considerable errors in the perturbed SPL when $M=100$ or even 300 ($<M_{conv}$ of 349). It is also evident that the entire SPL distribution has already started to converge when $M=500$, and a complete convergence can be achieved only when $M>1000$. As the discretization is only related to kL, ζ_L does not affect the convergences of the perturbed and radiated pressures, and the surface intensity of the strip. So, the value of M is selected on the basis that the surface pressure of the strip has already well converged for a given kL. In the following, based on a full convergence of this surface pressure for $kL=548$ which is the largest value considered, $M=1500$ is used for all values of $kL\leq 548$ and is well sufficient.

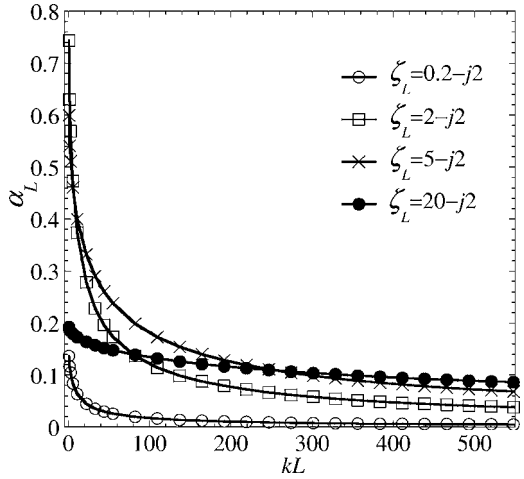


FIG. 3. Sound absorption coefficient of the strip vs kL for four different values of ζ_L.

B. Sound absorption coefficient of the strip

α_L vs kL for four different values of ζ_L is presented in Fig. 3, and two important features can be drawn from the figure. First, α_L is clearly nonzero (the strip absorbs sound) and different for different values of ζ_L even though the incident wave is a grazing plane wave. The variation of α_L with kL resembles an exponential decay (although it is not) where α_L decreases rapidly with kL. Due to this trend of variation, for any of the values of ζ_L shown, α_L is higher for a low frequency and a short strip than for a high frequency and a long strip. Also, $\alpha_L=0$ only if $kL\to\infty$ when the frequency and/or length of the strip is infinite. The latter case is consistent with Eq. (1) for an infinite impedance surface, which

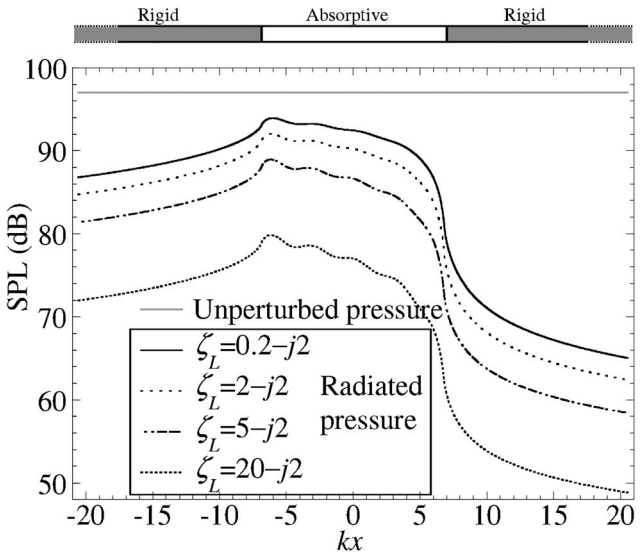


FIG. 4. Distributions of unperturbed and radiated sound pressure levels at $z=0$ for $P_{\text{inc}}=1.0$ Pa, $kL=13.7$, and four different values of ζ_L.

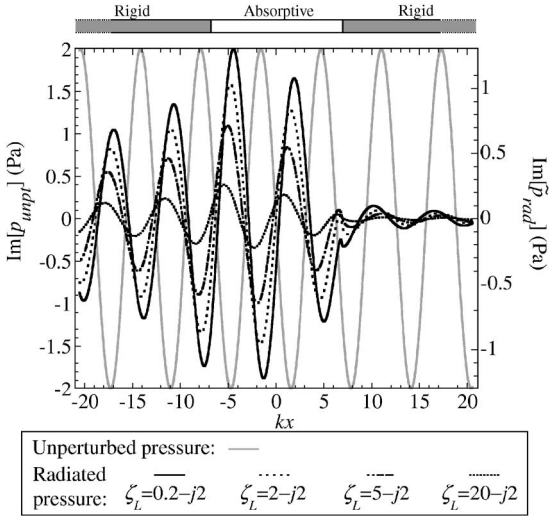


FIG. 5. Imaginary parts of the unperturbed and radiated sound pressures at $z=0$ for $P_{\text{inc}}=1.0$ Pa , $kL=13.7$, and four different values of ζ_L. Im[ ] denotes the imaginary part of the complex quantity.

gives a zero absorption coefficient ($R_L=-1$) for $\theta=90°$ even when ζ_L is finite and nonzero. As far as the second feature is concerned, in general, α_L for $kL<10$ is highest when the real part of ζ_L is close to 1, where the resistance (real part of impedance) of the strip is close to the characteristic impedance of air (i.e., in Fig. 3, at the first four values of $kL<10$, α_L is highest for $\zeta_L=2-j2$, followed by $\zeta_L=5-j2$, $20-j2$, and $0.2-j2$). However, it can be seen from the figure that α_L decreases with kL more rapidly for a small ζ_L than for a large ζ_L. As a result, α_L is eventually higher for a large ζ_L than for a small ζ_L (e.g., from $kL>10$ onwards, α_L is higher for $\zeta_L=5-j2$ than for $\zeta_L=2-j2$, from $kL>80$ onwards, α_L is higher for $\zeta_L=20-j2$ than for $\zeta_L=2-j2$, and from $kL>270$ onwards, α_L is higher for $\zeta_L=20-j2$ than for $\zeta_L=5-j2$). The above-mentioned two features of α_L are consecutively explained in the following, where the behavior of the radiated wave on the surface of the strip and the effect of the radiated pressure on the surface absorption of the strip are described in detail. The case of a small kL is first illustrated.

C. Variations of pressures and intensity with ζ_L

Figure 4 presents the distributions of radiated SPL at $z=0$ for $kL=13.7$ and the four values of ζ_L as before. Also shown is the uniform distribution of unperturbed SPL that is associated with the unperturbed wave. The radiation of the strip is stronger if the impedance difference between the strip and the baffle is greater (i.e., smaller ζ_L since the baffle has an infinite impedance) [note: From Eq. (12) or (13), $\tilde{p}_{\text{rad}} \propto 1/\zeta_L$]. Therefore, the radiated SPL is lowest for $\zeta_L=20-j2$ and highest for $\zeta_L=0.2-j2$, and it converges toward the unperturbed SPL as ζ_L decreases (see Fig. 4). When the plane wave propagates from the left baffle to the strip, it is strongly perturbed by a sudden change to a finite impedance at the baffle-strip interface where a strong radiation of the strip is induced. As there are no discontinuities in impedance when the plane wave propagates along the strip, the perturbation reduces gradually across the strip and a decreasing radiation is induced. When the plane wave propagates from the strip to the right baffle, the perturbation is further reduced by a sudden change back to an infinite impedance at the strip-baffle interface where the radiation of the strip is thus considerably weak. Consequently, it is obvious from Fig. 4 that when one proceeds from the left baffle to the strip, the radiated SPL increases until a maximum value at the baffle-strip interface. Then, it decreases slowly across the strip and exhibits a sharp drop around the strip-baffle interface after which a slow drop continues. This radiated pressure interferes with the unperturbed pressure, so the superposition of the radiated and unperturbed waves on the surfaces of the baffle and the strip is now examined.

Recall that $\tilde{p}(x_n^{(s)},0)$ obtained from Eq. (11) alternates in terms of kx because $p_{\text{unpt}}(x_n^{(s)},0)$ has the $e^{-jkx_n^{(s)}}$ term. From Eq. (12) or (13), $\tilde{p}_{\text{rad}}$ is dependent on $\tilde{p}(x_n^{(s)},0)$, and it follows that $\tilde{p}_{\text{rad}}$ also alternates in terms of kx. The alternating manner of p_{unpt} and $\tilde{p}_{\text{rad}}$ is described in their real and imaginary parts. Both parts are similar in wave form but different by 90° in phase, and they provide similar information of the pressures. Hence, only the imaginary parts of p_{unpt} and $\tilde{p}_{\text{rad}}$ at $z=0$ are presented in Fig. 5 for the same conditions as in Fig. 4. From Fig. 5, the unperturbed pressure has a constant amplitude that produces the uniform distribution of the unperturbed SPL in Fig. 4. Also, the amplitude of the radiated pressure decreases with ζ_L in Fig. 5, which corresponds to the drop of the radiated SPL with ζ_L observed in Fig. 4. Since the unperturbed wave is for the case as if the strip is rigid (i.e., $\zeta_L \to \infty$), in general, the unperturbed and radiated pressures are in phase to each other if the impedance difference between the strip and the baffle is small (i.e., large ζ_L), and out of phase if the impedance difference is large (i.e., small ζ_L). However, when ζ_L is large, the radiation of the strip will not have any significant influence on the unperturbed wave because the radiated pressure is negligibly low compared to the unperturbed pressure. In other words, ζ_L has to be small for the radiated pressure to be comparable to the unperturbed pressure (which is immediately clear from Figs. 4 and 5),

where both pressures are consequently out of phase to each other. Thus, as can be seen in Fig. 5 for all the values of ζ_L shown, the unperturbed and radiated waves superimpose out of phase. It is also evident that the latter shifts gradually as ζ_L increases, and its phase difference to the former slowly reduces. Hence, the phase difference between the unperturbed and radiated pressures is large when ζ_L is small (e.g., $\zeta_L=0.2-j2$, $2-j2$, and $5-j2$ in Fig. 5 where the pressures are mainly close to 180° out of phase to each other), and small when ζ_L is large (e.g., $\zeta_L=20-j2$ in Fig. 5 where the pressures are mainly away from 180° out of phase).

Figure 6 shows the distributions of perturbed SPL at $z=0$ (i.e., surface pressure) after the radiated and unperturbed waves have been superimposed. By comparing these distributions to those in Fig. 4, it can be seen that the behavior of the perturbed SPL is opposite to that of the radiated SPL, because the unperturbed and radiated waves superimpose out of phase as has been shown in Fig. 5 (i.e., the surface pressure is high when the radiated pressure is low and vice versa). For example, the surface pressure throughout the strip is highest for $\zeta_L=20-j2$ and lowest for $\zeta_L=0.2-j2$. At the baffle-strip interface where the maximum of the radiated SPL occurs, a minimum surface pressure is found. Also, around the strip-baffle interface where the sharp drop of radiated SPL occurs, a rapid increase of surface pressure is observed (compare Figs. 4 and 6). Since the radiated pressure is higher for a lower ζ_L, it affects the unperturbed pressure considerably more for a low ζ_L than for a high ζ_L. So, although the slopes of decrease in radiated SPL over the strip are almost the same for all four values of ζ_L (see Fig. 4), Fig. 6 indicates that the slopes of increase in surface pressure over the strip are substantially different (i.e., steepest for $\zeta_L=0.2-j2$, followed by $\zeta_L=2-j2$, $5-j2$, and $20-j2$). This surface pressure distribution determines two acoustical aspects of the strip, namely, the profile of the plane wave front above the strip and the absorbed intensity distribution on the surface of the strip.

In order to visualize the radiated wave effect on the profile of the plane wave front, the perturbed SPL at $z>0$ is also evaluated for the same conditions as in Fig. 6. Figure 7 shows the distributions of perturbed SPL on and above the baffle and the strip at five locations. Two of the locations are above the baffle (before the baffle-strip interface and after the strip-baffle interface), and the other three locations are above the strip (at the middle of the strip and in the vicinity of both interfaces). As $\tilde{p}(x,z)$ depends on $\tilde{p}(x_m^{(s)},0)$ and $H^{(1,0)}$ in Eq. (7), and the magnitude of $H^{(1,0)}$ is large when $kz \ll 1$, the perturbed pressure at $z>0$ around the strip is close to the surface pressure of the strip and hence, deviates from the unperturbed pressure [note: In Eq. (7), $\tilde{p}(x_m^{(s)},0)$ is weighted by $H^{(1,0)}$]. The perturbed pressure approaches the unperturbed pressure when kz increases because the magnitude of $H^{(1,0)}$ reduces with kz [i.e., the value of the last term in Eq. (7) that describes the radiated pressure, reduces with kz]. As a result of the deviation from the unperturbed pressure, pressure curvatures are generated on the perturbed wave and the plane wave front becomes distorted. The distortion of the propagating plane wave is clearly shown in Fig. 7 for the four values of ζ_L, where the curvature of the wave front is

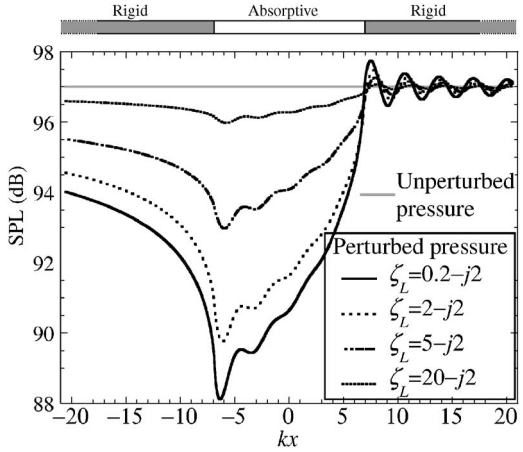


FIG. 6. Distributions of unperturbed and perturbed sound pressure levels at $z=0$ for $P_{inc}=1.0$ Pa, $kL=13.7$, and four different values of ζ_L.

large on and near the surfaces of the baffle and the strip, but decreases with the distance from the surfaces (i.e., z). Since the discrepancy between the perturbed and unperturbed SPLs in Fig. 6 reduces with ζ_L, the distortion also decreases with ζ_L [e.g., from Figs. 7(a) to 7(d), the distortion is largest for $\zeta_L=0.2-j2$, followed by $\zeta_L=2-j2$, $5-j2$, and $20-j2$]. Following the observations in Figs. 4 and 6, Fig. 7 also indicates that when one proceeds from the left baffle to the strip, the distortion increases to the largest at the baffle-strip interface, decreases gradually across the strip, and finally becomes insignificant around the strip-baffle interface and beyond. In other words, the surface pressure and the profile of the plane wave front in the vicinity of the baffle-strip interface are most significantly affected by the radiation of the strip.

Due to the nonzero surface pressure of the strip, which varies with ζ_L and has a nonuniform distribution as in Fig. 6, the absorbed intensity on the surface of the strip is also nonzero, varies with ζ_L, and has a nonuniform distribution as depicted in Fig. 8 [note: From Eq. (14), the surface intensity is directly proportional to the surface pressure]. This explains the first feature from Fig. 3 that α_L is nonzero and different for the four values of ζ_L. It can be seen from Fig. 8 that the slope of increase in intensity over the strip drops with ζ_L, which is caused by the similar variation of the slope of increase in surface pressure over the strip with ζ_L (see Fig. 6). So, the surface absorption of the strip is highly and most nonuniform for $\zeta_L=0.2-j2$, but nearly and most uniform for $\zeta_L=20-j2$. In addition, the intensity is lowest around the baffle-strip interface that has the maximum radiated pressure, and increases across the strip until a maximum value at the strip-baffle interface that has the minimum radiated pressure. As the product of the surface pressure and $\mathrm{Re}[1/\zeta_L^*]$ is largest for $\zeta_L=5-j2$ and smallest for $\zeta_L=0.2-j2$ {i.e., from Eq. (14), $\mathrm{Re}[I_z(x_m^{(s)})] \propto |\tilde{p}(x_m^{(s)},0)|^2 \mathrm{Re}[1/\zeta_L^*]$ can be deduced}, the overall absorbed intensity is highest for $\zeta_L=5-j2$, followed by $\zeta_L=2-j2$, $20-j2$, and $0.2-j2$ (see Fig. 8). So, at $kL=13.7$, α_L is higher for $\zeta_L=5-j2$ than for $\zeta_L=2-j2$ as observed in Fig. 3.

D. Variations of pressures and intensity with *kL*

In this section, the radiated wave effect for larger values of kL is considered. Generally, the plane wave for a larger kL

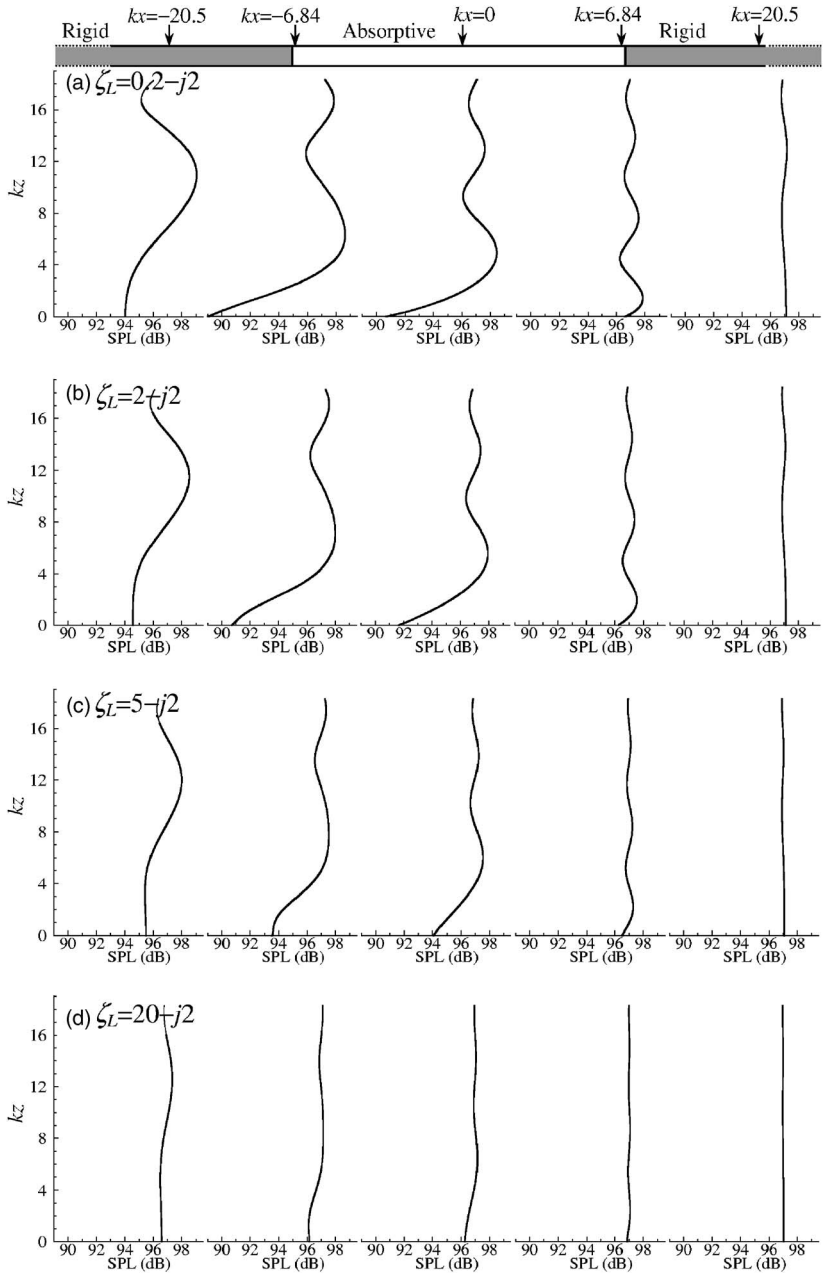


FIG. 7. Distributions of perturbed sound pressure level at $kx=-20.5$, -6.84, 0, 6.84, and 20.5 for $P_{\text{inc}}=1.0$ Pa, $kL=13.7$, and (a) $\zeta_L=0.2-j2$, (b) $\zeta_L=2-j2$, (c) $\zeta_L=5-j2$, and (d) $\zeta_L=20-j2$.

is easier to perturb because of a shorter λ compared to L, where a stronger radiation of the strip is induced. The radiated pressure is then higher and more uniform for a large kL than for a small kL [note: From Eq. (12) or (13), $\tilde{p}_{\text{rad}} \propto kh \propto kL$]. Consequently, when the perturbation of the plane wave near the strip-baffle interface is greatly reduced by a sudden change to an infinite impedance, the drop in the radiation of the strip around this discontinuity in impedance is shaper for the large kL. This phenomenon is illustrated in Fig. 9, which provides a comparison of radiated SPL at $z=0$ for $\zeta_L=2-j2$ and $20-j2$, and $kL=13.7$, 43.8, and 137 (i.e., associated with $f=250$, 800, and 2500 Hz for $L=3$ m). It is obvious that for both values of ζ_L, the sharp drop of radiated SPL in the proximity of the strip-baffle interface, increases with kL. The radiated SPL on the left baffle around the baffle-strip interface and on most of the length of the strip, also increases and converges toward the unperturbed SPL as kL increases [compare Figs. 9(a)–9(c)].

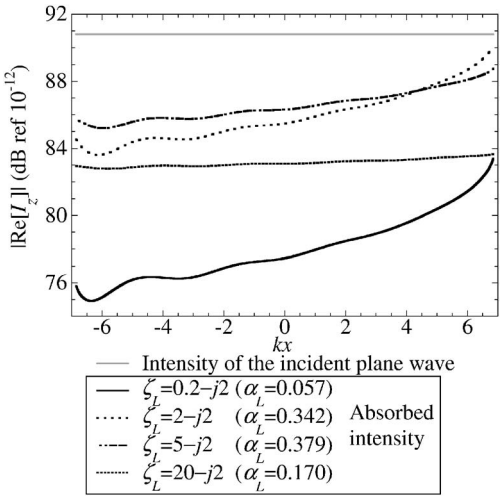


FIG. 8. Distributions of the intensity of the incident plane wave and absorbed intensity on the surface of the strip for $P_{\text{inc}}=1.0$ Pa, $kL=13.7$, and four different values of ζ_L.

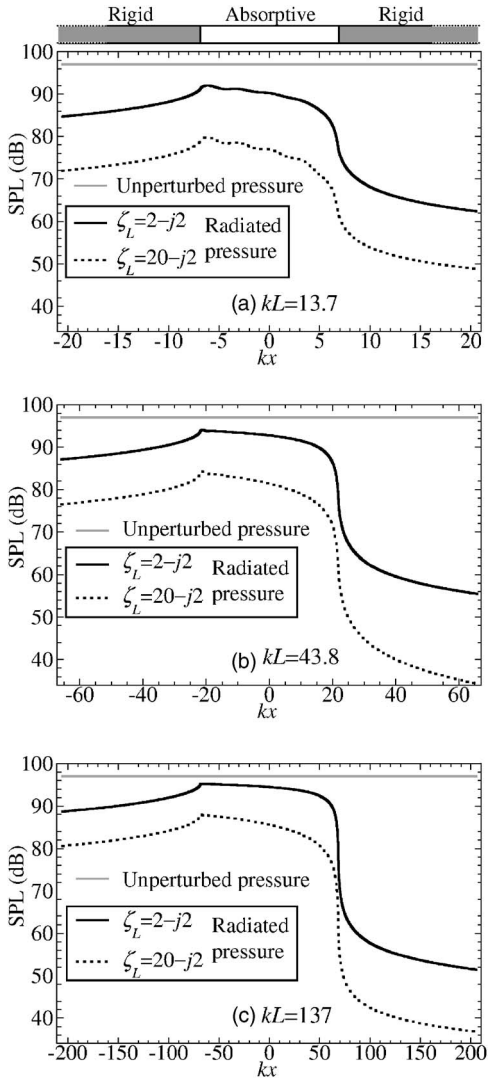


FIG. 9. Distributions of unperturbed and radiated sound pressure levels at $z=0$ for $P_{inc}=1.0$ Pa, $\zeta_L=2-j2$ and $20-j2$, and (a) $kL=13.7$, (b) $kL=43.8$, and (c) $kL=137$.

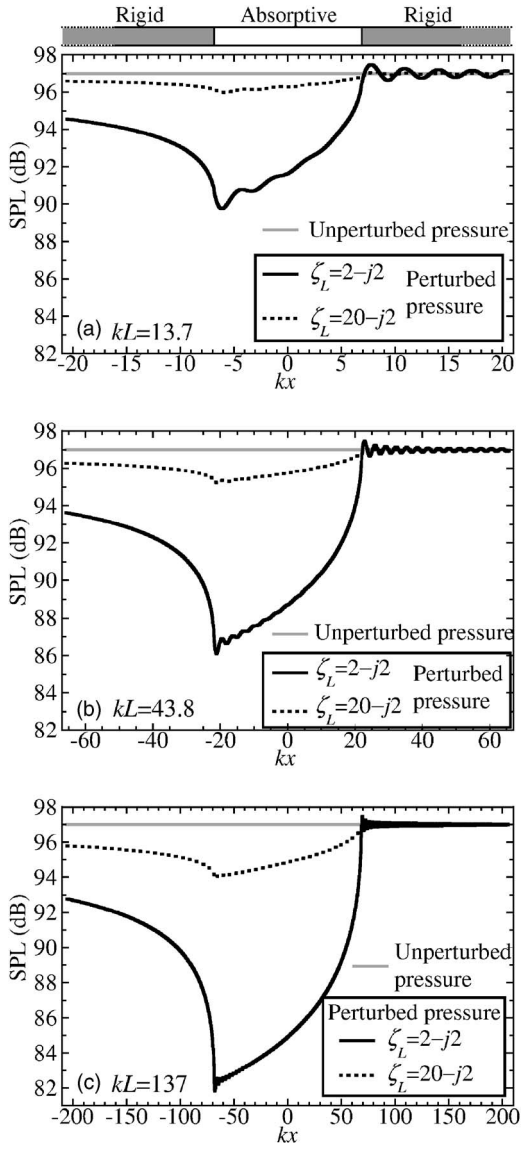


FIG. 10. Distributions of unperturbed and perturbed sound pressure levels at $z=0$ for $P_{inc}=1.0$ Pa, $\zeta_L=2-j2$ and $20-j2$, and (a) $kL=13.7$, (b) $kL=43.8$, and (c) $kL=137$.

As a result of the behavior of the radiated SPL in Fig. 9 as well as the above-mentioned out-of-phase superposition between the unperturbed and radiated waves, Figs. 10(a)–10(c) show that only the surface pressure of the strip and the right baffle around the strip-baffle interface is nearly unchanged with kL and close to the unperturbed pressure. Although the slope of increase in surface pressure over the strip rises considerably with kL, the surface pressure of the left baffle near the baffle-strip interface and of a large extent of the strip substantially decreases away from the unperturbed pressure. As the radiated SPL is much closer to the unperturbed SPL for $\zeta_L=2-j2$ than for $\zeta_L=20-j2$ (see Fig. 9), Fig. 10 indicates that the surface pressure for the low ζ_L decreases much faster with kL than that for the high ζ_L. Due to the deviation of the perturbed pressure from the unperturbed pressure, being smallest for $kL=13.7$ and largest for $kL=137$, the pressure curvature of the wave front on and near the surfaces of the baffle around both interfaces and of the strip is smallest for $kL=13.7$ and largest for $kL=137$. As an example for $\zeta_L=2-j2$, Figs. 11(a)–11(c) indicate that the distortion of the propagating plane wave is smallest for $kL=13.7$, followed by $kL=43.8$ and 137. Following the observations in Figs. 9 and 10 for $\zeta_L=2-j2$, it is also apparent in Fig. 11 that for all three values of kL, the profile of the plane wave front in the proximity of the baffle-strip interface is most significantly affected by the radiation of the strip.

Figure 12 shows the absorbed intensity of the strip for the same conditions as in Figs. 9–11. For both values of ζ_L, the intensity reduces with kL [compare Figs. 12(a)–12(c)] because the surface pressure decreases with kL (see Fig. 10). From this trend, it can be deduced that in the limit of $kL \rightarrow \infty$, the surface pressure and thus the intensity will tend to zero [i.e., corresponds to the radiated SPL that tends to the unperturbed SPL as in Figs. 9(a)–9(c), as well as the out-of-phase superposition between the radiated and unperturbed waves]. This explains the reason for the decreasing trend of α_L with kL that eventually leads to $\alpha_L=0$ when $kL \rightarrow \infty$, which is also the first feature from Fig. 3. The decrease of the surface pressure but the rise of the slope of increase in surface pressure over the strip with kL are faster for $\zeta_L=2-j2$

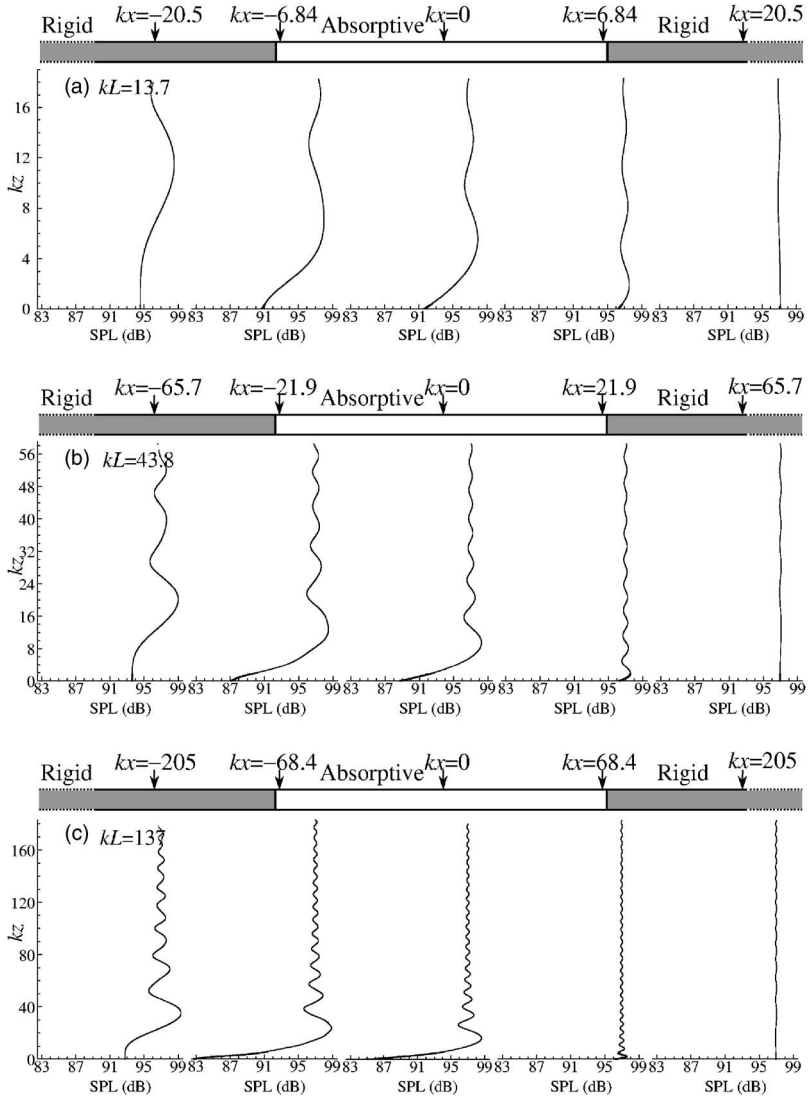


FIG. 11. Distributions of perturbed sound pressure level at $kx=-20.5$, -6.84, 0, 6.84, and 20.5 for $P_{\text{inc}}=1.0$ Pa, $\zeta_L=2-j2$, and (a) $kL=13.7$, (b) $kL=43.8$, and (c) $kL=137$.

than for $\zeta_L=20-j2$ (see Fig. 10). Hence, the product of the surface pressure and $\text{Re}[1/\zeta_L^*]$ (i.e., absorbed intensity), which is lower for $\zeta_L=20-j2$ than for $\zeta_L=2-j2$ at $kL=13.7$, becomes higher for the large ζ_L than for the small ζ_L at $kL=137$. For example, as can be seen in Fig. 12(a) for $kL=13.7$, the absorbed intensity throughout the strip and thus α_L are lower for $\zeta_L=20-j2$. When kL increases to 43.8, the intensity on the left half of the strip is higher for $\zeta_L=20-j2$ than for $\zeta_L=2-j2$, and the values of α_L for both values of ζ_L are close to each other [see Fig. 12(b)]. At $kL=137$, the intensity for $\zeta_L=20-j2$ is higher across most of the length of the strip, so α_L for $\zeta_L=20-j2$ becomes greater than that for $\zeta_L=2-j2$ [see Fig. 12(c)]. This behavior of the intensity explains the second feature from Fig. 3 that although α_L is lower for a large ζ_L than for a small ζ_L at a small kL, it can become greater for the former than for the latter at a large kL. The intensity distributions also suggest that a greater α_L for one value of ζ_L does not mean that the surface absorption across the entire strip is also larger than that for the other value of ζ_L [e.g., for $\zeta_L=20-j2$ in Fig. 12(b), α_L is lower but the intensity is higher on the baffle-strip interface side, and for $\zeta_L=2-j2$ in Fig. 12(c), α_L is lower but the intensity is higher around the strip-baffle interface].

IV. CONCLUSIONS

In this paper, characteristics of sound pressure and intensity on the surface of a finite impedance strip flush-mounted on a rigid baffle are studied for a plane wave excitation at grazing incidence. The boundary integral method is used, where the surface pressure of the strip is expressed in terms of the superposition between the unperturbed wave and the radiated wave from the strip (the total is called the perturbed wave). In order to avoid singularities in the numerical solutions of the surface pressure, the quadrature technique used to solve the boundary integral problem after the discretization of the strip into M small elements is modified by defining a mean value of Hankel function for each element. This mean value is the average of the two values of the Hankel function as when both end points (rather than the midpoint) of the element are treated as the radiation points. The solutions are well converged only when $M \gg 2kL/\pi$. A single value of M is used for all values of kL considered here, and it is based on a full convergence of the surface pressure of the strip for the largest of those values of kL.

The radiated pressure is highest at the baffle-strip interface, reduces across the strip, and is considerably low around

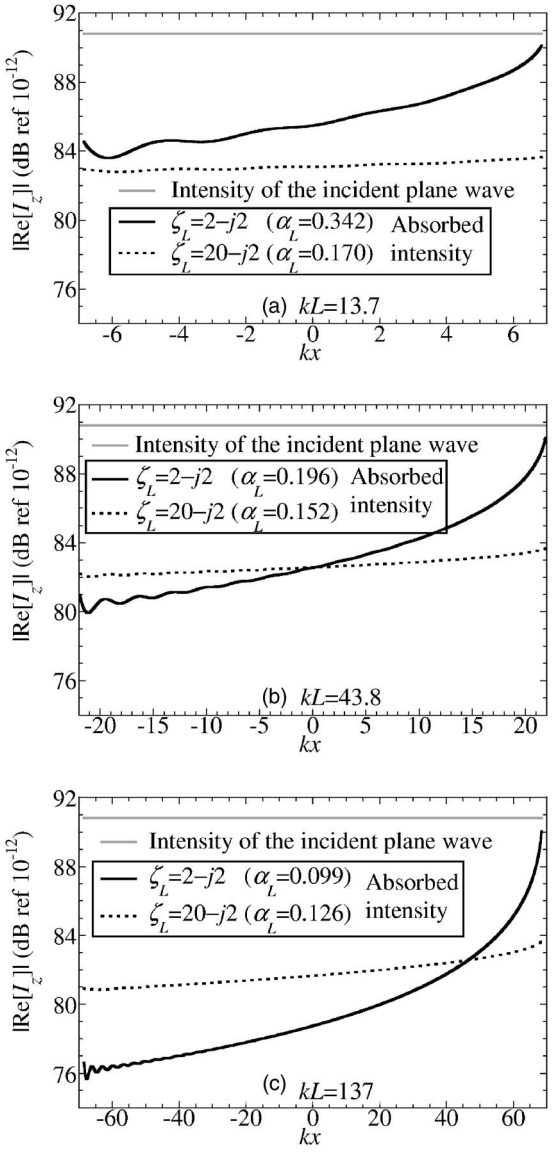


FIG. 12. Distributions of the intensity of the incident plane wave and absorbed intensity on the surface of the strip for P_{inc}=1.0 Pa, ζ_L=2−j2 and 20−j2, and (a) kL=13.7, (b) kL=43.8, and (c) kL=137.

the strip-baffle interface. Throughout most of the length of the strip, the radiated pressure decreases with ζ_L and increases with kL, but around the strip-baffle interface, it decreases with both ζ_L and kL. Since the unperturbed and radiated waves have different magnitudes and superimpose out of phase, the surface pressure of the strip is lowest at the baffle-strip interface and increases across the strip. It also increases with ζ_L and decreases with kL, except around the strip-baffle interface where it is nearly unchanged and close to the unperturbed pressure. Due to the deviation of the perturbed pressure on and above the surface of the strip from the unperturbed pressure, pressure curvatures are generated on the perturbed wave and the plane wave front becomes distorted. The distortion is large on and near the surface, but reduces with the distance from the surface. It is also largest at the baffle-strip interface, and reduces gradually toward the strip-baffle interface where it becomes very small. In other words, the surface pressure of the strip and the profile of the plane wave front near the baffle-strip interface are most significantly affected by the radiation of the strip. As the surface intensity is directly proportional to the surface pressure, the absorbed intensity of the strip is lowest around the baffle-strip interface and increases across the strip, where the slope of the increase (i.e., the nonuniformity of the intensity distribution) drops with ζ_L but rises with kL. Therefore, the intensity is nonzero, and varies with ζ_L and kL. Hence, α_L is also nonzero, and different for different values of ζ_L and kL. The decreasing trend of the surface pressure with kL implies that in the limit of $kL \to \infty$, the intensity will approach zero, which eventually leads to $\alpha_L=0$. As the intensity across a large extent of the strip decreases with kL faster for a small ζ_L than for a large ζ_L, the spatial average intensity over the entire strip (thus, α_L) for the small ζ_L, which is initially higher at a small kL, can become lower than that for the large ζ_L at a large kL.

In order to experimentally measure the distortion of a grazing plane wave front by an impedance strip as well as the distribution of surface intensity and resulting absorption coefficient of the strip, the incident wave needs to be planar and exactly graze the strip. Such excitation is very prone to errors, and the difficulty of its precise arrangement was obvious from previous measurements of impedance and/or absorption coefficient, which could only consider excitations that were far from grazing incidence for the incident wave to be planar.[9,32] Even when the incident wave was spherical, excitations could only be considered up to near grazing incidence.[7,18–20] So, in the case of a grazing plane wave as concerned here, it is necessary to discuss in detail the issues of deviation of the incident wave from planar and deviation of the excitation from grazing incidence. The effects of these deviations on the measured pressure, surface intensity, and absorption coefficient also need to be quantified, and the experiment is separately considered as a future work. Since the boundary integral method used here relies on the Green's function of the sound field, it can be further developed for predicting the size/edge effects of finite boundaries on the decay times of acoustic modes in ordinary rooms where the sound fields are not diffuse.

Given the great interest in the development of techniques for predicting impedances of absorptive materials in recent years, it is also important to establish some guidelines on the impact of edge diffraction (due to discontinuities in impedance) on the *in situ* measurement of impedance. For this purpose, one not only needs to deal with the grazing angle as concerned here, but also the normal and oblique angles as well as various materials of different absorptions and sizes (or frequencies) for these nongrazing angles. The reason is because the diffraction changes with any of these parameters. Although such investigation is beyond the scope of this paper, it could be useful as a future work.

[1]P. M. Morse and R. H. Bolt, "Sound waves in rooms," Rev. Mod. Phys. **16**, 69–150 (1944).
[2]H. Kuttruff, "The sound field in front of a wall," in *Room Acoustics*, 2nd ed. (Applied Science, London, 1979), Chap. 2, pp. 20–43.
[3]S. Thomasson, "On the absorption coefficient," Acustica **44**, 265–273 (1980).
[4]F. V. Hunt, L. L. Beranek, and D. Y. Maa, "Analysis of sound decay in

rectangular rooms," J. Acoust. Soc. Am. **11**, 80–94 (1939).
[5]B. Yegnanarayana, "Wave analysis of sound decay in rectangular rooms," J. Acoust. Soc. Am. **56**, 534–541 (1974).
[6]B. Yegnanarayana and B. S. Ramakrishna, "Diffusion of decaying sound field in a reverberation room with a highly absorbing sample," J. Acoust. Soc. Am. **56**, 706–708 (1974).
[7]C. Klein and A. Cops, "Angle dependence of the impedance of a porous layer," Acustica **44**, 258–264 (1980).
[8]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).
[9]J. F. Li and M. Hodgson, "Use of pseudo-random sequences and a single microphone to measure surface impedance at oblique incidence," J. Acoust. Soc. Am. **102**, 2200–2210 (1997).
[10]J. S. Suh and P. A. Nelson, "Measurement of transient response of rooms and comparison with geometrical acoustic models," J. Acoust. Soc. Am. **105**, 2304–2317 (1999).
[11]I. Rudnick, "The propagation of an acoustic wave along a boundary," J. Acoust. Soc. Am. **19**, 348–356 (1947).
[12]R. B. Lawhead and I. Rudnick, "Acoustic wave propagation along a constant normal impedance boundary," J. Acoust. Soc. Am. **23**, 546–549 (1951).
[13]C. F. Chien and W. W. Soroka, "Sound propagation along an impedance plane," J. Sound Vib. **43**, 9–20 (1975).
[14]K. M. Li, T. Waters-Fuller, and K. Attenborough, "Sound propagation from a point source over extended-reaction ground," J. Acoust. Soc. Am. **104**, 679–685 (1998).
[15]D. Takahashi, "Excess sound absorption due to periodically arranged absorptive materials," J. Acoust. Soc. Am. **86**, 2215–2222 (1989).
[16]A. R. Wenzel, "Propagation of waves along an impedance boundary," J. Acoust. Soc. Am. **55**, 956–963 (1974).
[17]S. Thomasson, "Reflection of waves from a point source by an impedance boundary," J. Acoust. Soc. Am. **59**, 780–785 (1976).
[18]C. Nocke, V. Mellert, T. Waters-Fuller, K. Attenborough, and K. M. Li, "Impedance deduction from broad-band, point source measurements at grazing incidence," Acust. Acta Acust. **83**, 1085–1090 (1997).
[19]C. Nocke, "Improved impedance deduction from measurements near grazing incidence," Acust. Acta Acust. **85**, 586–590 (1999).
[20]J. F. Allard, M. Henry, V. Gareton, G. Jansens, and W. Lauriks, "Impedance measurements around grazing incidence for nonlocally reacting thin porous layers," J. Acoust. Soc. Am. **113**, 1210–1215 (2003).
[21]G. Jansens, W. Lauriks, G. Vermeir, and J. F. Allard, "Free field measurements of the absorption coefficient for nonlocally reacting sound absorbing porous layers," J. Acoust. Soc. Am. **112**, 1327–1334 (2002).
[22]R. Lanoye, G. Vermeir, W. Lauriks, R. Kruse, and V. Mellert, "Measuring the free field acoustic impedance and absorption coefficient of sound absorbing materials with a combined particle velocity-pressure sensor," J. Acoust. Soc. Am. **119**, 2826–2831 (2006).
[23]F. V. Hunt, "Investigation of room acoustics by steady-state transmission measurements. I," J. Acoust. Soc. Am. **10**, 216–227 (1939).
[24]Y. W. Lam, "Issues for computer modelling of room acoustics in non-concert hall settings," Acoust. Sci. & Tech. **26**, 145–155 (2005) (available online from Acoustical Society of Japan).
[25]K. Sato and M. Koyasu, "Dependence of sound absorption coefficient upon area of acoustic materials," J. Acoust. Soc. Am. **31**, 628–629 (1959).
[26]E. D. Daniel, "On the dependence of absorption coefficients upon the area of the absorbent material," J. Acoust. Soc. Am. **35**, 571–573 (1963).
[27]T. Ten Wolde, "Measurement on the edge-effect in reverberation rooms," Acustica **18**, 207–212 (1967).
[28]A. Levitas and M. Lax, "Scattering and absorption by an acoustic strip," J. Acoust. Soc. Am. **23**, 316–322 (1951).
[29]R. K. Cook, "Absorption of sound by patches of absorbent materials," J. Acoust. Soc. Am. **29**, 324–329 (1957).
[30]T. D. Northwood, M. T. Grisaru, and M. A. Medcof, "Absorption of sound by a strip of absorptive material in a diffuse sound field," J. Acoust. Soc. Am. **31**, 595–599 (1959).
[31]A. De Bruijn, "A mathematical analysis concerning the edge effect of sound absorbing materials," Acustica **28**, 33–44 (1973).
[32]W. Lauriks, A. Cops, and Ph. Belien, "The influence of the edge effect on the statistical absorption coefficient," Acustica **70**, 155–159 (1990).
[33]M. Spivack, "Solution of the inverse scattering problem for grazing incidence upon a rough surface," J. Opt. Soc. Am. A **8**, 1892–1897 (1991).
[34]M. Spivack, "Coherent field and specular reflection at grazing incidence on a rough surface," J. Acoust. Soc. Am. **95**, 694–700 (1994).
[35]D. Habault, "Sound propagation above an inhomogeneous plane: Boundary integral equation methods," J. Sound Vib. **100**, 55–67 (1985).
[36]S. N. Chandler-Wilde and D. C. Hothersall, "Sound propagation above an inhomogeneous impedance plane," J. Sound Vib. **98**, 475–491 (1985).
[37]Y. Kawai and H. Meotoiwa, "Estimation of the area effect of sound absorbent surfaces by using a boundary integral equation," Acoust. Sci. & Tech. **26**, 123–127 (2005) (available online from Acoustical Society of Japan).
[38]P. M. Morse and K. U. Ingard, "The scattering of sound," in *Theoretical Acoustics* (McGraw-Hill, New York, 1968), Chap. 8, pp. 454–463.
[39]D. F. Mayers, "Quadrature methods for Fredholm equations of the second kind," in *Numerical Solutions of Integral Equations*, edited by L. M. Delves and J. Walsh (Oxford University Press, London, 1974), Chap. 6, pp. 64–79.
[40]M. C. Junger and D. Feit, "Elastic scatterers and waveguides," in *Sound, Structures, and Their Interaction* (Acoustical Society of America, New York, 1993), Chap. 11, pp. 342–387.

Reproducibility experiments on measuring acoustical properties of rigid-frame porous media (round-robin tests)

Kirill V. Horoshenkov[a] and Amir Khan
School of Engineering, Design and Technology, University of Bradford, Bradford, BD7 1DP, United Kingdom

François-Xavier Bécot, Luc Jaouen, Franck Sgard, Amélie Renault, and Nesrine Amirouche
Laboratoire des Sciences de l'Habitat, DGCB URA CNRS 1652, Ecole Nationale des Travaux, Publics de l'Etat, 69518 Vaulx-en-Velin Cedex, France

Francesco Pompoli, Nicola Prodi, and Paolo Bonfiglio
Department of Engineering (DIF), Universita di Ferrara, Via Saragat, 1-41100 Ferrara, Italy

Giulio Pispola and Francesco Asdrubali
Industrial Engineering Department, University of Perugia, via G. Duranti, 67, 06125 Perugia, Italy

Jörn Hübelt
Gesellschaft für Akustikforschung Dresden mbH (AFD), Stauffenbergallee 15, 01099 Dresden, Germany

Noureddine Atalla and Celse K. Amédin
Acoustics and Vibration Group, Faculty of Applied Sciences, University of Sherbrooke, Sherbrooke, J1K 2R1, Canada

Walter Lauriks and Laurens Boeckx[b]
Katholieke Universiteit Leuven, Celestijnenlaan 200 DB-3001 Leuven, Belgium

(Received 10 January 2007; revised 22 April 2007; accepted 24 April 2007)

This paper reports the results of reproducibility experiments on the interlaboratory characterization of the acoustical properties of three types of consolidated porous media: granulated porous rubber, reticulated foam, and fiberglass. The measurements are conducted in several independent laboratories in Europe and North America. The studied acoustical characteristics are the surface complex acoustic impedance at normal incidence and plane wave absorption coefficient which are determined using the standard impedance tube method. The paper provides detailed procedures related to sample preparation and installation and it discusses the dispersion in the acoustical material property observed between individual material samples and laboratories. The importance of the boundary conditions, homogeneity of the porous material structure, and stability of the adopted signal processing method are highlighted.
[DOI: 10.1121/1.2739806]

PACS number(s): 43.58.Bh, 43.50.Gf [KA] Pages: 345–353

I. INTRODUCTION

Porous materials are widely used to control the sound field in both interior and exterior spaces. Powders, wool, hair, soils, and vegetation are natural forms of porous media which affect sound field incident on the porous surface. Characterization of porous media is now a routine experiment which is carried out in many laboratories worldwide to determine the acoustical absorption performance of these materials and/or to deduce the fundamental nonacoustical data related to their porous microstructure. A standard technique to characterize these materials is the impedance tube method,[1] which typically allows one to measure the surface impedance and absorption coefficient of relatively small (e.g., 29–100 mm) samples of porous media. These data can then be used to deduce some of the nonacoustical (geometrical) parameters of porous materials. As a result, the characterization process relies heavily on the accuracy of experimental data on the acoustic surface impedance or absorption coefficient. The accuracy of this technique is affected by the quality and homogeneity of the material samples, their environmental, and operational conditions during the experiment, the quality of the setup, and the signal processing method. These conditions and measurement apparatus can vary from lab to lab and their effect on the measured values of the sound absorption coefficient is largely unknown.

There have been a number of studies into the accuracy of the standing wave tube method.[2–5] A majority of these studies are concerned with the effect of the mounting conditions on the measured values of the normal incidence acoustical impedance and sound absorption coefficient. Specifically, Pilon *et al.* suggested a practical criteria for the

[a]Electronic mail: k.horoshenkov@bradford.ac.uk
[b]Dr. Laurens Boeckx is a research mandatory of the IWT-Flanders.

assessment of the importance of the frame vibration effect in a sample with given elastic properties and of a given surface area. Song and Bolton[4] and Tsay and Yeh[6] developed advanced finite element models to study the effect of vibration of the sample installed in the impedance tube with arbitrary boundary constraints. Kino and Ueno[7] presented evidence from a series of laboratory experiments which suggests that the frame resonance effect may be overcome if the sample diameter is chosen 0.5–1.0 mm smaller than the inner diameter of the impedance tube. The latter effect will be dependent on the flow resistivity of the material sample.[2,5] These and other related studies suggest that the effects of the mounting conditions and of the circumferential air gap are expected to impact on the values of both the impedance surface and the absorption coefficient. In the case of weak edge constraints, the acoustical response of the sample tends to the response of a sample having infinite lateral dimensions.[3,6,7] The other extreme situation occurs when the sample edges are fixed.[4,5] In this case, the material is artificially "stiffened" and a shear resonance[4] may occur at the frequency which position would depend on the material elastic properties. This results in rapid and sharp variations in the impedance surface data and it corresponds to a sharp dip in the absorption data in the vicinity of the frequency of this resonance.

However, a majority of the relevant studies (e.g., Refs. 2–7) were carried out in individual laboratories and the authors are not aware of any works offering systematic experimental data on the performance of the impedance tube method between individual laboratories (i.e., interlaboratory data) for a particular set of material samples. These data should be obtained using independent sample preparation techniques, standing wave tubes of different diameters, different excitation stimuli, and signal processing methods. We also note that a majority of previous works focused on the acoustical properties of highly porous, light-weight foams and glasswool. As a result, there is a lack of data on the reproducibility of the standard impedance tube experiment on material samples made from granular-like media for which a limited porosity and a relatively low "air" permeability, are characteristic. Here the authors are referring to the static permeability, which is the ratio of the dynamic viscosity of air to the static air flow resistivity, i.e., the ability of a material to transmit continuous flow of air.

The objectives of this work are: (i) to use the standard standing wave tube method[1] to determine the dispersion of normal incidence, plane wave acoustic surface impedance and absorption coefficient data obtained for different samples of the same sheet of material in the same laboratory; and (ii) to determine the dispersion of acoustic surface impedance and absorption coefficient data for samples of the same material obtained between different laboratories. This paper is organized as follows. First, the methodology is detailed. Second, the results from individual laboratories are presented and dispersion in the results within each individual laboratory is discussed. Third, a comparison is made between the interlaboratory results. Finally, conclusions on the dispersion between the results are drawn.

TABLE I. A summary of the averaged characteristics of the investigated porous materials measured independently from the acoustic.

Material	Description	Mean porosity	Mean flow resistivity (kPa s m^{-2})	Mean density (kg/m^3)	Mean layer thickness (m)
A	Reconstituted porous rubber	0.80±0.02	141.4±44.0	242.0	0.0245
B	Reticulated foam	0.98±0.01	9.9±0.8	8.8	0.0197
C	Fibreglass	0.97±0.03	11.7±1.9	21.0	0.0290

II. METHODOLOGY

In total seven acoustic research centers were involved in this work. These are: University of Perugia (Italy), Katholieke Universiteit Leuven (Belgium), ENTPE (Lyon, France), Gesellschaft für Akustikforschung (Dresden, Germany), University of Bradford (UK), University of Ferrara (Italy), and Sherbrooke University (Canada). The paper reports the results for three porous media of different classes: reconstituted porous rubber, reticulated foam, and fiberglass, denoted, respectively, material A, B, and C. These materials were chosen to cover the range from relatively low (1.28×10^{-10} m^2) to relatively high (1.81×10^{-9} m^2) air permeability and to be representative of typical, commercially available acoustic porous materials. Table I provides a summary of some physical and geometrical characteristics of these materials. The values of parameters presented in Table I were averaged over all the range of data provided by all the partners for all the material specimens studied in this round-robin test. Table I also presents the standard deviation for the measured flow resistivity and porosity data. The criterion for the selection of the optimal sample area proposed by Pilon[5] was not adopted in this work because a detailed characterization of the elastic properties of the tested material was outside the scope of this investigation.

Each partner has been provided with a 400 mm $\times$ 400 mm sheet of the above-mentioned materials. Specimens of these materials have been cut individually by the partners using a circular cutting tool or water jet cutting machine to fit the diameter of the standing wave tube. The diameter of the standing wave tube, the measurement method, the sample preparation procedure, and the mounting method for the sample used by the partners are detailed in Table II.

TABLE II. Equipment and sample preparation procedures (HM—home-made tube; H—horizontally installed tube; V—vertically installed tube).

Partner	Tube diameter (m/tube manufacturer)	Tube length/microphone spacing (m)	Material preparation method
1	44 mm/HM/V	1/0.03	Water jet/circular tool
2	46 mm/HM/H	1.32/0.02	Rotating blade
3	38 mm/HM/H	1/0.02;0.03;0.05	Rotating blade
4	29 mm/BK4206/H	0.4225/0.02	Rotating blade
5	29 mm/BK4206/H	0.4225/0.02	Rotating blade
6	29 mm/HM/H	0.4225/0.02	Rotating blade
7	45 mm/HM/H	0.37/0.025	Water jet

TABLE III. A summary of the number of material specimens tested by individual laboratories and condition for specimen constraint (TF—tight fit; TC—tape constraint; GB—glue bonded; PF—perfect fit; R—repeated measurements with the reversed samples).

Partner	Material A	Material B	Material C	Method of support
1	6	4	6	TF/TC
2	6	6	3	TF/TC
3	1	1	1	GB
4	5	4	4	TF
5	10(R)	9(R)	10(R)	TF/TC
6	6	6	6	TF/TC
7	3	4	3	PF/TF

The measured properties were the surface impedance z_s and the absorption coefficient α of the material sample backed by a rigid wall, i.e.,

$$z_s = z_b \coth(-ik_b h), \quad \alpha = 1 - \left| \frac{z_s - \rho_0 c}{z_s + \rho_0 c} \right|^2, \tag{1}$$

respectively. Here ρ_0 and c are the equilibrium density and the sound speed in air, respectively, z_b is the characteristic impedance, k_b is the wave number within the material, and h denotes the material thickness.

Either of the following methods of sample mounting conditions were adopted (see Table III): (i) the diameter of the cut samples was 1 to 2 mm larger than the diameter of the tube in order to ensure their tight fit (TF); (ii) the diameter of the cut samples was close or slightly smaller than the diameter of the tube and the samples were wrapped in tape to prevent any leakage around the edge—tape constraint (TC); (iii) the diameter of the sample was exactly equal to that of the tube—perfect fit (PF); (iv) the diameter of the sample was exactly equal to that of the tube and the sample was glue bonded to the rigid backing—glue bonded (GB).

The partners used a range of commercially available impedance tube apparatus and impedance tubes especially designed and constructed for their laboratories. The characteristics of the impedance tube and the type of the acoustic stimuli used by individual partners are summarized in Tables II and IV. All the microphones used in these experiments were standard measurement 1/4 in. microphones. Six out of seven partners used identical or similar types of microphones provided by Brüel and Kjær. Partner 6 used specialized 1/4 in. microphones supplied by Microtech Gefell. The type of the acoustic stimulus was mainly restricted to random noise. Partners 1, 2, and 7 used maximum length[8] or phase-modulated random noise sequences.[9] The number of averages was adapted to the type of the acoustic stimulus and the signal-to-noise ratio observed during the experiment. This number varied from 8 in the case of the maximum length sequences to 100 in the case of random noise. The type of electronic hardware used for data acquisition varied from a specially dedicated commercially available Brüel and Kjær PULSE system (partners 5 and 6), general purpose A/D analyzers (partners 2 and 3), and high-quality sound cards (1 and 4). Each impedance tube was driven by a single loudspeaker which was adapted to the size and the frequency range of the impedance tube (typically in the range of 100–6000 Hz). It was assumed that the nonlinearity in the speaker response and tube vibration effect could be neglected. The sampling frequency and the sequence length used in the Fourier analysis were chosen so that to cover the desired frequency range and provide adequate frequency resolution in the transfer function spectrum as suggested in Ref. 1. The equipment was properly calibrated prior to the start of the experiments to compensate for the microphone channel mismatch using the procedure suggested in Ref. 1 for those partners who used two independent microphone channels. The effects of temperature and atmospheric pressure variations were compensated as suggested in Ref. 1. The material thickness was measured to ±0.1 mm using a set of calibrated calipers.

TABLE IV. Summary of the acoustic stimuli and the hardware used in the round-robin experiments (MLS—maximum length sequence; PRN—pseudorandom noise; RN—random noise; WN—"white" noise).

Partner	Type of acoustic stimulus	Number of averages	Electronic hardware	Microphone type
1	MLS	8	Marc-8 sound card	BK4187
2	PRN	16	NI card PXI-4461	BK4187
3	RN	30	HP-35060A analyzer	BK4135
4	WN	100	VX-Pocket PCMCIA	BK4187
5	RN	100	BK PULSE	BK4187
6	RN	100	BK PULSE	Microtech Gefell Type M360
7	PRN	6	MNS Tube-X	BK4187

III. RESULTS

A. Individual laboratory tests

The exact number of samples tested in the individual laboratories is presented in Table III. Up to ten sets of data for each of these materials were analyzed. In the case of laboratories 2 and 5 this number includes the data sets obtained when the tested samples were reversed. Figures 1–3 present the measured data for the real and imaginary parts of surface impedance together with their standard deviation obtained by each partner for materials labeled A, B, and C, respectively. The results obtained by laboratory 3 have been omitted from these figures since no statistics are available from a single set of data.

The results for the acoustic surface impedance obtained for material A indicate that the dispersion in both the real and imaginary part is considerable. Specifically, the maximum dispersion in the real part of the impedance is ±31% in the frequency range below 1000 Hz. The dispersion in the imaginary part of the impedance in this frequency range is limited and increases with the increasing frequency. This phenomenon is consistent in the results obtained in other laboratories as indicated in Fig. 1. This is a highly resistive material and the dispersion is likely to be attributed to dispersion in the values of the flow resistivity (σ) of the investigated material samples. The standard deviation in the measured values of the flow resistivity obtained for this material using the direct measurement[10] is relatively high (see Table

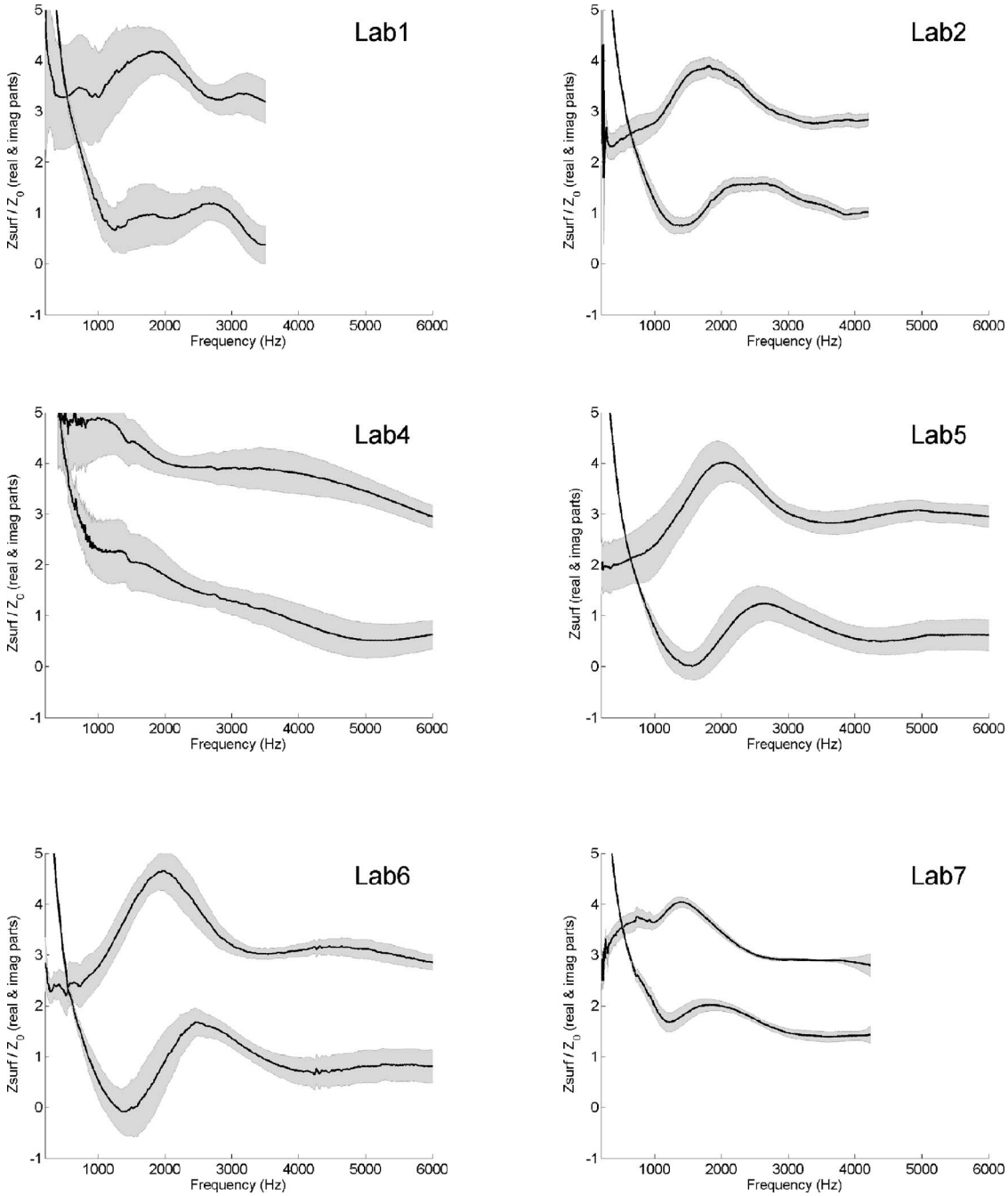


FIG. 1. Individual results of the measurements of the real and imaginary parts of the normalized surface impedance for material A for all partners.

I) and compares well with that observed in the measured data for the real part of the surface impedance ($\Delta\sigma$=31% vs Δ Re z_s=31%). This effect is expected and explained by the dominant $\sigma/3$ term in the expression [see exp. (4.4.5) in Ref. 11]

$$z_s \approx \frac{\sigma}{3} + \frac{i}{\phi\gamma kh}, \quad k \to 0, \tag{2}$$

where k is the wave number in air, γ is the ratio of specific heats, and ϕ is the material porosity.

The interlaboratory examination suggests that there is a similarity in the behavior of the mean impedance between laboratories 1, 2, and 7 but there are noticeable differences in the dispersion. These laboratories used the impedance tubes of similar diameter (44–46 mm) and similar type of acoustic excitation. The maximum dispersion in these results is observed in the case of laboratory 1 and the minimum is in the case of laboratory 7. This is likely to be attributable to the quality of the sample mounting conditions and the number of tested specimens (see Table III). The behavior of the results from laboratories 5 and 6 is comparable both in terms of the dispersion and mean values of the acoustic surface impedance. These two laboratories used identical type of the impedance tube (Brüel and Kjær 4206), signal analysis hardware (Brüel and Kjær PULSE), and the same software setup. The behavior of the impedance data obtained in laboratory 4 differs from that obtained in the other laboratories, but the level of dispersion is similar to that observed in the results

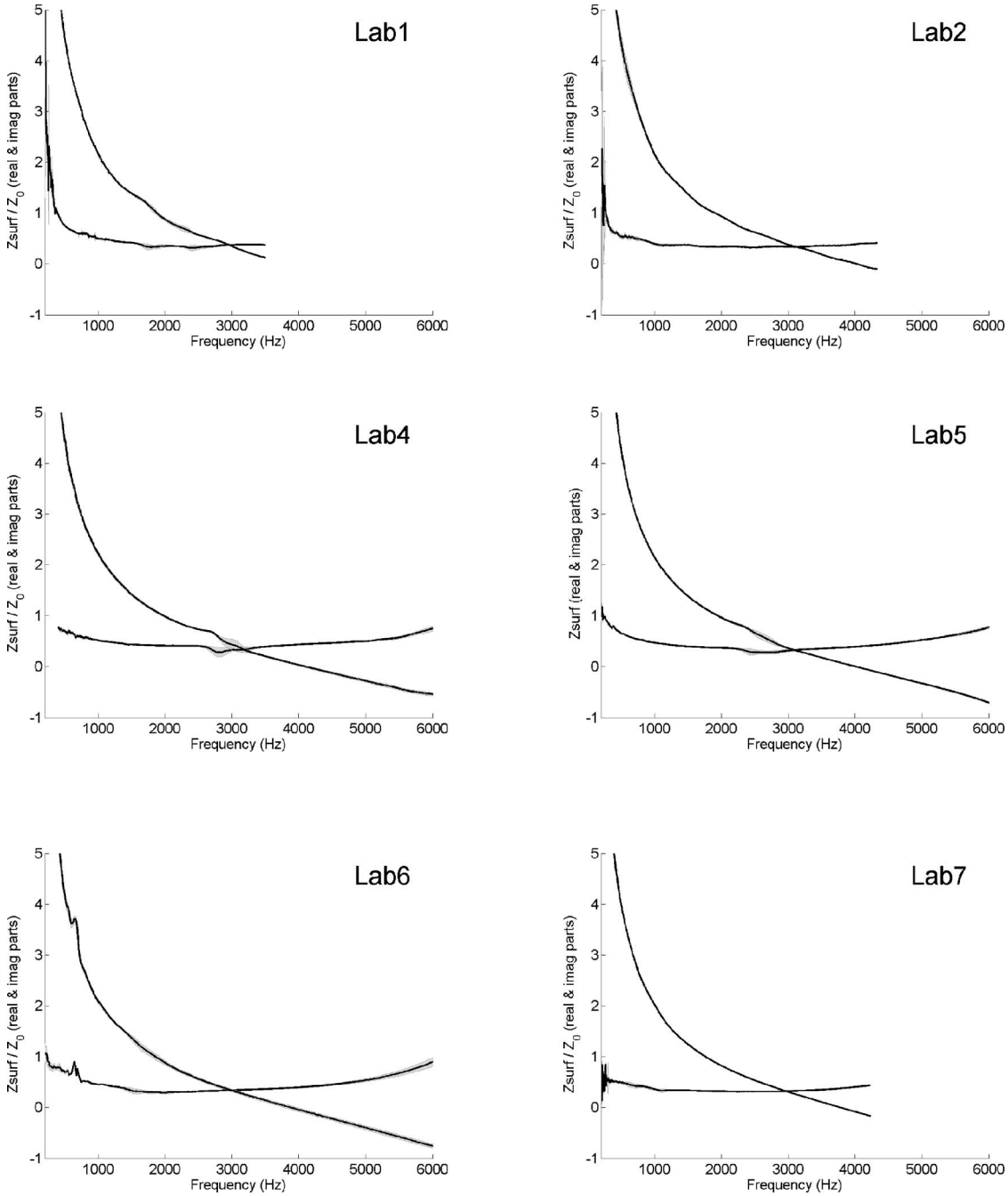


FIG. 2. Individual results of the measurements of the real and imaginary parts of the normalized surface impedance for material B for all partners.

from laboratories 5 and 6. These three laboratories use the identical impedance tubes but laboratory 4 uses a specialized sound card instead of a PULSE analyser (see Table IV).

Sample B (see Fig. 2) represents the case of a high permeability, relatively homogeneous foam for which the standard deviation in the airflow resistivity is within ±8% (excluding laboratory 7, for which this information is not available). The results show that the dispersion of the acoustic absorption spectra for this material is considerably less (<10% except around the frame resonance) than that in the case of material A, which relates to the consistent airflow resistivity values and the constant material thickness. Here the largest values of the standard deviation, reaching locally more than 30% for laboratory 4, occur near the structural resonance in the material frame which frequency depends on the mounting conditions attained during the measurement.[3–7] It is clear from the results obtained in laboratories 2 and 7 that the selected mounting conditions using the appropriate sample constraints enabled one to move the structural resonance frequency out of the measurement spectral range. Measured data from laboratories 1, 4, 5, and 6 suggest that the investigated samples were inconsistently mounted which resulted in the higher values of the standard deviation observed at frequencies of the frame resonance between 1500 and 3500 Hz (see also Fig. 4).

Sample C represents the case of a transversely isotropic fibrous material. The mean value of static permeability of this material measured in the direction normal to the fiber

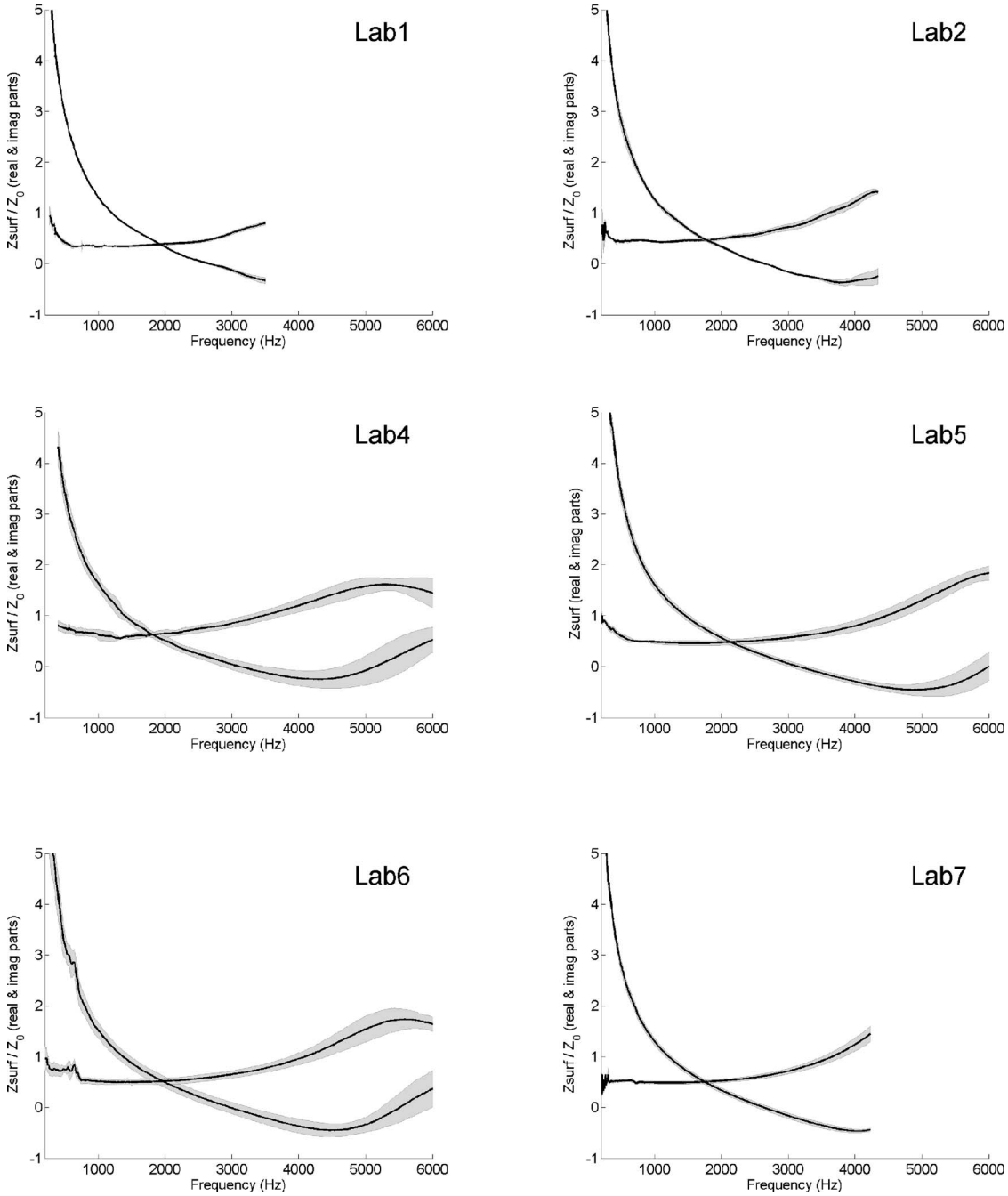


FIG. 3. Individual results of the measurements of the real and imaginary parts of the normalized surface impedance for material C for all partners.

orientation is similar to that measured in material B, but its standard deviation is higher being at ±16% (excluding laboratory 7, for which the individual value is given in the following). There is a greater thickness dispersion between individual specimens of material C. The density of this material is approximately 2.5 times greater than that of material B and the material exhibits a relatively low bulk modulus of approximately 100 kPa. The latter characteristics seem to drive the resonance frequency toward the lower spectral end so that none of the presented results show the distinctive frame resonant behavior. The dispersion in the presented data is less noticeable than that observed in the case of materials A but higher than that observed in the case of material B. The dispersion in the real part of the surface impedance lies between 10% and 20% on the frequency range from 1000 to 6000 Hz. The dispersion in the imaginary part of the acoustic surface impedance is noticeably greater across the considered frequency range. At the lower frequencies this behavior is explained by dominant effects of the material thickness as suggested by expression (2). In the higher frequency limit this behavior is governed by the oscillatory term $\coth(-ik_b h)$ in which the material thickness dispersion is likely to be dominant for low airflow resistivity glasswool. The value of the standard deviation is generally greater at the frequencies above 3000 to 4000 Hz, which is confirmed by the results from laboratories 2–6. The results from laboratory 1 do not extend to sufficiently high frequencies to demonstrate this effect. The results from laboratory 7 also show the

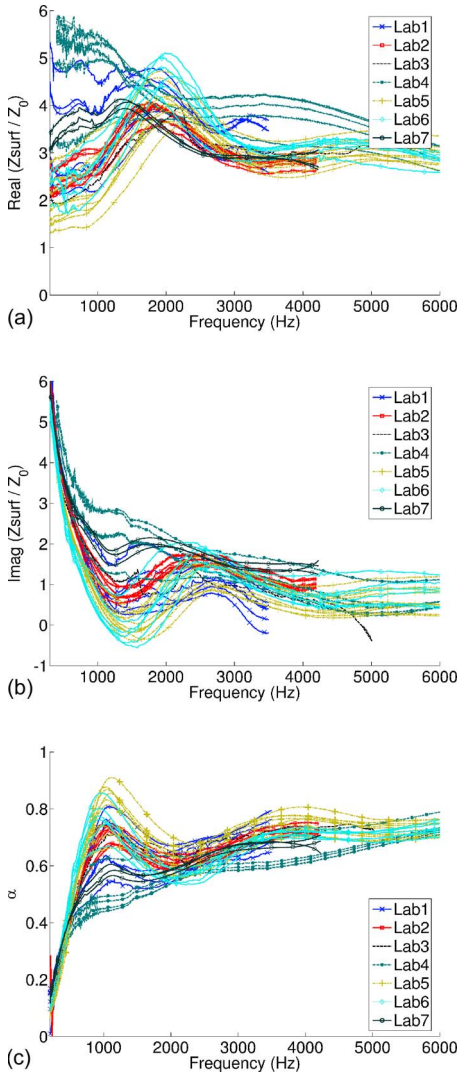


FIG. 4. (Color online) Interlaboratory measurements of the real part (top), imaginary part (middle), and the sound absorption coefficient (bottom) for material A.

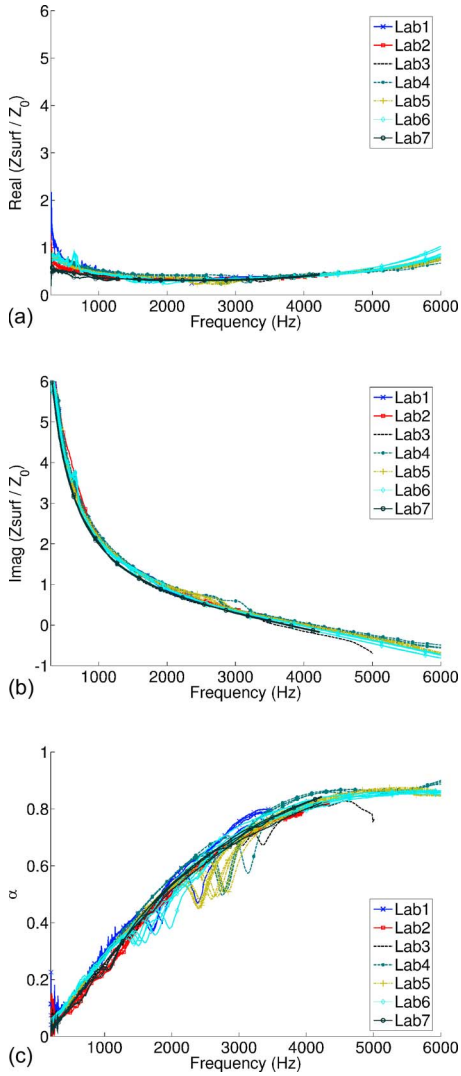


FIG. 5. (Color online) Interlaboratory measurements of the real part (top), imaginary part (middle), and the sound absorption coefficient (bottom) for material B.

increase in the dispersion in the real part of the acoustics surface impedance in the higher frequency range. This is explained by a relatively large standard deviation in the flow resistivity data for the batch of glasswool material presented to laboratory 7 ($\Delta\sigma$=27% for laboratory 7 vs $\Delta\sigma$=16% for all other laboratories) which is a consequence of the presence of the protection skin on the specimens tested by this laboratory. Note that partners 3, 4, and 7 did not remove the microperforated film covering the front surface of the fiberglass sample. The presence of the film slightly affects the measured data on the whole frequency range. This has been verified experimentally and numerically by comparing the acoustical properties for the material with and without the film. However, it is not possible to separate the effect of the screen and the sample thickness which can contribute similarly to the observed dispersion in the measured data.

B. Interlaboratory tests

Figures 4–6 present the combined results from all seven laboratories for all the specimens of materials A, B, and C. respectively. These figures (a)–(c) display data from the measurements of the real part (a) and the imaginary part (b) of the normalized acoustic surface impedance and the normal incidence, plane wave absorption coefficient (c).

The results obtained for material A show that there can be a maximum of five- to sixfold dispersion in the value of the real part of the surface impedance in the low frequency limit below 1000 Hz [Fig. 4(a)]. The agreement between the imaginary part data is poor in the medium frequency range of 1000–2000 Hz [Fig. 4(b)]. In this frequency range the imaginary part can take either negative (e.g., data from laboratory 6) or positive (e.g., data from laboratory 4) values. This dispersion is reflected in the erratic behavior of the absorption coefficient which values are summarized in Fig. 4(c). The data suggest that around these frequencies the absorption coefficient can vary within the 40%–95% range. This phenomenon is unlikely to be due to the quality of the impedance tube experiment and can rather be attributed to the dispersion in the airflow resistivity of the material specimens observed in the independent airflow resistivity tests.[10] The high flow resistivity and the acoustic penetration depth,

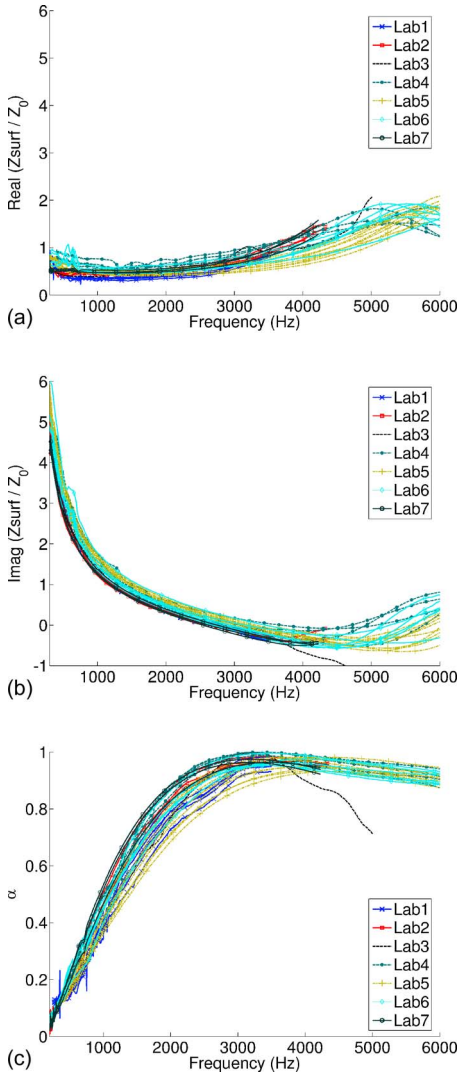


FIG. 6. (Color online) Interlaboratory measurements of the real part (top), imaginary part (middle), and the sound absorption coefficient (bottom) for material C.

$\mathrm{Re}(1/k_b)$, in this material are rather sensitive to the fluctuation of the material pore size distribution and the mounting conditions. The latter are likely to affect the thickness of the circumferential air gap and the compression rate of the investigated sample. These effects are reflected in the position and amplitude of the first interference maxima and minima in the absorption coefficient spectra which can disappear completely due to the low value of the penetration depth (e.g., data from laboratory 4).

Further examination of the obtained data suggests that the agreement between the imaginary part data for material A improves significantly in the low (e.g., below 500 Hz) and high frequency (above 5000 Hz) limits [see Fig. 4(b)] where the behavior of the absorption coefficient spectra appears much more consistent [within ±5% as illustrated in Fig. 4(c)]. This consistency in the measured acoustic absorption performance can be explained by a relatively small measured standard deviation in the material porosity [expression (2)] and by a relatively constant material layer thickness [Eq. (1)] (see also Table I) which dominate the behavior of the imaginary part.

The surface impedance and absorption coefficient spectra for material B are shown in Figs. 5(a)–5(c). There is good (approximately 10%–20%) agreement in all the results for the impedance obtained in the seven laboratories. The maximum dispersion in the real part of the impedance (within ±25%) is observed in the low and medium frequency range up to 3000 Hz [see Fig. 5(a)]. The dispersion in the imaginary part data for the impedance is limited and relatively independent of frequency [see Fig. 5(b)]. A noticeable increase of the dispersion in the absorption coefficient data can be observed around the frequency of the frame resonance in the range of 1500–3000 Hz [see Fig. 5(c)]. This suggests that the quality of the mounting conditions can control the position of the frame resonance frequency within 25% of the considered frequency range. The dispersion in the absorption coefficient due to the frame resonance can amount to more than 20% (see data from laboratories 1, 4, and 5).

Figures 6(a)–6(c) shows the results for material C measured by all seven partners. The maximum dispersion in the real part of the surface impedance can reach 100% at frequencies around 500 Hz. The dispersion in the imaginary part is comparable in this frequency range and the imaginary part can also fluctuate between the positive and the negative values [see Fig. 6(b)]. However, these fluctuations are not reflected in the dispersion of the absorption coefficient ($<10\%$) at frequencies above 4000 Hz [except data from laboratory 3 in Fig. 6(c)]. Although the behavior of the real and imaginary parts of the impedance is more consistent in the low and medium frequency range, the dispersion in the absorption coefficient is relatively large ($>20\%$). This relates to the strong dependence of the absorption coefficient of high-permeability fibrous material to the specimen thickness, which is straightforward to predict using a simple semi-empirical model (e.g., Ref. 12). There can be other complementary factors which can explain this phenomenon: (i) the quality of the material samples submitted to individual partners may not be consistent; (ii) the condition of the specimens can differ because of the damaged fibers or contamination which can occur during the transportation or cutting process; (iii) the microstructure of the sample in the standing wave tube can be affected when the sample is inserted.

IV. CONCLUSIONS

Interlaboratory standing wave tube measurements have been performed on samples of three commercial porous products which represent low- and high-permeability porous media. One standard method of testing has been used, namely the ISO 10534-2.[1] The maximum dispersion in the measured spectra for the surface impedance (five- to sixfold) and acoustic absorption coefficient (two fold) has been observed in the results between individual samples and individual laboratories in the case of low permeability, low homogeneity, broad pore size distribution, and reconstituted porous rubber (material A). The least dispersion ($<20\%$) in the data was observed in the case of high permeability porous foam (material B). This material is consistent in terms of its thickness and airflow resistivity values. Laboratories 2 and 7 demonstrated that it is possible to use the appropriate

sample constraints to move the structural resonance frequency out of the measurement spectral range to minimize the dispersion in the measured acoustical data. This effect can be predicted and avoided using the methods suggested in Refs. 4–6. However, similar mounting conditions for this material are difficult to reproduce in independent acoustic laboratories which results in the drift in the frequency of the frame resonance affecting the local value of the absorption coefficient spectrum. Intermediate level of dispersion in the measured acoustical absorption data ($>20\%$) is observed in the case of fibrous media (material C). This behavior has been attributed to the dispersion in the specimen thickness.

The current standard does not provide enough details on the procedure for sample preparation and the optimum method for sample support. In the view of the present and previous works,[2–7] it is suggested that the existing ISO10534-2[1] should be revised to define more precisely: (i) the procedure for sample preparation and minimum number of tested specimen; (ii) the minimum size of the sample as a function of the material density, bulk modulus of the material skeleton and flow resistivity; (iii) the sample mounting conditions; (iv) the type of stimuli and signal processing method; and (v) the procedure for merging material data obtained in tubes of different diameters, a procedure that has not been discussed here. The revised standard procedure should enable quantification of the intrinsic experimental errors.

This paper deals with the reproducibility of the acoustic measurements. It is proposed that a more systematic analysis of the results obtained should be carried out to investigate the dependence of the dispersion in the measured data on the geometrical and elastic properties of the porous structure and on the method of sample mounting. This will be the subject of a separate publication, which will be based on interlaboratory tests performed on the same samples of material.

ACKNOWLEDGMENTS

The authors would like to thank their industrial collaborators for the provision of material samples. The authors are grateful to the British Council and EGIDE in Paris for the support of this work. The authors would like to thank Professor Keith Attenborough for his technical comments and suggested corrections to the style and grammar.

[1]International Standards Organization, "Acoustics–Determination sound absorption coefficient and impedance in impedance tubes. 2. Transfer-unction method," ISO 10534–2 (Geneva).
[2]A. Cummings, "Impedance tube measurements on porous media—The effect of air gaps around the sample," J. Sound Vib. **151**, 63–75 (1991).
[3]T. E. Vigran, L. Kelders, W. Lauriks, P. Leclaire, and T. F. Johansen, "Prediction and measurements of the influence of boundary conditions in a standing wave tube," Acustica **83**, 419–423 (1997).
[4]B. H. Song and J. S. Bolton, "Investigation of the vibrational modes of edge-constrained fibrous samples placed in a standing wave tube," J. Acoust. Soc. Am. **113**, 1833–1849 (2003).
[5]D. Pilon, R. Panneton, and F. Sgard, "Behavioral criterion quantifying the effects of circumferential air gaps on porous materials in the standing wave tube," J. Acoust. Soc. Am. **116**(1), 344–356 (2004).
[6]H.-S. Tsay and F.-H. Yeh, "The influence of circumferential edge constraint on the acoustical properties of open-cell polyurethane foam samples," J. Acoust. Soc. Am. **19**, 2804–2814 (2006).
[7]N. Kino and T. Ueno, "Investigation of sample size effects in impedance tube measurements" Appl. Acoust., doi:10.1016/j.apacoust.2006.07.006.
[8]D. Rife and J. Vanderkooy, "Transfer-function measurement with maximum-length sequences," J. Audio Eng. Soc. **37**, 419–444 (1989).
[9]X. Olny, "Acoustic absorption of porous media with single and double porosity—Modeling and experimental validation," Ph.D. thesis, ENTPE-INSA, Lyon, 1999, p. 281 (in French).
[10]International Standards Organization, "Acoustics—Materials for acoustical applications—Determination of airflow resistance," ISO 9053, 1991.
[11]U. Ingard, *Notes on Sound Absorption Technology* (Noise Control Foundation, New York, 1994).
[12]J.-F. Allard and Y. Champoux, "New empirical equations for sound propagation in rigid frame fibrous materials," J. Acoust. Soc. Am. **91**(6), 3346–3353 (1992).

Active listening room compensation for massive multichannel sound reproduction systems using wave-domain adaptive filtering

Sascha Spors[a] and Herbert Buchner
Deutsche Telekom Laboratories, Berlin University of Technology, Ernst-Reuter-Platz 7, 10587 Berlin, Germany

Rudolf Rabenstein
Multimedia Communications and Signal Processing, University Erlangen-Nuremberg, Cauerstrasse 7, 91058 Erlangen, Germany

Wolfgang Herbordt
Center of Competence Signal Processing, Rohde & Schwarz GmbH & Co. KG, 81614 Munich, Germany

(Received 11 October 2006; revised 16 April 2007; accepted 16 April 2007)

The acoustic theory for multichannel sound reproduction systems usually assumes free-field conditions for the listening environment. However, their performance in real-world listening environments may be impaired by reflections at the walls. This impairment can be reduced by suitable compensation measures. For systems with many channels, active compensation is an option, since the compensating waves can be created by the reproduction loudspeakers. Due to the time-varying nature of room acoustics, the compensation signals have to be determined by an adaptive system. The problems associated with the successful operation of multichannel adaptive systems are addressed in this contribution. First, a method for decoupling the adaptation problem is introduced. It is based on a generalized singular value decomposition and is called *eigenspace adaptive filtering*. Unfortunately, it cannot be implemented in its pure form, since the continuous adaptation of the generalized singular value decomposition matrices to the variable room acoustics is numerically very demanding. However, a combination of this mathematical technique with the physical description of wave propagation yields a realizable multichannel adaptation method with good decoupling properties. It is called *wave domain adaptive filtering* and is discussed here in the context of wave field synthesis. © *2007 Acoustical Society of America.* [DOI: 10.1121/1.2737669]

PACS number(s): 43.60.Dh, 43.60.Pt, 43.60.Tj, 43.60.Ac [EJS] Pages: 354–369

I. INTRODUCTION

The ultimate goal of sound reproduction is to create the perfect acoustic illusion. Several generations of engineers have invented various sound reproduction systems in the past decades.[1–5] The perfect acoustic illusion has never been realized by either of them. Nevertheless, sound reproduction has improved considerably in terms of quality and spatial impression. However, there are still a number of open problems. One of these, common to most reproduction systems, is the influence of the room where the reproduction takes place, the so-called *listening room*. This contribution introduces a highly efficient approach to the active compensation of the listening room for reproduction systems with a high number of reproduction channels. The problems that arise for systems with high channel numbers pose fundamental restrictions on the realizability of active listening room compensation using standard algorithms. In order to distinguish these systems and their problems from conventional multichannel reproduction systems with low channel numbers (<10) these are referred to as massive multichannel sound reproduction systems in the context of this article. The following will introduce such massive multichannel reproduction systems, will illustrate the influence of the listening room on their performance, and will briefly review existing listening room compensation systems.

A. Massive multichannel sound reproduction systems

Spatial audio reproduction systems using loudspeakers can be roughly classified into so-called stereophony based and sound field reconstruction based. Stereophony-based systems operate typically with five or seven spatial channels.[6] The loudspeaker driving signals for each channel are obtained by carefully processing the source material (usually in a sound engineering studio) and then stored or transmitted as separate tracks, typically one for each loudspeaker, in a standardized way. Consequently, the loudspeaker setup at the user's site has to conform to the corresponding standard for optimal spatial reproduction. The techniques for processing the source material into tracks for each loudspeaker channel are borrowed from conventional two-channel stereo recording and mixing. They rely mainly on amplitude panning relative to the target position of the listener. An optimal spatial reproduction is only achieved in the vicinity of this target position, the so-called *sweet spot*.

[a]Electronic mail: sascha.spors@telekom.de

 0001-4966/2007/122(1)/354/16/$23.00

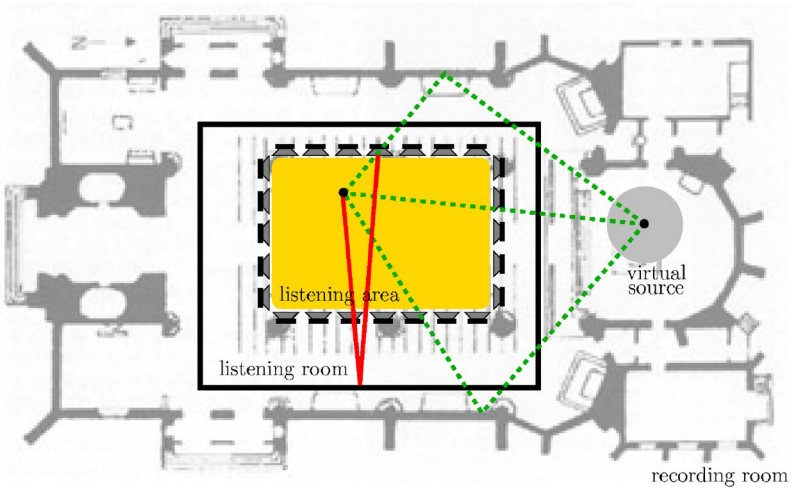


FIG. 1. (Color online) Simplified example that shows the effect of the listening room on the auralized wave field. The dashed lines from one virtual source to one exemplary listening position show the acoustic rays for the direct sound and one reflection off the sidewall of the virtual recording room. The solid line from one loudspeaker to the listening position shows a reflection of a loudspeaker wave field off the wall of the listening room.

In contrast, sound field reconstruction based systems avoid this spatial constraint. They are characterized by a high number of reproduction channels, typically some ten to several hundred. For these numbers it is neither possible nor desirable to store or transmit all the loudspeaker signals for playback. Since the loudspeaker setup often depends on the listening room architecture, each system will potentially have an individual speaker configuration. Therefore, the reproduction system cannot be built according to a certain standard for the number and placement of loudspeakers. In this situation, the driving signals for the loudspeakers have to be computed in real time by the reproduction system. The input for this processing step are the signals of the sound sources, their spatial position, and room acoustic information of the scene to be reproduced. This acoustic information is either collected during the recording of the source signals (e.g., in a concert hall) or derived from models of synthetic acoustic spaces.

Two well-known approaches for generating multichannel loudspeaker driving signals are ambisonics and wave field synthesis. Ambisonics systems represent the sound field in an enclosure by an expansion into three-dimensional basis functions. They typically use a speaker arrangement which resembles a sphere around the listening area. The contribution of the sound sources to each basis function may be recorded with special microphone arrangements.[7,8] Common realizations make use of only the contributions up to first order.[9] Wave field synthesis (WFS) is based on a physical description of wave propagation illustrated by Huygen's principle. The mathematical foundation is given by the Kirchhoff-Helmholtz integral. Certain simplifications and approximations lead to the description of a realizable multichannel sound reproduction system.[10–14]

Common to both these approaches for massive multichannel sound reproduction is that they rely on free-field acoustic wave propagation and do not consider the influence of reflections within the listening room. Since these reflections may impair the carefully designed spatial sound field, their influence should be minimized by taking appropriate countermeasures. This contribution presents an approach to addressing this issue. The following sections describe the influence of the listening room, discuss compensation methods, and give an outline of this article.

B. The influence of the listening room

The influence of the listening room on the sound scene reproduced by a multichannel reproduction system will be illustrated first in an intuitive fashion. For this purpose the simple reproduction scenario illustrated by Fig. 1 is considered in the following. The projection of an acoustic scene in a church (e.g., a singer performing in the choir) into the listening room is shown as an example. For simplicity the propagation of sound waves is illustrated by acoustic rays in Fig. 1. The dashed lines in Fig. 1 from the virtual source to one exemplary listening position show the acoustic rays for the direct sound and several reflections off the sidewalls of the church. The loudspeaker system in the listening room reproduces the direct sound and the reflections in order to create the desired spatial impression. The theory behind nearly all sound reproduction methods assumes an anechoic listening room which does not exhibit any reflections of its own. The solid line in Fig. 1 from one loudspeaker in the upper row to the listening position illustrates a possible reflection of the wave field produced by this loudspeaker off the wall of the listening room. This additional reflection caused by the listening room may impair the desired spatial impression.

The influence of the listening room on the performance of sound reproduction systems is a topic of active research.[15–26] Since the acoustic properties of the listening room and the reproduction system used may vary in a wide range, no generic conclusion can be given for the perceptual influence of the listening room. However, it is generally agreed that the reflections imposed by the listening room will have influence on the perceived properties of the reproduced scene. These influences may be, e.g., degradation of directional localization performance or sound coloration.

Summarizing, a reverberant listening room will superimpose its characteristics on the desired impression of the recorded room. Listening room compensation aims at eliminat-

ing or reducing the effect of the listening room. The following will classify and briefly introduce existing listening room compensation approaches.

C. Listening room compensation systems

Three basic classes of techniques can be identified:

1. Passive listening room compensation,
2. Consideration of the influence of the room in the rendering algorithm, and
3. Active listening room compensation.

Passive listening room compensation applies acoustic insulation materials to the listening room as countermeasure against its reflections. However, it is well known that acoustic insulation gets impractical and costly above an even rather modest level of sound absorption, especially for low frequencies. Thus, in practical setups passive room compensation alone cannot provide a sufficient suppression of listening room reflections.

The second class of approaches takes the influence of the room on the auralized field into account in calculating the loudspeaker driving signals. Almost every virtual scene will include at least some acoustic reflections, and the basic idea in this class of approaches is to modify the rendering of the scene according to the additional reflections produced by the listening room. However, no solutions or algorithms that successfully demonstrate the applicability of this idea are known to the authors at the time this article was written. A very rough sketch of some ideas can be found in Ref. 27.

The third class of approaches uses concepts from active control to perform the desired compensation. For sound reproduction, synergies with the reproduction system are used in order to control the wave field. Ideally, the desired control over the undesired reflections can be applied within the entire listening area by destructive interference. A concept shared by most active room compensation approaches, due to the causal nature of listening room reflections, is to prefilter the loudspeaker driving signals using suitable compensation filters. These filters are computed by analyzing the reproduced wave field. Most active compensation systems published in the past utilize only some few (<10) loudspeakers and analysis (microphone) positions.[28–31] As a consequence they are not able to provide sufficient control over the wave field or are they able to sufficiently analyze the reproduced wave field throughout the entire listening area. As a result, the influence of the listening room will be compensated mainly at the analyzed positions with the potential occurrence of severe artifacts at other positions.[32–35] These approaches are therefore termed *multipoint compensation* approaches.

Advanced reproduction systems like wave field synthesis and higher-order ambisonics provide an improvement in terms of control over the reproduced wave field up to a certain frequency (spatial aliasing frequency). They allow one to compensate for the reflections of the listening room throughout the entire listening area. Above the spatial aliasing frequency they may be supported easily by passive listening room compensation techniques. However, they require a large number of reproduction channels. Additionally an adequate analysis of the reproduced wave field will require a large number of analysis channels as well. Various active listening room compensation systems for such massive multichannel reproduction systems have recently been introduced.[36–46] The proposed approaches can roughly be classified into three classes depending on the algorithms used to compute the compensation filters: (1) noniterative, (2) iterative, (3) and adaptive room compensation techniques.

Most of the currently proposed listening room compensation schemes fall into the first two classes. They assume that the acoustical environment is time-invariant and has been characterized at some fixed time by impulse response measurements. The task of computing the compensation filters is typically formulated as a matrix inversion problem which is solved with noniterative or iterative methods. In general, the listening room characteristics may change over time. For instance as a result of a temperature variation in the listening room the speed of sound and hence the acoustic properties will change.[47,48] This calls for an adaptive computation of the compensation filters on the basis of an analysis of the reproduced wave field. Therefore we will focus on the derivation of an adaptive algorithm within this article.

A wide variety of problems are related to the algorithms used for the adaptation of the compensation filters. The most severe problem for a scenario with many playback and analysis channels is that the adaptation of the compensation filters is subject to fundamental problems.[49] Thus it seems that massive multichannel active room compensation may solve some of the problems of the multipoint room compensation systems like position dependence and inversion problems but only at the expense of new problems related to the multichannel adaptation.

This article will derive, discuss, and evaluate a novel efficient approach to adaptive active listening room compensation for spatial audio systems that yields an extended compensated area compared to the multipoint approaches. It also summarizes and extends our earlier work in Refs. 40–43. This approach is based on a technique that we denote as *wave domain adaptive filtering* (WDAF). In the following a short overview of this article is given.

D. Overview of this article

Section II discusses the concept of active listening room compensation in more detail and introduces the required notation. Section III presents a general concept for the adaptation of the compensation filters and lists its known problems. A framework for the solution to these problems is given in Sec. IV. It is based on a transformation into the so-called *eigenspace*, which allows a decoupling into a number of independent adaption problems. Unfortunately, this transformation turns out to be data dependent, which renders it unsuitable for practical computing. An escape is shown in Sec. V by considering the physical meaning of this transformation. Based on acoustic intuition, a new inverse filtering approach based on WDAF is presented. It is not optimal in the sense of totally decoupling the problem, but it is suitable for practical application. Finally, its application to a widely used

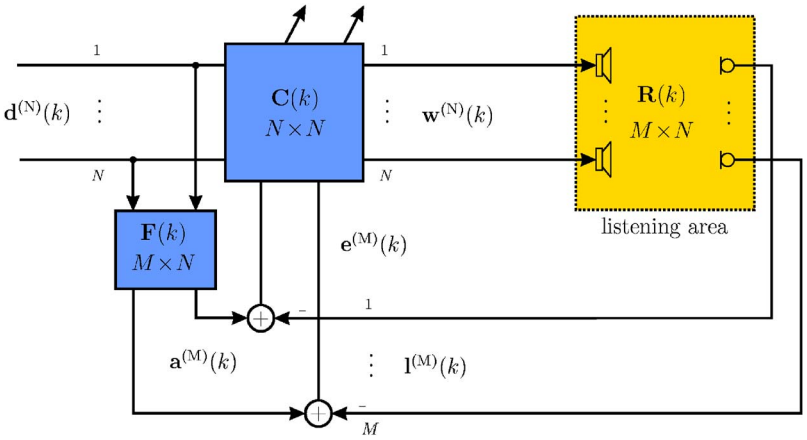


FIG. 2. (Color online) Block diagram illustrating the adaptive MIMO inverse filtering approach to room compensation.

spatial rendering technique (wave field synthesis) is shown in Sec. VI.

II. ACTIVE LISTENING ROOM COMPENSATION

The following presents the concept of active room compensation by prefiltering of the loudspeaker signals in more detail. The need for adaptation of the prefilters arises in all practical listening room situations. Moderate variations of the room temperature or the movement of a single person may change the acoustic properties of a listening room to such an extent that a previously working compensation based on fixed compensation filters is no longer valid.[48] First, the basic block diagram of an adaptive system for active room compensation system is presented and then the required notation is introduced.

A. Adaptive system for multichannel active room compensation

The block diagram of an adaptive system for multichannel active room compensation is shown in Fig. 2. The vector of input signals

$$\mathbf{d}^{(N)}(k) = [d_1(k)\ d_2(k)\ \cdots\ d_N(k)]^T \tag{1}$$

represents the N loudspeaker driving signals $d_n(k)$, $n = 1, \ldots, N$ provided by a multichannel spatial rendering system which is based on the assumption of a reflection-free listening room. Throughout this article the discrete time index is denoted by k. The matrices of impulse responses $\mathbf{C}(k)$, $\mathbf{R}(k)$, and $\mathbf{F}(k)$ represent the compensation filters, the listening room impulse responses, and the desired listening room impulses responses, respectively. In order to eliminate the effect of the listening room, free-field conditions are desired for reproduction. Hence, the matrix $\mathbf{F}(k)$ is composed from the free-field impulse responses from each loudspeaker to each microphone. This matrix is derived analytically from the solution of the wave equation for the free-field case. After the compensation filters there are N prefiltered driving signals for the loudspeakers, that are combined into a column vector $\mathbf{w}^{(N)}(k)$. A total of M microphones are used for the analysis of the resulting wave field. Their signals are combined in the vector $\mathbf{l}^{(M)}(k)$, representing spatial samples of the desired wave field on which the undesired listening room reflections are superimposed. The $N \times 1$ column vector $\mathbf{w}^{(N)} \times(k)$ and the $M \times 1$ vector $\mathbf{l}^{(M)}(k)$ are defined in an analogous way as given by Eq. (1). The output vector $\mathbf{a}^{(M)}(k)$ results from filtering the deriving signals $\mathbf{d}^{(N)}(k)$ with the idealized $M \times N$ matrix $\mathbf{F}(k)$ of free-field impulse responses.

The error $\mathbf{e}^{(M)}(k)$ between free-field propagation $\mathbf{a}^{(M)} \times(k)$ and the actual microphone signals $\mathbf{l}^{(M)}(k)$ describes the deviation of the rendered wave field from the reflection-free case. It is used to adapt the matrix of compensation filters $\mathbf{C}(k)$. After convergence of $\mathbf{C}(k)$, the response of the prefiltered reproduction system in the listening room produces the desired free-field sound field.

The matrices of impulse responses $\mathbf{R}(k)$, $\mathbf{F}(k)$, and $\mathbf{C}(k)$ describe discrete linear multiple-input/multiple-output (MIMO) systems. Assuming point source like propagation for the loudspeakers, the room $\mathbf{R}(k)$ and free-field $\mathbf{F}(k)$ matrices of impulse responses can be related to the so-called Green's function $G(\mathbf{x}|\mathbf{x}_0, \omega)$. This function characterizes the solutions of the wave equation with respect to boundary conditions. The Green's function can be regarded as the acoustic transfer function from a spatial excitation point $\mathbf{x}_0$ to a measurement point $\mathbf{x}$. Thus, its counterpart in the temporal domain $g(\mathbf{x}|\mathbf{x}_0, t) = \mathcal{F}_t^{-1}\{G(\mathbf{x}|\mathbf{x}_0, \omega)\}$ can be interpreted as the corresponding continuous-time room impulse response between the positions $\mathbf{x}$ and $\mathbf{x}_0$. This impulse response may be of infinite length, but for practical purposes it is truncated at a reasonable time (e.g., when its energy has decayed below a suitably chosen threshold) resulting in a finite impulse response (FIR). The discretized impulse response between the nth loudspeaker position and the mth analysis (microphone) position is defined as

$$r_{m,n}(k) := g(\mathbf{x}_m|\mathbf{x}_n, kT_s), \tag{2}$$

with the temporal sampling interval T_s. The $M \times N$ matrix $\mathbf{R}(k)$ captures all these impulse responses, and thus all sampling points of the Green's function $G(\mathbf{x}|\mathbf{x}_0, \omega)$ in the temporal domain. If the respective impulse responses in $\mathbf{R}(k)$ are actually measured, then these will also contain the nonideal loudspeaker and microphone characteristics (e.g., directivity and frequency response) as well as the influence of the employed hardware.[50]

Analogous definitions, as given earlier for $\mathbf{R}(k)$, apply also to the $M \times N$ matrix $\mathbf{F}(k)$ of free-field impulse responses. The matrix $\mathbf{C}(k)$ contains the sequences $c_{n,n'}(k)$ of compensation filters which have to be determined by the

adaptation algorithm introduced in Sec. III. Since the respective impulse responses are finite, they will be referred to as MIMO FIR systems in the following.

B. Frequency-domain representation of signals and systems

For the temporal frequency-domain description of the signals and systems used in this article, the discrete time Fourier transform (DTFT)[51] is used. The DTFT transform and its inverse for the spatially discrete loudspeaker driving signal $d_n(k)$ are defined as follows:

$$D_n(\omega) = \sum_{k=-\infty}^{\infty} d_n(k) e^{-j\omega k T_s}, \tag{3a}$$

$$d_n(k) = \frac{T_s}{2\pi} \int_0^{2\pi/T_s} D_n(\omega) e^{j\omega k T_s} d\omega. \tag{3b}$$

Analogous definitions apply to the other signals $w_n(k)$, $l_m(k)$, $a_m(k)$, and $e_m(k)$. The vector of loudspeaker driving signals $\mathbf{d}^{(N)}(k)$ is transformed into the temporal frequency domain by transforming each element $d_n(k)$ separately using the DTFT (3a). The resulting vector is denoted by $\underline{\mathbf{d}}^{(N)}(\omega)$. Vectors and matrices of frequency-domain signals are underlined in the following. Analogous definitions as for $\mathbf{d}^{(N)}(k)$ apply to the other signals used.

The matrix of room impulse responses $\mathbf{R}(k)$ is transformed into the temporal frequency domain in the same way as $\mathbf{d}^{(N)}(k)$. The resulting matrix in the frequency domain is referred to as *room transfer matrix* and denoted by $\underline{\mathbf{R}}(\omega)$. The matrices $\mathbf{F}(k)$ and $\mathbf{C}(k)$ are transformed analogously. The transfer matrix $\underline{\mathbf{F}}(\omega)$ is referred to as *free-field transfer matrix*.

III. ADAPTATION OF THE ROOM COMPENSATION FILTERS

Room compensation can be understood as an inverse MIMO FIR filtering problem. For only few synthesis and analysis channels numerous solutions to this problem have been developed in the past.[32–35,50] However, algorithms for massive multichannel systems still remain a challenge. The following discusses the problems that arise when standard adaptive filtering algorithms are used for the computation of the matrix of compensation filters $\mathbf{C}(k)$. To this end, it is shown how to formulate the adaptation problem such that a powerful state-of-the-art least-squares algorithm can be applied. The presentation is just detailed enough to highlight the problems associated with the application of conventional adaptation approaches to active listening room compensation. It is not intended to provide all the details necessary for an implementation. Instead the discussion of the problems will lead to an alternative approach which is presented in Secs. IV and V.

A. Least-squares error adaptation of the compensation filter

The derivation of the normal equation for the adaptive pre-equalization problem introduced in Sec. II A is briefly reviewed in the following section. A detailed discussion for acoustic MIMO systems can be found, e.g., in Refs. 43 and 50. The solution of the normal equation with the filtered-x recursive least-squares algorithm (X-RLS) is additionally shown.

The adaptation of the compensation filters $\mathbf{C}(k)$ is driven by the vector of error signals $\mathbf{e}^{(M)}(k)$. Each component $e_m(k)$ is determined by the components of the signal vector $\mathbf{d}^{(N)}(k)$ and the respective impulse response matrices as

$$e_m(k) = \sum_{n'} f_{m,n'}(k) d_{n'}(k) - \sum_{n} \sum_{n'} r_{m,n}(k) \hat{c}_{n,n'}(k) d_{n'}(k), \tag{4}$$

where the error $e_m(k)$ depends on the estimates $\hat{c}_{n,n'}(k)$ of the ideal compensation filters $c_{n,n'}(k)$. With this error signal, the following cost function is defined:

$$\xi(\hat{\mathbf{c}}, k) = \sum_{\kappa=0}^{k} \lambda^{k-\kappa} \sum_{m=1}^{M} |e_m(\kappa)|^2, \tag{5}$$

where $0 < \lambda \leq 1$ denotes an exponential weighting factor. The cost function can be interpreted as the time-averaged energy of the error $e_m(k)$ between the desired wave field $a_m(k)$ and the actual reproduced wave field $l_m(k)$ averaged over all M analysis positions. The optimal filter coefficients in the mean-squared error (MSE) sense are found by setting the gradient of the cost function $\xi(\hat{\mathbf{c}}, k)$ to zero with respect to the estimated filter coefficients $\hat{\mathbf{c}}(k)$. The normal equation is derived by expressing the error $\mathbf{e}^{(M)}(k)$ in terms of the filter coefficients, introducing the result into the cost function (5) and calculating its gradient. To express the normal equation as a linear system of equations, the estimated compensation filters are arranged in vector form,

$$\hat{\mathbf{c}}_{n,n'}(k) = [\hat{c}_{n,n'}(0), \hat{c}_{n,n'}(1), \ldots, \hat{c}_{n,n'}(N_c - 1)]^T, \tag{6a}$$

$$\hat{\mathbf{c}}_n^{(N)}(k) = [\hat{\mathbf{c}}_{n,1}^T(k) \hat{\mathbf{c}}_{n,2}^T(k) \cdots \hat{\mathbf{c}}_{n,N}^T(k)]^T, \tag{6b}$$

$$\hat{\mathbf{c}}(k) = [\hat{\mathbf{c}}_1^{(N)}(k)^T \hat{\mathbf{c}}_2^{(N)}(k)^T \cdots \hat{\mathbf{c}}_N^{(N)}(k)^T]^T, \tag{6c}$$

where N_c denotes the number of filter coefficients. Typically this number has to be chosen higher than the number of coefficients N_r of the room impulse responses $\mathbf{r}_{m,n}(k)$ since we are dealing with an inverse identification problem.[52] The resulting normal equation of the inverse filtering problem is then given as

$$\hat{\mathbf{\Phi}}_{dd}(k) \hat{\mathbf{c}}(k) = \hat{\mathbf{\Phi}}_{da}(k). \tag{7}$$

The $N^2 N_c \times N_c N^2$ matrix $\hat{\mathbf{\Phi}}_{dd}(k)$ is the time and analysis position-averaged autocorrelation matrix of the filtered driving signals

$$\hat{\mathbf{\Phi}}_{dd}(k) = \sum_{\kappa=0}^{k} \lambda^{k-\kappa} \mathbf{D}_R(\kappa) \mathbf{D}_R^T(\kappa) = \lambda \hat{\mathbf{\Phi}}_{dd}(k-1) + \mathbf{D}_R(k) \mathbf{D}_R^T(k), \tag{8}$$

where $\mathbf{D}_R(\kappa)$ denotes the matrix of filtered driving signals. The matrix of filtered driving signals is given as follows:

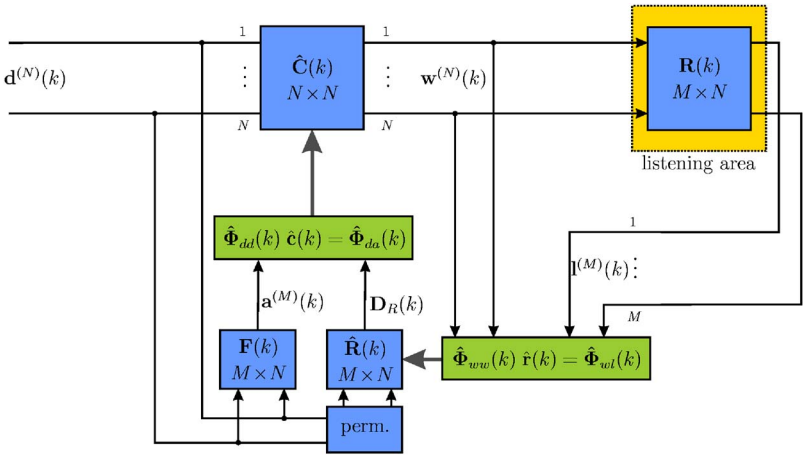


FIG. 3. (Color online) Block diagram illustrating the application of the X-RLS algorithm to active listening room compensation.

$$d_{m,n,n'}(k) = d_{n'}(k) r_{m,n}(k), \tag{9a}$$

$$\mathbf{d}_{m,n,n'}(k) = [d_{m,n,n'}(k) d_{m,n,n'}(k-1) \cdots d_{m,n,n'}(k - N_c + 1)]^T, \tag{9b}$$

$$\mathbf{d}_{m,n}^{(N)}(k) = [\mathbf{d}_{m,n,1}^T(k) \mathbf{d}_{m,n,2}^T(k) \cdots \mathbf{d}_{m,n,N}^T(k)]^T, \tag{9c}$$

$$\mathbf{d}_{M}^{(N,N)}(k) = [\mathbf{d}_{m,1}^{(N)}(k)^T \mathbf{d}_{m,2}^{(N)}(k)^T \cdots \mathbf{d}_{m,N}^{(N)}(k)^T]^T, \tag{9d}$$

$$\mathbf{D}_R(k) = [\mathbf{d}_1^{(N,N)}(k) \mathbf{d}_2^{(N,N)}(k) \cdots \mathbf{d}_M^{(N,N)}(k)]. \tag{9e}$$

Note that $\mathbf{D}_R(k)$ is composed from all N^2 combinations resulting from filtering the driving signals $d_{n'}(k)$ with all inputs of the MIMO room responses $r_{m,n}(k)$ for all possible combination of n and n'. This results from rearranging Eq. (4) in order to isolate the filter coefficients $\hat{\mathbf{c}}(k)$.

The $N^2 N_c \times 1$ vector $\hat{\mathbf{\Phi}}_{da}(k)$ can be interpreted as the time and analysis position-averaged cross-correlation vector between the filtered driving signals and the desired signals which is defined as

$$\hat{\mathbf{\Phi}}_{da}(k) = \sum_{\kappa=0}^{k} \lambda^{k-\kappa} \mathbf{D}_R(\kappa) \mathbf{a}^{(M)}(\kappa) = \lambda \hat{\mathbf{\Phi}}_{da}(k-1) + \mathbf{D}_R(k) \mathbf{a}^{(M)}(k). \tag{10}$$

The optimal pre-equalization filter with respect to the cost function (5) is given by solving the normal equation (7) with respect to the filter coefficients $\hat{\mathbf{c}}(k)$. The normal equation is typically not solved in a direct manner but by recursive updates of the filter coefficients. A recursive update equation for the filter coefficients can be derived from the normal equation (7), the recursive definitions of the correlation matrices given by Eqs. (8) and (10), and the definition of the error signal $\mathbf{e}^{(M)}(k)$ as follows:

$$\hat{\mathbf{c}}(k) = \hat{\mathbf{c}}(k-1) - \hat{\mathbf{\Phi}}_{dd}^{-1}(k) \mathbf{D}_R(k) \mathbf{e}^{(M)}(k). \tag{11}$$

Equation (11) together with the recursive definition of the correlation matrix $\hat{\mathbf{\Phi}}_{dd}(k)$ given by Eq. (8) constitutes the basis of the X-RLS algorithm. The inverse $\hat{\mathbf{\Phi}}_{dd}^{-1}(k)$ of the autocorrelation matrix is typically computed in a recursive fashion by applying the matrix inversion lemma.[53]

The X-RLS algorithm deviates from the standard RLS algorithm by using a filtered version of the driving signal for adaptation. The calculation of the filtered driving signals requires knowledge of the room response $\mathbf{R}(k)$, which is in general not known *a priori* and will be time variant. Hence, the room characteristics have to be identified additionally using a multichannel RLS algorithm. Its normal equation can be derived in the same manner as shown earlier for the compensation filters. It is given by

$$\hat{\mathbf{\Phi}}_{ww}(k) \hat{\mathbf{r}}(k) = \hat{\mathbf{\Phi}}_{wl}(k), \tag{12}$$

where $\hat{\mathbf{\Phi}}_{ww}(k)$ denotes the autocorrelation matrix of the filtered loudspeaker driving signals $\mathbf{w}^{(N)}(k)$, $\hat{\mathbf{\Phi}}_{wl}(k)$ the cross-correlation matrix between the filtered loudspeaker driving signals $\mathbf{w}^{(N)}(k)$ and the analysis signals $\mathbf{l}^{(M)}(k)$, and $\hat{\mathbf{r}}(k)$ the estimated coefficients of the room transfer matrix. Figure 3 illustrates the X-RLS algorithm applied to the active listening room compensation scenario.

B. Fundamental problems of adaptive inverse filtering

Three fundamental problems of massive multichannel adaptive pre-equalization can be identified from the normal equation (7) and the definition (8) of the autocorrelation matrix $\hat{\mathbf{\Phi}}_{dd}(k)$. These are:

1. Nonuniqueness of the solution,
2. Ill-conditioning of the autocorrelation matrix $\hat{\mathbf{\Phi}}_{dd}(k)$, and
3. Computational complexity for massive MIMO systems.

The first problem is related to minimization of the cost function $\xi(\hat{\mathbf{c}}, k)$. Minimization of the cost function $\xi(\hat{\mathbf{c}}, k)$ may not provide the optimal solution in terms of identifying the inverse system to the room transfer matrix. Depending on the driving signals $\mathbf{d}^{(N)}(k)$ there may be multiple solutions for $\hat{\mathbf{c}}(k)$ that minimize $\xi(\hat{\mathbf{c}}, k)$.[54] This problem is often referred to as *nonuniqueness problem*.

The second and the third fundamental problems are related to the solution of the normal equation (7). The normal equation has to be solved with respect to the coefficients $\hat{\mathbf{c}}(k)$ of the room compensation filter. However, the autocorrelation matrix $\hat{\mathbf{\Phi}}_{dd}(k)$ is typically ill-conditioned for the considered multichannel reproduction scenarios.[49] The filtered driv-

TABLE I. Complexity of adaptive listening room compensation using a multichannel filtered-x RLS (X-RLS) algorithm in comparison to the proposed approach of eigenspace adaptive filtering (EAF) in conjunction with single-channel X-RLS algorithms. For the complexity of the transformations $\underline{\mathbf{X}}^H(\omega)$ and $\underline{\mathbf{Y}}(\omega)$ only their application has been considered but not their computation by the GSVD.

		X-RLS	EAF
Adaption of $\hat{\underline{\mathbf{C}}}(\omega)$	Dimension of $\hat{\mathbf{\Phi}}_{dd}(k)$	$N^2N_c \times N_cN^2$	$N_c \times N_c$
	Complexity	$O(N^4N_c^2)$	$M \cdot O(N_c^2)$
Identification of $\hat{\underline{\mathbf{R}}}(\omega)$	Dimension of $\hat{\mathbf{\Phi}}_{ww}(k)$	$NN_c \times N_cN$	$N_c \times N_c$
	Complexity	$O(N^2N_c^2)$	$M \cdot O(N_c^2)$
Transformations $\underline{\mathbf{X}}^H(\omega)$ and $\underline{\mathbf{Y}}(\omega)$	Complexity		$O(NM)+O(M^2)$

ing signals $\mathbf{D}_R(k)$ will contain cross-channel (spatial) correlations due to the deterministic nature of most auralization algorithms. Also temporal correlations may be present for typical virtual source signals. Besides the ill-conditioning also the dimensionality of the autocorrelation matrix $\hat{\mathbf{\Phi}}_{dd}(k)$ poses problems. Massive multichannel systems exhibit a high number of reproduction channels and additionally the length N_c of the inverse filter has to be chosen quite long for a suitable suppression of reflections.[52] As a consequence, the derivation of the compensation filters will get computationally very demanding. Table I illustrates the complexity of the X-RLS algorithm. The complexity is given on the basis of a straightforward solution of the normal equation without any optimizations. $O(N)$ denotes "in the order of N" as measure of complexity. A more detailed complexity analysis of the X-RLS algorithm can be found, e.g., in Ref. 55.

The same problems as discussed earlier apply to the identification of the room transfer matrix using a multichannel RLS algorithm.[54]

IV. EIGENSPACE ADAPTIVE FILTERING

In this section we will derive a generic framework for pre-equalization which explicitly solves the problems of complexity and cross-channel correlations by utilizing signal and system transformations. It will be shown additionally that the other problems mentioned earlier are highly alleviated by the proposed approach. The basic idea is to perform a joint decoupling of the MIMO systems $\underline{\mathbf{R}}(\omega)$ and $\underline{\mathbf{F}}(\omega)$ in the temporal frequency domain. This decoupling yields a decoupling of the MIMO adaptation problem and the autocorrelation matrix $\hat{\mathbf{\Phi}}_{dd}(k)$ as will also be shown.

A. Nonadaptive computation of room compensation filters

In order to gain more insight into the solution of the pre-equalization problem the nonadaptive case will be regarded first. For this purpose a frequency-domain description of the pre-equalization problem depicted in Fig. 2 is used.

The error $\underline{\mathbf{e}}^{(M)}(\omega)$ between the desired $\underline{\mathbf{a}}^{(M)}(\omega)$ and the actual $\underline{\mathbf{l}}^{(M)}(\omega)$ signal at the M analysis positions in the frequency domain can be derived by a DTFT of Eq. (4) as

$$\underline{\mathbf{e}}^{(M)}(\omega) = \underline{\mathbf{a}}^{(M)}(\omega) - \mathbf{I}^{(M)}(\omega) = \underline{\mathbf{F}}(\omega)\underline{\mathbf{d}}^{(N)}(\omega) - \underline{\mathbf{R}}(\omega)\underline{\mathbf{C}}(\omega)\underline{\mathbf{d}}^{(N)}(\omega). \tag{13}$$

Optimal pre-equalization is obtained by minimizing this error: $\underline{\mathbf{e}}^{(M)}(\omega) \rightarrow \mathbf{0}$. The least-squares solution to Eq. (13) in this sense, with respect to the compensation filters, is given by[53,56]

$$\underline{\mathbf{C}}(\omega) = \underline{\mathbf{R}}^{+}(\omega)\underline{\mathbf{F}}(\omega), \tag{14}$$

where $\underline{\mathbf{R}}^{+}(\omega)$ denotes the Moore-Penrose pseudoinverse of $\underline{\mathbf{R}}(\omega)$.

B. Generalized singular value decomposition

It will be assumed in the following that $\underline{\mathbf{R}}(\omega)$ and $\underline{\mathbf{F}}(\omega)$ have the dimensions $M \times N$ with $N \geqslant M$. Hence, a reproduction system with equal or less analysis positions than loudspeakers is assumed. This restriction is meaningful with respect to the solution of the pre-equalization problem, since in the general case an exact reproduction at M positions can only be gained for such scenarios.[57] However, the derived results can be generalized straightforwardly to arbitrary channel numbers M and N.

The singular value decomposition (SVD) states that any matrix can be decomposed into two unitary matrices and a diagonal matrix.[53,56] The concept of the SVD can be generalized to the diagonalization of a pair of matrices. This decomposition is known as generalized singular value decomposition (GSVD).[56] The GSVD of the matrices $\underline{\mathbf{R}}(\omega)$ and $\underline{\mathbf{F}}(\omega)$ is given as follows:

$$\underline{\mathbf{R}}(\omega) = \underline{\mathbf{X}}(\omega)\underline{\tilde{\mathbf{R}}}(\omega)\underline{\mathbf{Y}}^H(\omega), \tag{15a}$$

$$\underline{\mathbf{F}}(\omega) = \underline{\mathbf{X}}(\omega)\underline{\tilde{\mathbf{F}}}(\omega)\underline{\mathbf{U}}^H(\omega). \tag{15b}$$

The matrices $\underline{\mathbf{X}}(\omega)$, $\underline{\mathbf{Y}}(\omega)$, and $\underline{\mathbf{U}}(\omega)$ are unitary matrices with the dimensions $M \times M$, $N \times M$, and $N \times M$, respectively. The matrix $\underline{\mathbf{X}}(\omega)$ is the generalized singular matrix of $\underline{\mathbf{R}}(\omega)$ and $\underline{\mathbf{F}}(\omega)$, the matrices $\underline{\mathbf{Y}}(\omega)$ and $\underline{\mathbf{U}}(\omega)$ the respective right singular matrices of $\underline{\mathbf{R}}(\omega)$ and $\underline{\mathbf{F}}(\omega)$. The matrices $\underline{\tilde{\mathbf{R}}}(\omega)$ and $\underline{\tilde{\mathbf{F}}}(\omega)$ are diagonal matrices constructed from the singular values of $\underline{\mathbf{R}}(\omega)$ and $\underline{\mathbf{F}}(\omega)$. The diagonal matrix $\underline{\tilde{\mathbf{R}}}(\omega)$ is defined as

$$\underline{\tilde{\mathbf{R}}}(\omega) = \mathrm{diag}\{[\tilde{R}_1(\omega), \tilde{R}_2(\omega), \ldots, \tilde{R}_M(\omega)]\}, \tag{16}$$

where $\tilde{R}_1(\omega) \geqslant \tilde{R}_2(\omega) \geqslant \cdots \geqslant \tilde{R}_B(\omega) > 0$ denote the B nonzero singular values $\tilde{R}_m(\omega)$ of $\underline{\mathbf{R}}(\omega)$. Their total number B is given by the rank of the matrix $\underline{\mathbf{R}}(\omega)$ with $1 \leqslant B \leqslant M$. For $B < M$ the remaining singular values $\tilde{R}_{B+1}(\omega), \tilde{R}_{B+2}(\omega), \ldots, \tilde{R}_M(\omega)$ are zero. Similar definitions as given earlier for $\underline{\tilde{\mathbf{R}}}(\omega)$ apply to the matrix $\underline{\tilde{\mathbf{F}}}(\omega)$.

The relation given by Eq. (15a) can be inverted by exploiting the unitary property of the joint and right singular matrices. This results in

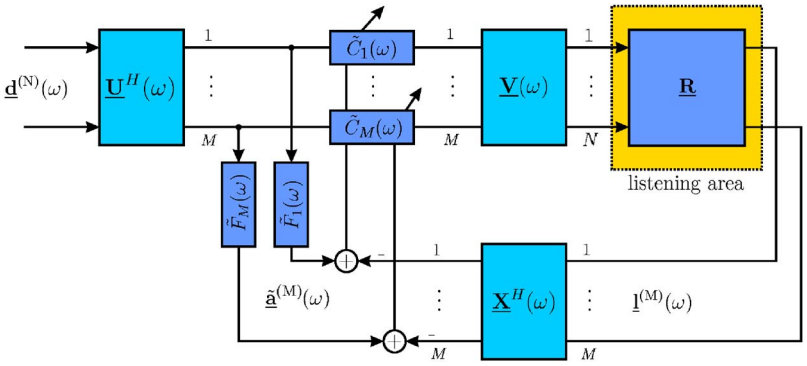


FIG. 4. (Color online) Block diagram illustrating the eigenspace adaptive inverse filtering approach to room compensation.

$$\tilde{\underline{\mathbf{R}}}(\omega) = \underline{\mathbf{X}}^H(\omega)\underline{\mathbf{R}}(\omega)\underline{\mathbf{V}}(\omega). \tag{17}$$

Hence each matrix $\underline{\mathbf{R}}(\omega)$ can be transformed into a diagonal matrix $\tilde{\underline{\mathbf{R}}}(\omega)$ using the joint and right singular matrix $\underline{\mathbf{X}}(\omega)$ and $\underline{\mathbf{V}}(\omega)$. A similar relation as given by Eq. (17) can be derived straightforwardly for $\tilde{\underline{\mathbf{F}}}(\omega)$. The GSVD transforms the matrices $\mathbf{R}(\omega)$ and $\mathbf{F}(\omega)$ into their joint eigenspace using the singular matrices $\underline{\mathbf{X}}(\omega)$, $\underline{\mathbf{V}}(\omega)$, and $\underline{\mathbf{U}}(\omega)$. In general, these singular matrices depend on the matrices $\underline{\mathbf{R}}(\omega)$ and $\underline{\mathbf{F}}(\omega)$. The GSVD is a *data-dependent transformation*. The transformation of $\underline{\mathbf{R}}(\omega)$ into its diagonal representation $\tilde{\underline{\mathbf{R}}}(\omega)$, as given by Eq. (17), can be interpreted as pre- and postfiltering the room transfer matrix $\underline{\mathbf{R}}(\omega)$ by the MIMO systems $\underline{\mathbf{V}}(\omega)$ and $\underline{\mathbf{X}}^H(\omega)$.

The SVD can be used to define the pseudoinverse $\underline{\mathbf{R}}^+(\omega)$ of the matrix $\underline{\mathbf{R}}(\omega)$,[53]

$$\underline{\mathbf{R}}^+(\omega) = \underline{\mathbf{V}}(\omega)\tilde{\underline{\mathbf{R}}}^{-1}(\omega)\underline{\mathbf{X}}^H(\omega). \tag{18}$$

Equation (15b) and Eq. (18) can be combined to derive the following result:

$$\underline{\mathbf{R}}^+(\omega)\underline{\mathbf{F}}(\omega) = \underline{\mathbf{V}}(\omega)\tilde{\underline{\mathbf{R}}}^{-1}(\omega)\tilde{\underline{\mathbf{F}}}(\omega)\mathbf{U}^H(\omega), \tag{19}$$

where it is assumed that $\underline{\mathbf{R}}(\omega)$ and $\underline{\mathbf{F}}(\omega)$ both have full rank. Equation (19) will be utilized in the following to derive a decoupling of the room compensation filters.

C. Decoupling of the MIMO adaptation problem

The decompositions of the transfer matrices $\underline{\mathbf{F}}(\omega)$ and $\underline{\mathbf{R}}(\omega)$ are given by Eq. (15). It remains to choose a suitable decomposition of the compensation filter matrix $\underline{\mathbf{C}}(\omega)$. The nonadaptive solution for the pre-equalization filter is given by Eq. (14) in terms of the pseudoinverse $\underline{\mathbf{R}}^+(\omega)$ of the room transfer matrix. Hence, an eigenspace expansion of $\underline{\mathbf{C}}(\omega)$ is given by Eq. (19). However, the system transfer matrix $\underline{\mathbf{R}}(\omega)$ is not known in general and has to be identified additionally. An expansion of the pre-equalization filter can be given by utilizing Eq. (19) but with unknown expansion coefficients $\tilde{\underline{\mathbf{C}}}(\omega)$,

$$\underline{\mathbf{C}}(\omega) = \underline{\mathbf{V}}(\omega)\tilde{\underline{\mathbf{C}}}(\omega)\underline{\mathbf{U}}^H(\omega), \tag{20}$$

where $\tilde{\underline{\mathbf{C}}}(\omega)$ denotes a diagonal matrix, where some diagonal elements may be zero. Using Eq. (20) together with Eq. (15a) yields the transformed signal $\tilde{\underline{\mathbf{l}}}^{(M)}(\omega)$ at the analysis points

$$\tilde{\underline{\mathbf{l}}}^{(M)}(\omega) = \tilde{\underline{\mathbf{R}}}(\omega)\tilde{\underline{\mathbf{C}}}(\omega)\tilde{\underline{\mathbf{d}}}^{(M)}(\omega), \tag{21}$$

where $\tilde{\underline{\mathbf{l}}}^{(M)}(\omega) = \underline{\mathbf{X}}^H(\omega)\underline{\mathbf{l}}^{(M)}(\omega)$ and $\tilde{\underline{\mathbf{d}}}^{(M)}(\omega) = \underline{\mathbf{U}}^H(\omega)\underline{\mathbf{d}}^{(N)}(\omega)$. Decomposition of the desired system response according to Eq. (15b) yields the desired signal in the transformed domain as

$$\tilde{\underline{\mathbf{a}}}^{(M)}(\omega) = \tilde{\underline{\mathbf{F}}}(\omega)\tilde{\underline{\mathbf{d}}}^{(M)}(\omega), \tag{22}$$

where $\tilde{\underline{\mathbf{a}}}^{(M)}(\omega) = \underline{\mathbf{X}}^H(\omega)\underline{\mathbf{a}}^{(M)}(\omega)$. Equation (21) together with Eq. (22) allows one to express the error $\underline{\mathbf{e}}^{(M)}(\omega)$ in the transformed domain

$$\tilde{\underline{\mathbf{e}}}^{(M)}(\omega) = \tilde{\underline{\mathbf{a}}}^{(M)}(\omega) - \tilde{\underline{\mathbf{l}}}^{(M)}(\omega) = \tilde{\underline{\mathbf{F}}}(\omega)\tilde{\underline{\mathbf{d}}}^{(M)}(\omega) - \tilde{\underline{\mathbf{R}}}(\omega)\tilde{\underline{\mathbf{C}}}(\omega)\tilde{\underline{\mathbf{d}}}^{(M)}(\omega), \tag{23}$$

where $\tilde{\underline{\mathbf{e}}}^{(M)}(\omega)$ denotes the error signal for all M components in the transformed domain. Since $\tilde{\underline{\mathbf{R}}}(\omega)$, $\tilde{\underline{\mathbf{C}}}(\omega)$, and $\tilde{\underline{\mathbf{F}}}(\omega)$ are diagonal matrices, the mth component of the error signal $\tilde{E}_m(\omega)$ in the transformed domain is given by the following relation:

$$\tilde{E}_m(\omega) = \tilde{F}_m(\omega)\tilde{D}_m(\omega) - \tilde{R}_m(\omega)\tilde{C}_m(\omega)\tilde{D}_m(\omega), \tag{24}$$

where $\tilde{R}_m(\omega)$, $\tilde{C}_m(\omega)$, and $\tilde{F}_m(\omega)$ denote the mth component of the main diagonal of $\tilde{\underline{\mathbf{R}}}(\omega)$, $\tilde{\underline{\mathbf{C}}}(\omega)$, and $\tilde{\underline{\mathbf{F}}}(\omega)$, respectively. The error $\tilde{E}_m(\omega)$ depends only on the mth component of the respective signals and systems in the transformed domain. Thus, Eq. (24) states that the MIMO adaptive inverse filtering problem can be decomposed into M single-input/single-output (SISO) adaptive inverse filtering problems using the GSVD. The computation of the pre-equalization filters can be performed independently for each of the M transformed components. The transformation of the systems and signals is performed by transforming them into the joint eigenspace of $\underline{\mathbf{R}}(\omega)$ and $\underline{\mathbf{F}}(\omega)$ using the GSVD. Therefore this approach will be referred to as *eigenspace inverse adaptive filtering*. Please note that the transformation is not dependent on the driving signals. Figure 4 illustrates the eigenspace inverse adaptive filtering approach.

D. Eigenspace adaptive filtering

In the following, the normal equation of the multichannel adaptive pre-equalization problem presented in Sec. III A will be specialized to the decoupled MIMO system. Due to the decoupling, the cost function $\xi(\hat{\mathbf{c}}, k)$ given by Eq. (5) can

be minimized independently for each component $m = 1, \ldots, M$. The normal equation in the transformed domain is then given as

$$\hat{\tilde{\mathbf{\Phi}}}_{dd,m}(k)\hat{\tilde{\mathbf{c}}}_m(k) = \hat{\tilde{\mathbf{\Phi}}}_{da,m}(k), \tag{25}$$

where $\hat{\tilde{\mathbf{\Phi}}}_{dd,m}(k)$ denotes the time-averaged autocorrelation matrix of the mth component of the transformed filtered loudspeaker driving signal, $\hat{\tilde{\mathbf{\Phi}}}_{da,m}(k)$ the corresponding cross-correlation matrix between the filtered loudspeaker driving signal and the desired signal, and $\hat{\tilde{\mathbf{c}}}_m(k)$ the filter coefficients. The autocorrelation matrix $\hat{\tilde{\mathbf{\Phi}}}_{dd,m}(k)$ has the dimensions $N_c \times N_c$. Due to this reduction in dimensionality, the solution of the M equations given by Eq. (25) is much more efficient than for the adaptation using the original (not transformed) signals. Table I summarizes the complexity reduction of the eigenspace approach to MIMO inverse adaptive filtering in contrast to the X-RLS algorithm.

Equation (25) corresponds to the well-known single channel normal equation.[53] The cross-channel correlations present in $\hat{\mathbf{\Phi}}_{dd}(k)$ have been removed in the transformed domain by the spatial decoupling of the MIMO systems. Thus, the nonuniqueness and ill-conditioning problem discussed in Sec. III B are highly alleviated. There may still be time-domain correlations present in the filtered input signals which cause problems when solving the normal equation (25). However, there are numerous approaches known in the literature on single-channel adaptive (inverse) filtering to overcome these problems.[53]

In practice, however, a number of problems emerge from the GSVD-based transformations used to decouple the MIMO systems. These will be discussed in the following.

E. Fundamental problems of eigenspace adaptive filtering

The major problems of eigenspace adaptive filtering are that

1. The singular matrices $\underline{\mathbf{X}}(\omega)$, $\underline{\mathbf{V}}(\omega)$, and $\underline{\mathbf{U}}(\omega)$ depend on the room $\underline{\mathbf{R}}(\omega)$ and free-field transfer matrix $\underline{\mathbf{F}}(\omega)$, and
2. Their computation using the GSVD is quite complex.

The contents of the room transfer matrix will be time variant in general due to changes in the acoustic conditions caused by, e.g., temperature changes or persons entering the listening room. This requires that the room transfer matrix $\underline{\mathbf{R}}(\omega)$ has to be identified additionally and that the singular matrices have to be updated each time the room transfer matrix changes. Both tasks are computationally very demanding for massive multichannel systems. In general no optimizations can be performed in practice without placing restrictions on the structure of the transfer matrices $\underline{\mathbf{F}}(\omega)$ and $\underline{\mathbf{R}}(\omega)$. Section V will introduce the concept of wave-domain adaptive filtering in order to overcome these problems.

V. WAVE-DOMAIN ADAPTIVE FILTERING

In Sec. IV an eigenspace approach to adaptive inverse filtering was proposed. Its main feature was the decoupling of the MIMO adaptive inverse filtering problem into a series of single channel adaptive inverse filtering problems. This way most of the fundamental problems (see Sec. III B) of adapting the room compensation filters for massive multichannel systems were solved. However, the computation of suitable transformations using the GSVD may become very complex. In order to overcome this problem the concept of wave-domain adaptive filtering is introduced.

A. Generic concept of wave-domain adaptive filtering

Wave-domain adaptive filtering is based on two basic ideas:

1. Explicit consideration of the characteristics of the propagation medium for the derivation of suitable transformations, and
2. Approximation of the concept of perfect decoupling of the MIMO adaptation problem.

The listening room transfer matrix describes the sound transmission from the loudspeakers to the analysis positions with respect to the characteristics of the propagation medium and the boundary conditions imposed by the listening room. Hence it has to fulfill the wave equation and the homogeneous boundary conditions imposed by the room. This knowledge can be used to construct efficient transformations. Since these transformations inherently have to account for the wave nature of sound in order to perform well, this approach will be referred to as *wave domain adaptive (inverse) filtering* (WDAF) and the transformed domain as *wave domain* in the following.

The second idea is to approximate the perfect decoupling of the MIMO adaptation problem in favor of generic transformations which are to some degree independent of the listening room characteristics. In order to keep the complexity low, these generic transformations need not to be strictly diagonalizing, but they should still represent the MIMO system with as few paths as possible.

The combination of both ideas allows one to derive fixed transformations that provide nearly the same favorable properties as the optimal GSVD-based transformations used for eigenspace adaptive inverse filtering with the benefit of computational efficiency.

Based on the approach of eigenspace adaptive filtering and the above-mentioned considerations a generic block diagram of the WDAF approach can be developed. Figure 5 displays this generic block diagram. The signal and system transformations are performed by three generic transformations. Their structure is not limited to the MIMO FIR systems derived from the GSVD. Transformation $\mathcal{T}_1$ transforms the driving signals into the wave domain, $\mathcal{T}_2$ inversely transforms the filtered loudspeaker driving signals from the wave domain, and $\mathcal{T}_3$ transforms the signals at the analysis points into the wave domain. As previously for the eigenspace domain, the signals and transfer functions in the wave domain are denoted by a tilde over the respective variable, since suitable transforms will be based on the idea of a transformation into the eigenspace of the respective systems. The adaptation is performed entirely in the wave domain.

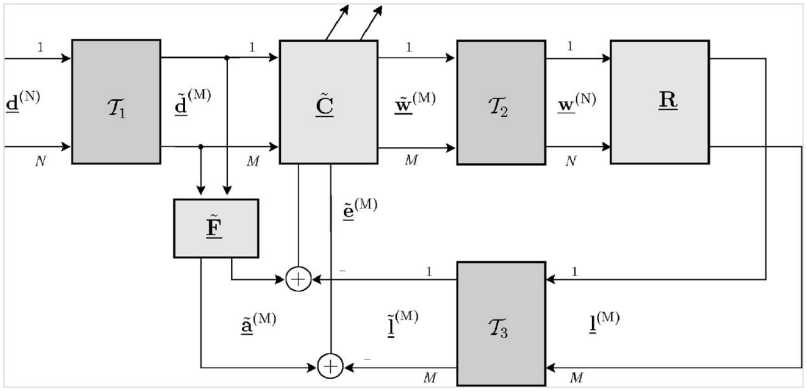


FIG. 5. Block diagram illustrating the wave domain adaptive inverse filtering approach to active room compensation.

Note that the generic block diagram depicted by Fig. 5 also includes the eigenspace adaptive inverse filtering approach. In this special case the transformations are given as the following MIMO FIR systems (see Sec. IV): $\mathcal{T}_1 = \underline{\mathbf{U}}^H(\omega)$, $\mathcal{T}_2 = \underline{\mathbf{V}}(\omega)$, and $\mathcal{T}_3 = \underline{\mathbf{X}}^H(\omega)$.

The following section introduces wave field expansions which are potential candidates as wave domain transformations. Since they are based upon the fundamental free-field solutions of the wave equation they will only decouple the free-field transfer matrix.

B. Wave field representations

A first choice to characterize an acoustic wave field is given by the acoustic pressure $P(\mathbf{x}, \omega)$ captured at all positions $\mathbf{x}$. For a wide variety of applications it is convenient to represent acoustic wave fields with respect to an orthogonal basis. This basis is typically constructed from the fundamental solutions of the wave equation considering the particular problem. The general solution of the wave equation is then given as a weighted superposition of all elementary solutions. Of special interest in the following are decompositions which are based on the fundamental free-field solutions of the wave equation. These depend on the underlying coordinate system and its dimensionality. The basis functions connected to spherical, cylindrical, and Cartesian coordinates are known as spherical harmonics, cylindrical harmonics, and plane waves. An arbitrary wave field can be represented by the expansion coefficients with respect to these basis functions. If two-dimensional wave fields are considered then the polar and Cartesian coordinate systems are common. Here the basis functions are circular harmonics and (two-dimensional) plane waves. Please note that according to the term "spherical harmonics" used for the elementary solutions of the wave equation in spherical coordinates,[58–60] the solutions in cylindrical coordinates have been termed as *cylindrical harmonics* and the solutions in polar coordinates as *circular harmonics* within this article. Typical sound reproduction systems aim at the reproduction in a plane only. The analysis of the reproduced wave field is then typically performed in the reproduction plane. We will therefore limit ourselves to two-dimensional wave field representations for the following discussion of the circular harmonics and plane wave decomposition. However, since the reproduction will take place in a three-dimensional environment several artifacts of two-dimensional sound reproduction and analysis have to be considered.[61] A generalization of the proposed decompositions to three-dimensional representations can be derived straight-forwardly.[58]

1. Circular harmonics expansion

The elementary solutions of the wave equation using cylindrical coordinates are given in Refs. 58 and 62. The elementary solutions using polar coordinates can be derived from these by discarding the components which depend on the z coordinate. The circular harmonics decomposition of an arbitrary wave field is then given by

$$P_P(\mathbf{x}_p, \omega) = \sum_{\nu=-\infty}^{\infty} \left(\breve{P}^{(1)}(\nu, \omega) H_\nu^{(1)}\left(\left| \frac{\omega}{c} \right| r \right) e^{j\nu\alpha} + \breve{P}^{(2)}(\nu, \omega) H_\nu^{(2)}\left(\left| \frac{\omega}{c} \right| r \right) e^{j\nu\alpha} \right), \tag{26}$$

where $\mathbf{x}_p = [\alpha\ r]^T$ denotes the position vector in polar coordinates, $H_\nu^{(1),(2)}(\cdot)$ the νth order Hankel function of first/second kind, and ν the angular frequency. The polar coordinates are defined as $x = r\cos\alpha$ and $y = r\sin\alpha$ within this article. Quantities and functions whose arguments are given in polar coordinates are denoted by a P in their index. The infinite sum over ν can be interpreted as Fourier series with respect to the angle α. The coefficients $\breve{P}^{(1),(2)}(\nu, \omega)$ are referred to as circular harmonics expansion coefficients in the following and will be denoted by a breve over the respective variable. The Hankel function $H_\nu^{(1)}(|\omega/c|r)$ belongs to an incoming (converging) and $H_\nu^{(2)}(|\omega/c|r)$ to an outgoing (diverging) cylindrical wave.[58] Thus, the expansion coefficient $\breve{P}^{(1)}(\nu, \omega)$ describes the incoming wave field, whereas $\breve{P}^{(2)}(\nu, \omega)$ describes the outgoing wave field. According to Eq. (26) the total wave field is given as a superposition of incoming and outgoing contributions.

In order to get more insight into the circular harmonics expansion a closer look at the basis functions is taken. Each spatial variable has its own basis function. The angular coordinate α has an exponential function as basis. Figure 6 illustrates the angular basis functions for different angular frequencies. For the sake of illustration the plots only show the absolute value of the real part. It can be seen clearly that the angular basis functions exhibit a spatial selectivity in the angular coordinate. They can be interpreted as directivity

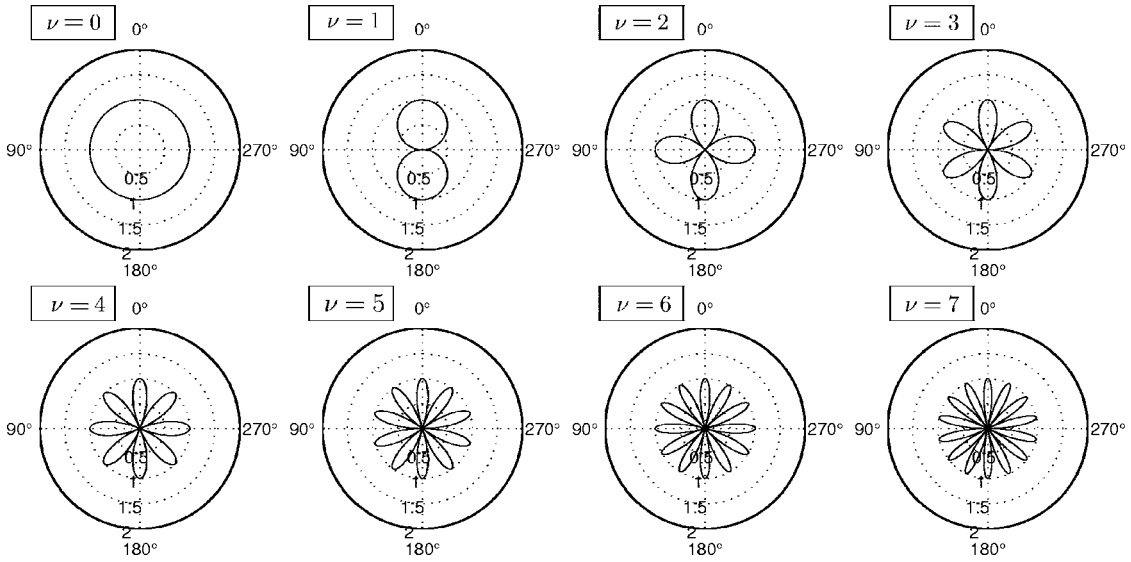


FIG. 6. Illustration of the angular basis functions $e^{j\nu\alpha}$ of the circular harmonics. The plots show the absolute value of the real part ($|\Re\{e^{j\nu\alpha}\}|$) for different angular frequencies ν.

patterns. Hankel functions are the basis functions of the radial coordinate r. See Ref. 63 for a detailed discussion of their properties.

The circular harmonics expansion coefficients for a particular problem can be derived by introducing the expansion (26) into the underlying wave equation. However, it is also possible to measure an arbitrary wave field in a bounded region and derive the expansion coefficients from these measurements. Due to the underlying geometry circular microphone arrays are well suited for this task.[64]

2. Plane wave expansion

Arbitrary solutions of the wave equation can be expressed alternatively as superposition of plane waves traveling into all possible directions. The plane wave expansion coefficients $\breve{P}^{(1)}(\theta,\omega)$ and $\breve{P}^{(2)}(\theta,\omega)$ describe the spectrum of incoming/outgoing plane waves with incidence angle θ. They can be derived by a plane wave decomposition.[43,64] The plane wave expansion coefficients exhibit a direct link to the expansion coefficients in circular harmonics

$$\breve{P}^{(1)}(\theta,\omega) = \frac{4\pi}{k}\sum_{\nu=-\infty}^{\infty} j^{\nu}\check{P}^{(1)}(\nu,\omega)e^{j\nu\theta}, \tag{27a}$$

$$\breve{P}^{(2)}(\theta,\omega) = \frac{4\pi}{k}\sum_{\nu=-\infty}^{\infty} j^{\nu}\check{P}^{(2)}(\nu,\omega)e^{j\nu\theta}. \tag{27b}$$

Equation (27) states that the plane wave decomposition of a wave field is, up to the factor j^{ν}, given by the Fourier series of the expansion coefficients in terms of circular harmonics.

C. Circular harmonics as wave domain transformation

The optimal choice for the transformed signal representation depends on the acoustic characteristics of the listening room and the desired free-field response. Two different wave field representations have been introduced in Sec. IV: the decomposition into plane waves and the decomposition into circular harmonics. Both are based upon orthogonal basis functions derived from the free-field wave equation and will therefore provide a decoupling of the free-field transfer matrix $\underline{\mathbf{F}}(\omega)$ but not necessarily of the room transfer matrix $\underline{\mathbf{R}}(\omega)$. Either representation may be nevertheless a suitable choice for the wave domain representation as long as they provide an approximate decoupling of the room transfer matrix. In order to qualitatively evaluate the two proposed decompositions a closer look is taken at the room response with respect to the basis functions. In the ideal case the reproduction of a basis function will result only in contributions belonging to the same basis function after analysis. Consider now typical listening rooms with a rectangular shape as a first approximation and their response to either plane waves or circular harmonics.

The response to a plane wave of a certain angle consists in general of a mixture of plane waves with all kinds of different angles. This means that the desired effect of decoupling is not achievable with plane waves in more or less rectangular rooms.

On the other hand, the response to a circular wave depends on the wavelength. Since high frequencies can be damped effectively by passive methods, active room compensation can be restricted to low frequencies. For the corresponding long wavelengths, scattering at the corners of the listening room does not play a major role. Therefore each circular harmonics component emitted by a well-designed reproduction system will lead to reflections which can be described mainly by the same component. Consequently the desired decoupling effect will be more prominent than for plane waves. This intuitive consideration suggests that circular harmonics seem to be a more suitable choice for the wave domain representation than plane waves. This assumption holds also for reverberant rooms as long as the energy of other components than the emitted one is considerably attenuated.

Please note that the representation of a field in a transformed domain requires adequate sampling of the measured field. Sampling theorems for circular microphone arrays can be found, e.g., in Refs. 36, 43, 65, and 66.

VI. APPLICATION TO WAVE FIELD SYNTHESIS

Section V introduced wave domain adaptive filtering. Now its application to spatial reproduction techniques is shown using wave field synthesis as an example. First, this reproduction technique is discussed shortly. Then the application of wave domain adaptive filtering to wave field synthesis is shown by specifying the transformations $\mathcal{T}_1$ through $\mathcal{T}_3$ in Fig. 5. Finally some selected results of performance evaluations are presented.

A. Wave field synthesis

Wave field synthesis (WFS) is a massive multichannel sound reproduction technique which overcomes the "sweet spot" limitation well known from stereophonic surround sound methods. WFS techniques are based on Huygen's principle and are formulated in terms of the Kirchhoff-Helmholtz integral.[62] These foundations have been initially developed by the Technical University of Delft[10–12] and were later extended within the European project CARROUSO.[67] Detailed descriptions of WFS can be found in Refs. 13, 14, 43, and 68–72.

B. Application of wave domain adaptive filtering to wave field synthesis

The generic concept of WDAF is now specialized to the reproduction utilizing a WFS system and the circular harmonics decomposition as wave domain transformation. The transformations $\mathcal{T}_1$ through $\mathcal{T}_3$ are based on the decomposition into circular harmonics of the respective wave fields. For active room compensation only the incoming parts (see Sec. V B 1) of the respective wave fields are of interest. For the particular scenario considered these transformations are specialized as follows.

(1) **Transformation** $\mathcal{T}_1$: This transformation transforms the virtual source signal $S(\omega)$ using a spatial model of the virtual source into the loudspeaker driving signals $\breve{\underline{\mathbf{d}}}^{(M)}$ in the wave domain. Suitable models for the virtual source characteristics are line/point sources or plane waves.
(2) **Transformation** $\mathcal{T}_2$: This transformation generates the loudspeaker driving signals from the filtered driving signals $\breve{\underline{\mathbf{w}}}^{(M)}$. Equation (26) together with a suitable loudspeaker selection criterion[73] can be used for this purpose.
(3) **Transformation** $\mathcal{T}_3$: This transformation calculates the circular harmonics decomposition coefficients of the wave field within the listening area from the microphone array measurements.[64]

The free-field transfer matrix $\breve{\underline{\mathbf{F}}}$ models the free-field propagation in terms of circular harmonics from the loudspeakers to the microphone array. In the ideal case this matrix would only model the propagation delay. Besides this, the transfer matrix $\breve{\underline{\mathbf{F}}}$ should also include an additional delay to ensure the computation of causal room compensation filters. If desired, certain artifacts of WFS systems like, e.g., amplitude errors (see Ref. 61) can be considered in the construction of the matrix $\breve{\underline{\mathbf{F}}}$.

It was proposed by the authors in previous publications[40–42,74–76] to perform a Fourier transformation with respect to the angular variable of the plane wave decomposition to derive the signals in the wave domain. This Fourier transformation of the plane wave decomposed signals is up to the factor j^{ν} (and a frequency correction) equivalent to calculating the circular harmonics expansion coefficients, since the plane wave decomposition in terms of circular harmonics is given by the Fourier series (27).

C. Performance evaluation

After describing the application of wave domain adaptive filtering for wave field synthesis, its performance for decoupling the adaptation will now be evaluated. The evaluation of the circular harmonics decomposition for listening room compensation with wave domain adaptive filtering has been described in detail in Ref. 77. The following quote some selected results.

1. Performance measures

The performance of the proposed transformations and of the resulting active listening room compensation system is evaluated by the ability to compact the room characteristics into less coefficients than using the microphone signals directly. This ability is described by two measures: (1) the energy of the elements of the room transfer matrix and (2) the energy compaction performance. The first measure is defined by calculating the energy of each spatial transmission path for the room transfer matrix in its different representations. For its representation in the pressure domain (pressure microphones) the energy of the elements of the room transfer matrix is defined as

$$E(m,n) = \frac{1}{2\pi}\int_{-\infty}^{\infty} |R_{m,n}(\omega)|^2 d\omega. \tag{28}$$

Similar definitions apply to the room transfer matrix in its plane wave and circular harmonics representation yielding the energy representations $\bar{E}(\theta,\theta_0)$ and $\breve{E}(\nu,\nu_0)$. The second measure, the energy compaction, measures the ability of a transform to compact the energy of the room transfer matrix to as few coefficients as possible. For one particular room transfer matrix this measure is defined by calculating the ratio between the energy of the first i dominant elements and the total energy of all elements. For this purpose the energies $E(m,n)$ are sorted in descending order yielding the sorted elements $E_{\mathrm{sort}}(\eta)$. Then the ratio between the energy of the first i sorted elements and the total energy of all elements is calculated as follows:

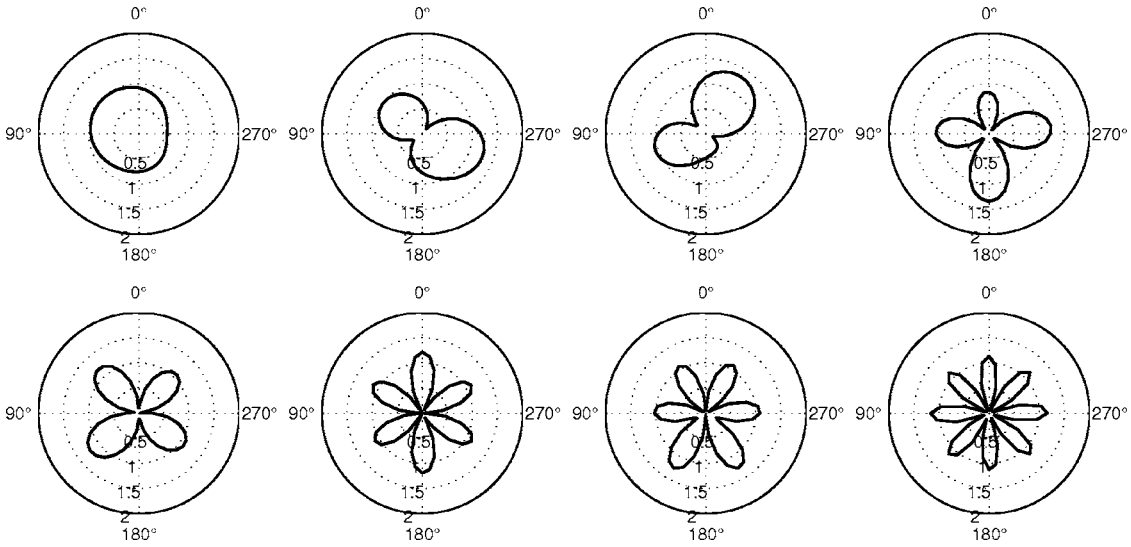


FIG. 7. Absolute value of the first eight right singular vectors ($f=80$ Hz) for the simulated circular WFS/wave field analysis system sorted by their descending singular values (top left to bottom right). The singular vectors for a plane wave reflection factor $R_{\mathrm{pw}}=0.8$ at the walls of the simulated room are shown.

$$\mathrm{EC}(i)=\frac{\sum_{\eta=0}^{i} E_{\mathrm{sort}}(\eta)}{\sum_{\eta=0}^{MN-1} E_{\mathrm{sort}}(\eta)}, \tag{29}$$

where $0\leqslant \mathrm{EC}(i)\leqslant 1$. Equivalent definitions apply for the transformed representations. The larger portion of the total energy that is captured by the first i elements, the better the performance in terms of energy compaction. The energy of the elements $E(m,n)$, $\bar{E}(\theta,\theta_0)$, and $\breve{E}(\nu,\nu_0)$ illustrate the distribution of the energy in the different transformed domains, while $\mathrm{EC}(i)$, $\overline{\mathrm{EC}}(i)$, and $\breve{\mathrm{EC}}(i)$ illustrate the ability of a particular transformation to compact the energy in few coefficients.

2. Evaluation results

The following illustrates the performance of active listening room compensation for WFS using the proposed WDAF approach for simulated acoustical environments. The main benefits of using simulated versus real acoustical environments are that the parameters of the simulated environment can be changed easily in order to simulate different scenarios and that practical aspects such as noise, transducer mismatch, and misplacement can be excluded for a first proof of the WDAF concept.

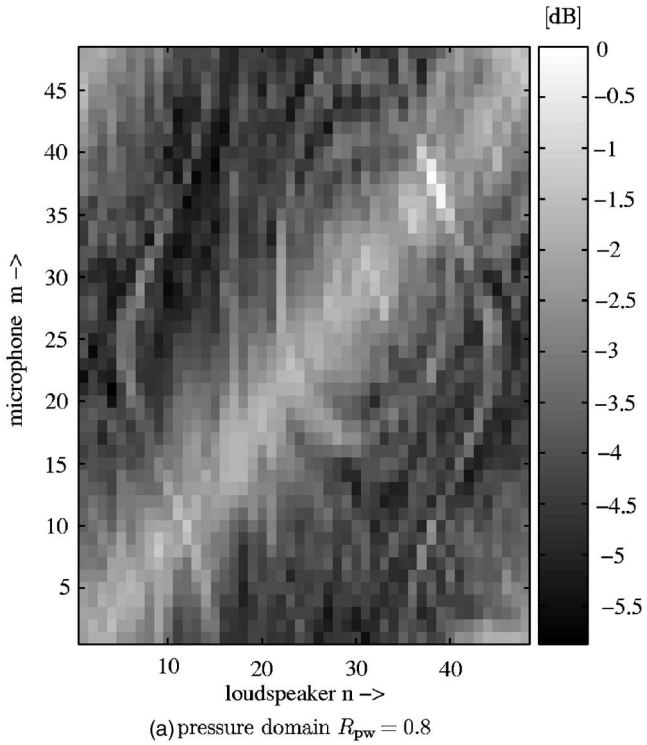
For an accurate simulation of the acoustic environment, a numerical simulation of wave propagation based on the functional transformation method was used.[78] The particular implementation used (wave2d) simulates the wave equation in two dimensions.[79] This method requires no spatial discretization and thus allows one to place the virtual microphones and speakers at their exact spatial positions. This is not possible when using methods which perform a spatial discretization, such as, e.g., the finite element method.[80]

The geometry of the simulated acoustical environment is a simplification of a real environment at the multimedia laboratory of the Multimedia Communications and Signal Processing Group at the University of Erlangen-Nuremberg.[81] Its geometrical and acoustical details as well as the loudspeaker and microphone setups are given in Refs. 43 and 77. The simulated room has the dimensions $5.90\times 5.80\times 3.10$ m $(w\times l\times h)$.

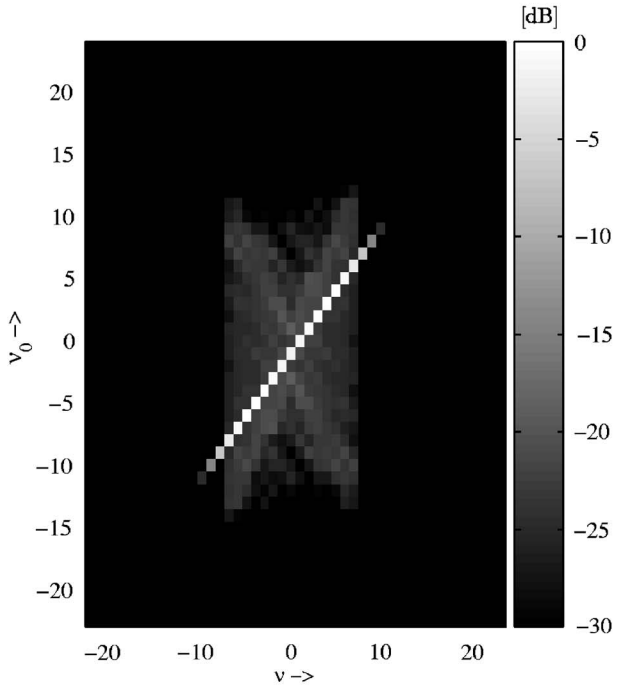
A WFS system with 48 loudspeakers placed equidistant on a circle with a radius of $R_{\mathrm{LS}}=1.50$ m and a wave field analysis system consisting of 48 angular sampling positions at a radius of $R_{\mathrm{mic}}=0.75$ m was simulated. The simulations were performed for a reverberant environment with a plane wave reflection factor of $R_{\mathrm{pw}}=0.8$ at the walls. Due to the spatial aliasing frequency of the WFS system and the microphone array all signals were low-pass filtered to a bandwidth of 650 Hz to calculate the results shown.

An indication for the suitability of circular harmonics is given by calculating the right singular vectors of the room transfer matrix $\boldsymbol{R}(\omega)$ (for the pressure microphones) using the GSVD. Figure 7 shows the absolute value of the first eight right singular vectors for the reverberant case $R_{\mathrm{pw}}=0.8$. The singular vectors have been sorted by their descending singular values and thus by their energy. It can be seen that they look very similar to the circular harmonics illustrated in Fig. 6. This result and the considerations in Sec. V C give justification for using the circular harmonics representation as wave domain transformation.

Figure 8 shows the energy of the elements of the room transfer matrix for the pressure microphones and its circular harmonics representation. Figure 8(a) shows $E(m,n)$ for the pressure microphones. The direct path from the loudspeakers to the microphones can be seen clearly along the main diagonal but also the reflections of the loudspeaker wave fields at the walls of the listening room. Figure 8(b) shows the energy of the room transfer matrix represented in circular harmonics. As desired, the main diagonal elements represent a major portion of the energy. The off-diagonal elements are a result of the reverberant enclosure. Please note the different scales used for Figs. 8(a) and 8(b).



(a) pressure domain $R_{pw} = 0.8$



(b) circular harmonics domain $R_{pw} = 0.8$

FIG. 8. Energy of the room transfer matrix of the signals captured by the pressure microphones $E(m,n)$ and in the circular harmonics domain $\check{E}(\nu,\nu_0)$.

Not only the performance of the circular harmonics in terms of energy compaction toward the main diagonal is of interest, but also its performance to represent the MIMO system by as few coefficients as possible. This can be measured by the energy compaction of the different representations. Results are shown in Fig. 9. Figure 9 illustrates the energy compaction according to Eq. (29) for the room transfer matrix in the pressure domain $\mathrm{EC}(i)$, in the plane wave decomposed domain $\overline{\mathrm{EC}}(i)$, and in the circular harmonics domain

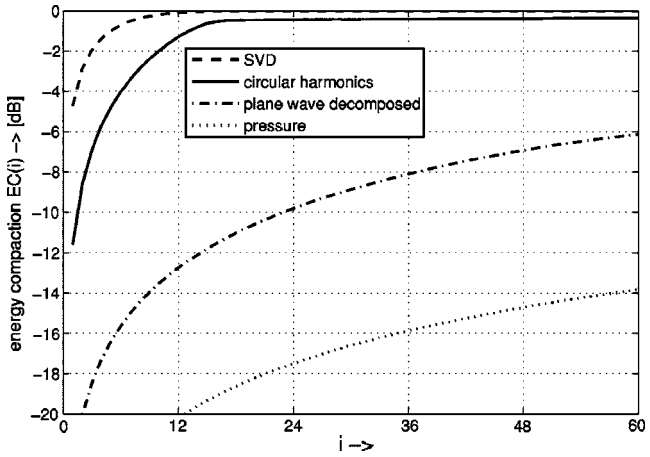


FIG. 9. Energy compaction performance $\mathrm{EC}(i)$ for the different representations of the room transfer matrix. The index i denotes the number of sorted elements which capture most of the total energy.

$\widetilde{\mathrm{EC}}(i)$. The energy compaction of the GSVD is shown for reference. It can be seen clearly that the circular harmonics representation of the room transfer matrix compacts the energy much better than its pressure and plane wave representation. It can also be seen that the GSVD provides the optimal transformation in this sense.

More simulation results for other values of the reflection factor and also for a rectangular loudspeaker array geometry have been reported in Ref. 77. Furthermore, measurements have been conducted for the circular wave field synthesis/analysis system described earlier. These measurements confirm the simulation results.

The presented results indicate that the circular harmonics decomposition is not the optimal transformation for the reverberant case. However, the results also show that this transformation provides a quite reasonable approximation to the optimal GSVD transformation, especially when compared to the plane wave decomposition and pressure microphone representations.

Further results of the proposed algorithm with respect to the adaptation performance and the resulting wave field inside the listening area after compensation can be found in Refs. 41 and 48. They reveal that the proposed algorithm provides fast and stable adaptation even for nonstationary virtual scenes and results in a reproduced wave field without listening room reflections throughout the entire listening area.

VII. CONCLUSIONS

This contribution has derived a method for active listening room compensation for massive multichannel sound reproduction systems. It has been shown that a major challenge is the adaptation of the matrix of compensation filters when straightforwardly applying multichannel adaptation algorithms (like the X-RLS algorithm) to massive multichannel reproduction scenarios. In addition to the computation of the compensation filters also the matrix of room impulse responses has to be estimated. The numerical expense for such a straightforward implementation would be immense: For reproduction systems with even only a few ten loudspeakers, the computation of many thousands of single adaptation

problems would be required for each time step. Furthermore, the convergence of the adaptation will be slow due to a high correlation between the channels.

Suitable mathematical techniques can help to reduce the numerical expense and to improve convergence through decoupling the adaptation problem. More specifically, the matrix of compensation filters, the matrix of room impulse responses, and the free field propagation matrix can be jointly diagonalized through a GSVD. This method has been called eigenspace adaptive filtering because the diagonalization can be seen as a transformation into the eigenspace of the problem. Although eigenspace adaptive filtering provides optimal decoupling, there is a major drawback for the practical application: The GSVD matrices for the transformation into the eigenspace are data dependent, since they depend on the time-varying acoustic room response characteristics. Calculating these matrices anew in each time step would create another numerical problem.

At this point, the situation is similar as for coding of audio or image data. The optimal energy compaction and decorrelation could be achieved by the Karhunen-Loéve transform, but the correct transformation matrix is data dependent and hence not known. Thus practical coders are realized with a different transformation which does not depend on the input data. For a wide variety of problems in audio and video coding, the discrete cosine transformation has proven to approximate the properties of the Karhunen-Loève transform reasonably well. A similar route has been taken here. The search for an approximation of the transformation into the eigenspace by GSVD has been motivated by physical intuition. Inspection of the eigenvectors of the GSVD from simulations and measurements has shown that they resemble the circular harmonics decomposition of free field acoustics. This notion led us to the idea that an approach in which the optimal transformations into the eigenspace is replaced by a transformation into the space of circular harmonics, the so-called *wave domain*, may give results that are only slightly suboptimal.

The implementation of the transformation into the wave domain depends on the specific spatial reproduction/analysis technique. Details have been given for reproduction by wave field synthesis. It turned out that the decoupling properties of the transformation into the wave domain are still quite good and close to the theoretical optimum of the GSVD. In summary, it has been shown that active listening room compensation for massive multichannel sound reproduction is a tractable problem. For spatial reproduction with wave field synthesis, wave domain adaptive filtering provides a solution which is both numerically feasible and almost optimal. It remains to extend these investigations to other spatial reproduction techniques.

[1]E. Torick, "Highlights in the history of multichannel sound," J. Audio Eng. Soc. **46**, 27–31 (1998).

[2]G. Steinke, "Surround sound—The new phase. An overview," in 100th Audio Engineering, Society Convention, Copenhagen, Denmark, May 1996.

[3]D. Griesinger, "Surround: The current technological situation," in 108th Audio Engineering Society Convention, Paris, France, February 2000.

[4]S. Kiriakakis, P. Tsakalides, and T. Holman, "Surrounded by sound," IEEE Signal Process. Mag. **16**, 55–66 (1999).

[5]M. Poletti, "A unified theory of horizontal holographic sound systems," J. Audio Eng. Soc. **48**, 1155–1182 (2000).

[6]F. Rumsey, *Spatial Audio* (Focal Press, 2001).

[7]J. Meyer, "Spherical microphone arrays for 3D sound reproduction," in *Audio Signal Processing for Next-Generation Multimedia Communication Systems*, edited by Y. Huang and J. Benesty (Kluwer Academic, Dordrecht, 2004).

[8]S. Moreau and J. Daniel, "3D sound field recording with higher order ambisonics—Objective measurements and validation of a 4th order spherical microphone," in 120th Audio Engineering Society Convention, Paris, France, 2006.

[9]M. Gerzon, "Ambisonics in multichannel broadcasting and video," J. Acoust. Soc. Am. **33**, 859–871 (1985).

[10]A. Berkhout, "A holographic approach to acoustic control," J. Audio Eng. Soc. **36**, 977–995 (1988).

[11]D. de Vries, E. Start, and V. Valstar, "The wave field synthesis concept applied to sound reinforcement: Restrictions and solutions," in 96th Audio Engineering Society Convention, Amsterdam, The Netherlands, February 1994.

[12]J. Sonke, D. de Vries, and J. Labeeuw, "Variable acoustics by wave field synthesis: A closer look at amplitude effects," in 104th Audio Engineering Society Convention, Amsterdam, The Netherlands, May 1998.

[13]S. Spors, H. Teutsch, A. Kuntz, and R. Rabenstein, "Sound field synthesis," in *Audio Signal Processing for Next-Generation Multimedia Communication Systems*, edited by Y. Huang and J. Benesty (Kluwer Academic, Dordrecht, 2004).

[14]R. Rabenstein and S. Spors, "Wave field synthesis techniques for spatial sound reproduction," in *Topics in Acoustic Echo and Noise Control*, edited by E. Haensler and G. Schmidt (Springer, New York, 2006), Chap. 13, pp. 517–545.

[15]D. Griesinger, "Multichannel sound systems and their interaction with the room," in 16th International Conference on Audio, Acoustics, and Small Places, Audio Engineering Society, 1998, pp. 159–173.

[16]M. Dewhirst, S. Zielinski, P. Jackson, and F. Rumsey, "Objective assessment of spatial localization attributes of surround-sound reproduction systems," in 118th Audio Engineering Society Convention, Barcelona, Spain, May 2005.

[17]B. Klehs and T. Sporer, "Wave field synthesis in the real world. 1. In the living room," in 114th Audio Engineering Society Convention, Amsterdam, The Netherlands, March 2003.

[18]E. Völker, "To nearfield monitoring of multichannel reproduction—Is the acoustics of the living room sufficient?," in Proceedings of the Tonmeistertagung, Hannover, Germany, 1998.

[19]E. Völker, W. Teuber, and A. Bob, "5.1 in the living room—On acoustics of multichannel reproduction," in Proceedings of the Tonmeistertagung, Hannover, Germany, 2002.

[20]E.-J. Voelker, "Home cinema surround sound—Acoustics and neighbourhood," in 100th Audio Engineering Society Convention, Copenhagen, Denmark, May 1996.

[21]T. Sporer and B. Klehs, "Wave field synthesis in the real world. 2. In the movie theatre," in 116th Audio Engineering Society Convention, Berlin, Germany, May 2005.

[22]P. Gauthier and A. Berry, "Sound-field reproduction in-room using optimal control techniques: Simulations in the frequency domain," J. Acoust. Soc. Am. **117**, 662–678 (2005).

[23]S. Bech, "Timbral aspects of reproduced sound in small rooms. I," J. Acoust. Soc. Am. **97**, 1717–1726 (1995).

[24]S. Bech, "Timbral aspects of reproduced sound in small rooms. II," J. Acoust. Soc. Am. **99**, 3539–3549 (1996).

[25]S. Bech, "Spatial aspects of reproduced sound in small rooms," J. Acoust. Soc. Am. **103**, 434–445 (1998).

[26]T. Caulkins and O. Warusfel, "Characterization of the reverberant sound field emitted by a wave field synthesis driven loudspeaker array," in 120th Audio Engineering Society Convention, Paris, France, May 2006.

[27]E. Corteel and R. Nicol, "Listening room compensation for wave field synthesis. What can be done?," in 23rd Audio Engineering Society Convention, Copenhagen, Denmark, May 2003.

[28]L. Fielder, "Practical limits for room equalization," in 111th Audio Engineering Society Convention, New York, September 2001.

[29]L. Fielder, "Analysis of traditional and reverberation-reducing methods of room equalization," J. Audio Eng. Soc. **51**, 3–26 (2003).

[30]P. Hatziantoniou and J. Mourjopoulos, "Errors in real-time room acoustics dereverberation," J. Audio Eng. Soc. **52**, 883–899 (2004).

[31]J. Mourjopoulos, "Digital equalization of room acoustics," J. Audio Eng.

Soc. **42**, 884–900 (1994).
[32]F. Talantzis and D. Ward, "Multi-channel equalization in an acoustic reverberant environment: Established of robustness measures," in Institute of Acoustics Spring Conference, Salford, UK, March 2002.
[33]F. Talantzis and D. Ward, "Robustness of multichannel equalization in an acoustic reverberant environment," J. Acoust. Soc. Am. **114**, 833–841 (2003).
[34]S. Bharitkar, P. Hilmes, and C. Kyriakakis, "Sensitivity of multichannel room equalization to listener position," in IEEE International Conference on Multimedia and Expo (ICME), Baltimore, MD, 2003.
[35]P. Nelson, F. Orduna-Bustamante, and H. Hamada, "Inverse filter design and equalization zones in multichannel sound reproduction," IEEE Trans. Speech Audio Process. **3**, 185–192 (1995).
[36]T. Betlehem and T. Abhayapala, "Theory and design of sound field reproduction in reverberant rooms," J. Acoust. Soc. Am. **117**, 2100–2111 (2005).
[37]T. Betlehem and T. Abhayapala, "A modal approach to soundfield reproduction in reverberant rooms," in IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), Philadelphia, PA, 2005, Vol. **III**, pp. 289–292.
[38]A. Sontacchi and R. Hoeldrich, "Enhanced 3D sound field synthesis and reproduction system by compensating interfering reflections," in Conference on Digital Audio Effects (DAFX-00), Verona, Italy, 2000.
[39]M. Poletti, "Three-dimensional surround sound systems based on spherical harmonics," J. Audio Eng. Soc. **53**, 1004–1025 (2005).
[40]S. Spors, H. Buchner, and R. Rabenstein, "A novel approach to active listening room compensation for wave field synthesis using wave-domain adaptive filtering," in IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), (2004).
[41]S. Spors, H. Buchner, and R. Rabenstein, "Adaptive listening room compensation for spatial audio systems," in European Signal Processing Conference (EUSIPCO) (2004).
[42]S. Spors, H. Buchner, and R. Rabenstein, "Efficient active listening room compensation for wave field synthesis," in 116th Audio Engineering Society Convention, Berlin, Germany, 2004.
[43]S. Spors, "Active listening room compensation for spatial sound reproduction systems," Ph.D. thesis, University of Erlangen-Nuremberg, 2006.
[44]J. Lopez, A. Gonzalez, and L. Fuster, "Room compensation in wave field synthesis by means of multichannel inversion," in IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, New Paltz, NY, 2005.
[45]R. von Zon, E. Corteel, D. de Vries, and O. Warusfel, "Multi-actuator panel (map) loudspeakers: How to compensate for their mutual reflections," in 116th Audio Engineering Society Convention, Berlin, Germany, 2004.
[46]P. Gauthier and A. Berry, "Adaptive wave field synthesis with independent radiation mode control for active sound field reproduction: Theory," J. Acoust. Soc. Am. **119**, 2721–2737 (2006).
[47]M. Omura, M. Yada, H. Saruwatari, S. Kajita, K. Takeda, and F. Itakura, "Compensating of room acoustic transfer functions affected by change of room temperature," in IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), Phoenix, AZ, 1999.
[48]S. Petrausch, S. Spors, and R. Rabenstein, "Simulation and visualization of room compensation for wave field synthesis with the functional transformation method," in 119th Audio Engineering Society Convention, New York, 2005.
[49]M. Sondhi, D. Morgan, and J. Hall, "Stereophonic acoustic echo cancellation—An overview of the fundamental problem," IEEE Signal Process. Lett. **2**, 148–151 (1995).
[50]J. Garas, *Adaptive 3D Sound Systems* (Kluwer Academic, Dordrecht, 2000).
[51]A. Oppenheim and R. Schafer, *Discrete-Time Signal Processing* (Prentice-Hall, Englewood Cliffs, NJ, 1999).
[52]M. Miyoshi and Y. Kaneda, "Inverse filtering of room acoustics," IEEE Trans. Acoust., Speech, Signal Process. **36**, 145–152 (1988).
[53]S. Haykin, *Adaptive Filter Theory* (Prentice-Hall, Englewood Cliffs, NJ, 1996).
[54]J. Benesty, D. Morgan, and M. Sondhi, "A better understanding and an improved solution to the specific problems of stereophonic acoustic echo cancellation," IEEE Trans. Speech Audio Process. **6**, 156–165 (1998).
[55]M. Bouchard and S. Quednau, "Multichannel recursive-least-squares algorithms and fast-transversal-filter algorithms for active noise control and sound reproduction systems," IEEE Trans. Speech Audio Process. **8**, 606–618 (2000).
[56]G. Golub and C. Loan, *Matrix Computations* (The Johns Hopkins University Press, Baltimore, MD, 89).
[57]P. Flikkma, "An algebraic theory of 3D sound synthesis with loudspeakers," in 22nd International Conference on Virtual, Synthetic and Entertainment Audio, Audio Engineering Society, 2002.
[58]E. Williams, *Fourier Acoustics: Sound Radiation and Nearfield Acoustical Holography* (Academic, New York, 1999).
[59]A. Pierce, *Acoustics. An Introduction to its Physical Principles and Applications* (Acoustical Society of America, New York, 1991).
[60]D. Blackstock, *Fundamentals of Physical Acoustics* (Wiley, New York, 2000).
[61]S. Spors, M. Renk, and R. Rabenstein, "Limiting effects of active room compensation using wave field synthesis," in 118th Audio Engineering Society Convention, Barcelona, Spain, 2005.
[62]P. Morse and H. Feshbach, *Methods of Theoretical Physics. Part I* (McGraw-Hill, New York, 1953).
[63]M. Abramowitz and I. Stegun, *Handbook of Mathematical Functions* (Dover, New York, 1972).
[64]E. Hulsebos, D. de Vries, and E. Bourdillat, "Improved microphone array configurations for auralization of sound fields by Wave Field Synthesis," in 110th Audio Engineering Society Convention, Amsterdam, The Netherlands, May 2001.
[65]H. Teutsch, "Wavefield decomposition using microphone arrays and its application to acoustic scene analysis," Ph.D. thesis, University of Erlangen-Nuremberg (2005), http://www.lnt.de/lms/publications (last viewed on 16 April 2007).
[66]T. Ajdler, L. Sbaiz, and M. Vetterli, "Plenacoustic function on the circle with application to HRTF interpolation," in IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), Philadephia, PA (2005).
[67]S. Brix, T. Sporer, and J. Plogsties, "CARROUSO—An European approach to 3D-audio," in 110th Audio Engineering Society Convention, 2001.
[68]E. Start, "Direct sound enhancement by wave field synthesis," Ph.D. thesis, Delft University of Technology, Delft, The Netherlands, 1997.
[69]E. Verheijen, "Sound reproduction by wave field synthesis," Ph.D. thesis, Delft University of Technology, Delft, The Netherlands, 1997.
[70]P. Vogel, "Application of wave field synthesis in room acoustics," Ph.D. thesis, Delft University of Technology, Delft, The Netherlands, 1993.
[71]W. de Bruijn, "Application of wave field synthesis in videoconferencing," Ph.D. thesis, Delft University of Technology, Delft, The Netherlands, 2004.
[72]E. Hulsebos, "Auralization using wave field synthesis," Ph.D. thesis, Delft University of Technology, Delft, The Netherlands, 2004.
[73]S. Spors, "An analytic secondary source selection criteria for wave field synthesis," in 33rd German Annual Conference on Acoustics (DAGA), Stuttgart, Germany, March 2007.
[74]H. Buchner, S. Spors, and W. Kellermann, "Wave-domain adaptive filtering: Acoustic echo cancellation for full-duplex systems based on wavefield synthesis," in IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), 2004.
[75]H. Buchner, S. Spors, and W. Kellermann, "Wave-domain adaptive filtering for acoustic human-machine interfaces based on wavefield analysis and synthesis," in European Signal Processing Conference (EUSIPCO), 2004.
[76]H. Buchner, S. Spors, and W. Kellermann, "Full-duplex systems for sound field recording and auralization based on wave field synthesis," in 116th Audio Engineering Society Convention, Berlin, Germany, 2004.
[77]S. Spors and R. Rabenstein, "Evaluation of the circular harmonics decomposition for WDAF-based active listening room compensation," in 28th AES Conference: The Future of Audio Technology—Surround and Beyond, Audio Engineering Society, Pitea, Sweden, June/July 2006, pp. 134–149.
[78]S. Petrausch and R. Rabenstein, "Highly efficient simulation and visualization of acoustic wave fields with the functional transformation method," in *Simulation and Visualization*, (Otto von Guericke Universität, Magdeburg, 2005), pp. 279–290.
[79]"Wave2D: Wave field simulation in 2D," http://www.lnt.de/~stepe/demo/Wave2D (last viewed on 16 April 2007).
[80]J. Reddy, *Introduction to the Finite Element Method* (McGraw Hill, New York, 2005).
[81]"Multimedia communications and signal processing at the University of Erlangen-Nuremberg," http://www.lnt.de/LMS (last viewed on 16 April 2007).

Steering vector sensor array elements with linear cardioids and nonlinear hippioids

Kevin B. Smith[a)]
Department of Physics, Naval Postgraduate School, Monterey, California 93943

A. Vincent van Leijen
Combat Systems Department, Netherlands Defence Academy, Den Helder, Netherlands

(Received 18 August 2006; revised 9 March 2007; accepted 9 March 2007)

The purpose of this work is to study the impact due to individual vector sensor element steering patterns on linear array beamforming. Standard, linear beamformers employ cardioid beampatterns for each vector sensor. In this work, a class of vector sensor element steering patterns beyond the standard cardioid was examined. The element weighting is nonlinear but nonadaptive, making it simple to implement in hardware processing. The new sensor steering patterns, referred to as hippioids, are products of cardioids and various powers of hippopedes. The angular resolution of individual sensors, the impact on angular resolution from arrays of varying aperture, and peak-to-sidelobe levels will serve as performance measures. An example of the differences in vector sensor steering patterns is provided using measured directional frequency and recorded buoy data.
 [DOI: 10.1121/1.2722054]

PACS number(s): 43.60.Fg, 43.58.Fm [EJS] Pages: 370–377

I. INTRODUCTION

Standard array processing techniques for vector sensors[1] are based on a (properly scaled) linear summation of the different signals measured by pressure and three orthogonal components of particle velocity. This leads to the so-called cardioid pattern for each vector sensor element in the array. However, the angular resolution of the cardioid peak is poor, and the final resolution of the array is generally defined by the array properties (e.g., element spacing and aperture). Thus, cardioid patterns do little to impact the ability of a towed array to localize a target beyond the right-left ambiguity discrimination near broadside and potentially several dB of signal gain. (Note that there are benefits from steering the more highly resolved null to reject certain acoustic noise paths.) In this work, we examine other vector sensor element steering patterns beyond the standard cardioid, specifically a new beam shape referred to as a hippioid (pronounced "hippie-oid").

In the following sections, we develop the fundamental expressions defining these new beampatterns in general, three-dimensional coordinates. Examples of vector sensor plane-wave responses are provided, and comparisons are made between the standard, linear cardioid shape and the higher order, nonlinear hippioid structures. We then present a simple analysis of some directional frequency and recording (DIFAR) buoy data that compare the performance of the cardioid with the second order hippioid. This is followed by some numerical examples using simple, analytical representations of incoming signals on arrays of vector sensors. The effects of varying array aperture are also examined. The results of this analysis are used to compare the performance of the different approaches, and to highlight issues with multipath resolution. We conclude the paper with a summary of findings.

II. THEORETICAL TREATMENT

For the theoretical development of processing routines described below, we shall assume that the output of a vector sensor is the acoustic pressure, p, and the three components of vector velocity, $\mathbf{v}=(v_x,v_y,v_z)$. A Cartesian coordinate system is assumed here for simplicity, but can be adapted to cylindrical or spherical coordinates, if needed. For the data presented here, the z axis is coincident with the array axis and azimuthal angles are measured relative to broadside, as depicted in Fig. 1. The scaling factor between pressure and velocity is the acoustic impedance, ρc, where ρ is the local ambient density of the fluid medium and c is the local sound speed. All acoustic quantities are assumed to have the same $e^{-i\omega t}$ time dependence, and

$$\mathbf{v}=-\frac{i}{\omega\rho}\nabla p. \tag{1}$$

For plane-wave beamforming, the processor is typically defined by[1,2]

$$B^{(pv)}(\theta_s,\phi_s)=\left|\sum_n (w_{xn}v_{xn}+w_{yn}v_{yn}+w_{zn}v_{zn}+w_{pn}v_{pn})e^{-i\mathbf{k}_s\cdot\mathbf{r}_n}\right|^2, \tag{2}$$

where $\mathbf{k}$ is the wave number vector of the propagating plane wave, $\mathbf{v}_n=\mathbf{V}_n e^{i\mathbf{k}\cdot\mathbf{r}_n}$ (e.g., $v_{nx}=V_{nx}e^{i\mathbf{k}\cdot\mathbf{r}_n}$, etc), $v_{pn}=V_{pn}e^{i\mathbf{k}\cdot\mathbf{r}_n}$, $V_{pn}=P_n/\rho c$ and $\mathbf{k}_s\cdot\mathbf{r}_n=kz_n\sin\phi_s$ indicates the steering direction of the array. This is a linear processor, as the in-phase elements are simply summed. Note that, although individual

a)Author to whom correspondence should be addressed. Electronic mail: kbsmith@nps.edu. Currently serving as Visiting Professor at the Naval Undersea Warfare Center, Division Newport.

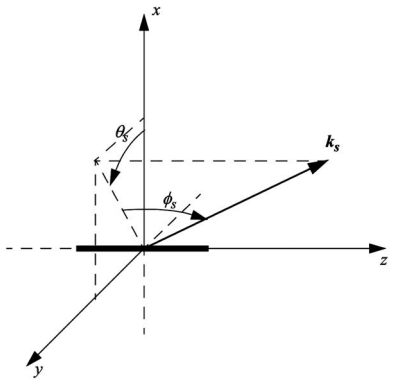


FIG. 1. Coordinate system used in derivations.

elements can be steered in θ and ϕ, the array steering is only accomplished by relative phases along the array axis, which only corresponds to steering in ϕ. (Any linear array of omnidirectional sensors can only delineate along a single axis.)

The individual weights can then be steered independently (so each element "looks" in a different direction, although the array can only steer a plane wave coherently in a single direction), or a common vector steering can be applied across the array. Assuming the latter is chosen for simplicity, the relative magnitude of the weighting must also be defined. For example, one approach would be to use *equal* weighting, where $w_{xn}=w_n\cos\theta_s\cos\phi_s$, $w_{yn}=w_n\sin\theta_s\cos\phi_s$, $w_{zn}=w_n\sin\phi_s$, and $w_{pn}=w_n$. (Note that the trigonometric functions operating on ϕ_s are switched from that listed in Ref. 1 since we will assume the array is along the z axis and ϕ is then measured relative to broadside.) Such weighting of the components produces the standard cardioid pattern. Additional directivity gain can be achieved by using *optimal* weighting[1] of individual sensors, defined by $w_{xn}=3w_n\cos\theta_s\cos\phi_s$, $w_{yn}=3w_n\sin\theta_s\cos\phi_s$, $w_{zn}=3w_n\sin\phi_s$, and $w_{pn}=w_n$. However, since our focus is on the reduction of influence of interfering sources from the backside of the beam shape, we will maintain the deep null of the standard cardioid shape by using equal weighting. The array element weights, w_n, are then defined simply in terms of a standard window function (e.g., Hanning).

Linear array beamforming is based on matching the relative signal phases along the array as measured by each element. The output of the beamformer is maximized when the steering direction, (θ_s,ϕ_s), matches the direction of the incoming plane wave. This can be further illustrated by rewriting Eq. (2) as

$$B^{(pv)}(\theta_s,\phi_s)=\left|\sum_n (w_{xn}V_{xn}+w_{yn}V_{yn}+w_{zn}V_{zn}+w_{pn}V_{pn})e^{i(\mathbf{k}\cdot\mathbf{r}_n-\mathbf{k}_s\cdot\mathbf{r}_n)}\right|^2. \quad (3)$$

Since the response of each velocity component to a single, incident plane wave acts like a dipole, then $V_x=V\cos\theta\cos\phi$, $V_y=V\sin\theta\cos\phi$, and $V_z=V\sin\phi$. When $(\theta,\phi)=(\theta_s,\phi_s)$, Eq. (3) shows that the phase mismatch term vanishes (i.e., $\mathbf{k}=\mathbf{k}_s$), and the array is properly steered. For the standard approach to vector sensor beamforming with cardioids, the weighted sum in parentheses simplifies to $\Sigma_n w_n V_n(\cos^2\theta_s\cos^2\phi_s+\sin^2\theta_s\cos^2\phi_s+\sin^2\phi_s+1)=2\Sigma_n w_n V_n$. In arriving at this simplification, use has been made of the relationship $V_{pn}=V_n$, which follows from Eq. (1). This twofold increase in the signal amplitude, as compared to a single hydrophone response, provides a theoretical 3 dB enhancement on a single vector sensor. (Note that optimal weighting provides 6 dB theoretical gain.)[1]

In order to consider other vector sensor steering patterns, let us rewrite the steering of each element. For the cardioid used above, the steering on the nth element is defined by

$$\begin{aligned} b^{(n)}_{\mathbf{cardioid}}(\theta,\phi)&=(w_{xn}V_{xn}+w_{yn}V_{yn}+w_{zn}V_{zn}+w_{pn}V_{pn})e^{i\mathbf{k}\cdot\mathbf{r}_n}\\ &=w_nV_ne^{i\mathbf{k}\cdot\mathbf{r}_n}(\cos\theta\cos\phi\cos\theta_s\cos\phi_s\\ &\quad+\sin\theta\cos\phi\sin\theta_s\cos\phi_s+\sin\phi\sin\phi_s+1)\\ &=w_nV_ne^{i\mathbf{k}\cdot\mathbf{r}_n}(\cos\phi\cos\phi_s\cos(\theta-\theta_s)\\ &\quad+\sin\phi\sin\phi_s+1). \end{aligned} \quad (4)$$

Note how this element steering maintains the relative phase along the array (defined by $\mathbf{k}\cdot\mathbf{r}_n$) and has a maximum when $(\theta,\phi)=(\theta_s,\phi_s)$.

As an alternative, we now consider the beampattern defined by first creating a rescaled hippopede, or "horse's fetter" (defined by the Greek mathematician/astronomer Eudoxus of Cnidos around the 4th century BC), given by

$$\begin{aligned} &\left[\frac{(w_{xn}v_{xn}+w_{yn}v_{yn}+w_{zn}v_{zn})}{(w_nv_{pn})}\right]\\ &\quad\times\left[\frac{(w_{xn}v_{xn}+w_{yn}v_{yn}+w_{zn}v_{zn})^*}{(w_nv^*_{pn})}\right]\\ &\quad=(\cos\phi\cos\phi_s\cos(\theta-\theta_s)+\sin\phi\sin\phi_s)^2. \end{aligned} \quad (5)$$

More general definitions of hippopedes may be found on various websites.[3,4] The general form in polar coordinates, $r^2=4b(a-b\sin^2\theta)$, is obtained when we set $\phi=\phi_s=\theta_s=0$. Note that we are defining the parameters $a=b=0.5$ in this work.

Multiplying the expression in Eq. (5) by the cardioid produces what we shall refer to as a second order "hippioid," defined by

$$\begin{aligned} b^{(n)}_{\mathbf{hippioid\_2}}(\theta,\phi)&=[(w_{xn}v_{xn}+w_{yn}v_{yn}+w_{zn}v_{zn}+w_{pn}v_{pn})]\\ &\quad\times\left[\frac{(w_{xn}v_{xn}+w_{yn}v_{yn}+w_{zn}v_{zn})}{(w_nv_{pn})}\right]\\ &\quad\times\left[\frac{(w_{xn}v_{xn}+w_{yn}v_{yn}+w_{zn}v_{zn})^*}{(w_nv^*_{pn})}\right]\\ &=w_nV_ne^{i\mathbf{k}\cdot\mathbf{r}_n}(\cos\phi\cos\phi_s\cos(\theta-\theta_s)\\ &\quad+\sin\phi\sin\phi_s+1)\\ &\quad\times(\cos\phi\cos\phi_s\cos(\theta-\theta_s)\\ &\quad+\sin\phi\sin\phi_s)^2. \end{aligned} \quad (6)$$

The cardioid of Eq. (4) and the hippioid of Eq. (6) have plane wave responses as depicted in Fig. 2. Note in both cases, the incoming plane wave is assumed to be incident on the $(\theta,\phi)=(0,0)$ axis. Both figures have been normalized to provide a peak response of unity.

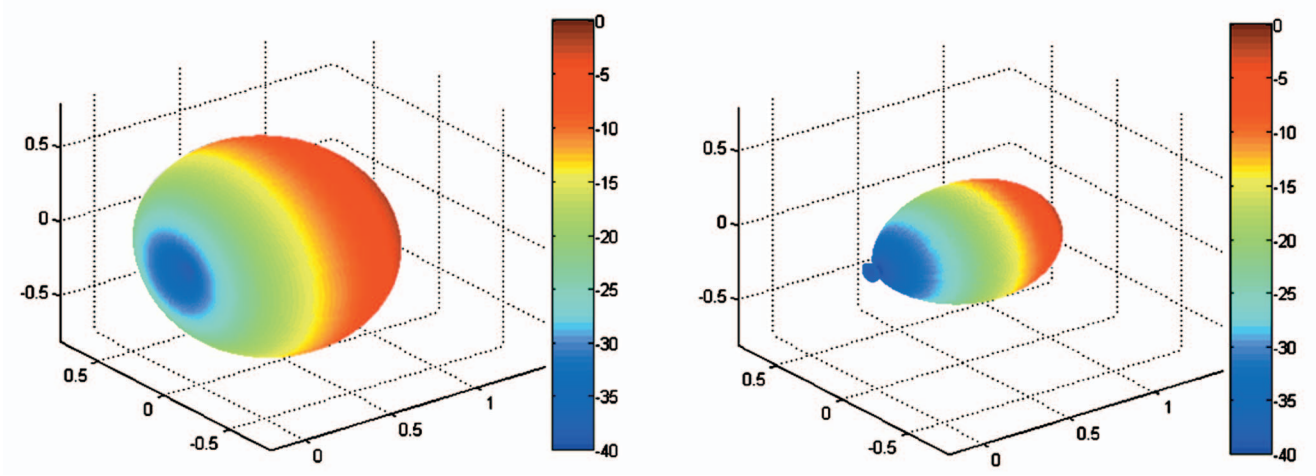


FIG. 2. Cardioid beam pattern (left) and second order hippioid beam pattern (right) for incoming plane wave incident along $(\theta,\phi)=(0,0)$. Beam patterns are presented on a linear scale and have been normalized to unity in the peak direction. The axes represent the linear scale response of the beam. The color scale maps the associated dB units onto the shape.

This beampattern is one of a class of beamshapes defined generically as

$$b^{(n)}_{\textbf{hippioid\_m}}(\theta,\phi) = w_n V_n e^{i\mathbf{k}\cdot\mathbf{r}_n}(\cos\phi\cos\phi_s\cos(\theta-\theta_s) + \sin\phi\sin\phi_s + 1) \times (\cos\phi\cos\phi_s\cos(\theta-\theta_s) + \sin\phi\sin\phi_s)^m, \quad (7)$$

where the integer factor m determines the final form of the shape. When $m=0$ (zeroth order hippioid), we retain the standard cardioid shape. For $m=2$ (second order hippioid), we obtain the shape just defined. If we examine values of $m=0\ldots5$, and compare the beampatterns along the $\phi=0$ plane, we obtain the curves plotted in Fig. 3.

In Fig. 3, we note that the great disadvantage of the cardioid shape is its broad main beam with a 3 dB down beamwidth of ~130°. When $m=1$, this beamwidth reduces to ~76°, and when $m=2$ it reduces further to ~60°. The directivity of the beamshapes can be shown to satisfy

$$D(\theta) = 20\log\left|\frac{1}{2}(\cos\theta+1)\cos^m\theta\right|, \quad (8)$$

which provides an expression for the −3 dB beamwidth of

$$\theta(-3\text{ dB}) = 2\cos^{-1}(10^{-3/20m}). \quad (9)$$

This expression leads to values consistent with the estimates just given. Furthermore, all beam shapes presented still provide a strong, narrow null of the same width (~24° at 40 dB down from the main peak). Thus, the reduction in the main beam width does not require a degradation of the null, which may be useful in other processing schemes.

On the other hand, the expressions for the higher-order beamshapes require a division by the data and, therefore, will introduce some processing noise. Furthermore, the equations for the beam shapes provided in Eqs. (6) and (7) are no longer based on a linear summation of weighted components, but rather some nonlinear product with cross terms. (This should be distinguished from nonlinear, "adaptive" processing whereby the measured data affects the definition of the weighting vectors.) Still, this only affects the local beamshape at the vector sensor, and does not impact the linear array beamforming yet to be computed.

It is perhaps worth noting that even a single vector sensor may be considered a superdirective array inasmuch as it does provide a degree of angular resolution at a single point in space. Furthermore, the nonlinear hippioids are analogous in many ways to multiplicative arrays.[5] Such arrays are known to suffer performance in the presence of low signal-to-noise ratio, where multiplicative noise destroys the directional gain. An analysis on the effects of noise on the hippioid response will not be performed here, but is being considered in future work.

Preliminary analysis does show that hippioids and cardioids also both suffer performance degradations in the presence of correlated interferers (e.g., multipaths). These effects can be mitigated with arrays of sensors, but directionality

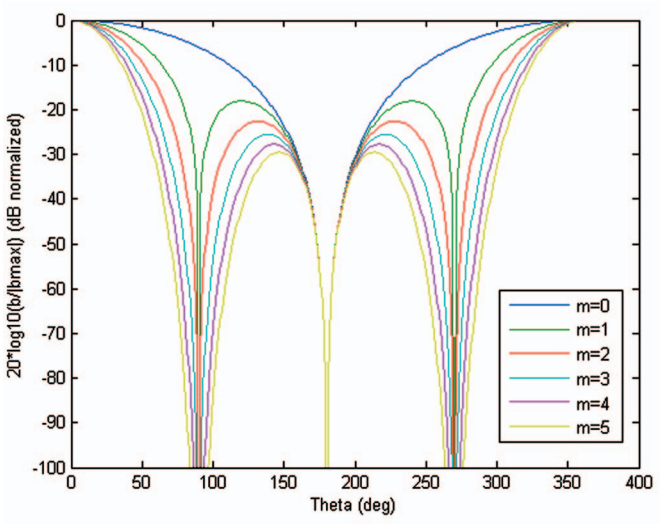


FIG. 3. Beam pattern along symmetrical axis for zeroth order hippioid (standard cardioid) and first five nonzero orders. Responses are presented on a dB scale, normalized to 0 dB in the peak direction.

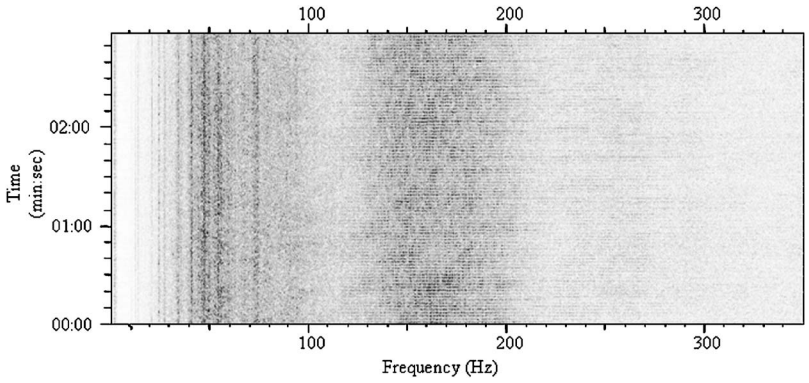


FIG. 4. LOFAR gram, a spectral decomposition with time versus frequency of data recorded by the omnidirectional sensor. Sound pressure intensities are measured at the sensor and scaled to fit a color map of 256 gray scales.

then relies on array aperture rather than individual sensor performance. However, in many cases, interfering sources fall into different frequency bins or can be separated using more sophisticated techniques than are described here. This is particularly true in the horizontal plane where we do not need to concern ourselves so much with multipaths due to waveguide propagation.

III. DATA ANALYSIS

A. Vector steering of DIFAR buoy data

In previous work,[6] van Leijen employed a two-dimensional version of the second order hippioid (Eq. (6) with $\phi=\phi_s=0$) on DIFAR buoy data, and compared that with results from a standard cardioid. Some results from this analysis are depicted in Figs. 4–7. In Fig. 4, the pressure-only, omni-directional response of the buoy is presented as a function of frequency and time. There appear to be some tones present around 50 Hz or so, with some broadband, noisy signals covering the band 130–200 Hz.

In Fig. 5, the same data are displayed but now each data point is color coded to represent the arrival angle in the horizontal (azimuthal) plane. The value of the angle is computed by taking ratios of components of intensity,[7] then rounded to a resolution of 30°. The tonal energy appears to be coming primarily from southerly directions, whereas the broadband, diffuse energy seems to come mostly from the west. The pie chart next to the figure displays the color segmentation of directions.

Figure 6 shows the same data after steering the vector components to the north-northeast, along the direction 30°, with the standard cardioid shape. Note that this points the cardioid null in the direction 210°, towards the south-southwest, where most of the tonal energy seemed to be generated. Perhaps surprisingly, this does not remove the observed tonal energy, but actually helps to enhance the observation of several tonals, some that were barely observable before steering the null to the south-southwest. In other words, there are additional tones of interest in the signal that could not be discriminated without some form of spatial (angular) resolution. The pie chart next to the figure shows a cardioid shape superimposed to give some sense of what directions are included and excluded from this processing. It is also worth noting that, although the direction to the broadband, diffuse energy is nearly perpendicular to the direction of steering, the broad sidelobes of the cardioid do little to diminish the influence of this portion of the signal.

Figure 7 then shows the same data after steering towards 30° with the second order hippioid. Not only are more tones better resolved (allowing for better classification and/or tracking algorithms to be applied), but the narrower beam of the hippioid is able to significantly reduce the influence of the broadband energy from the west. There are even some transient events around 180, 270, and 300 Hz that were unobservable in the previous plots. Thus, the improved resolution of the hippioid appears to provide significant performance improvements over the basic cardioid pattern.

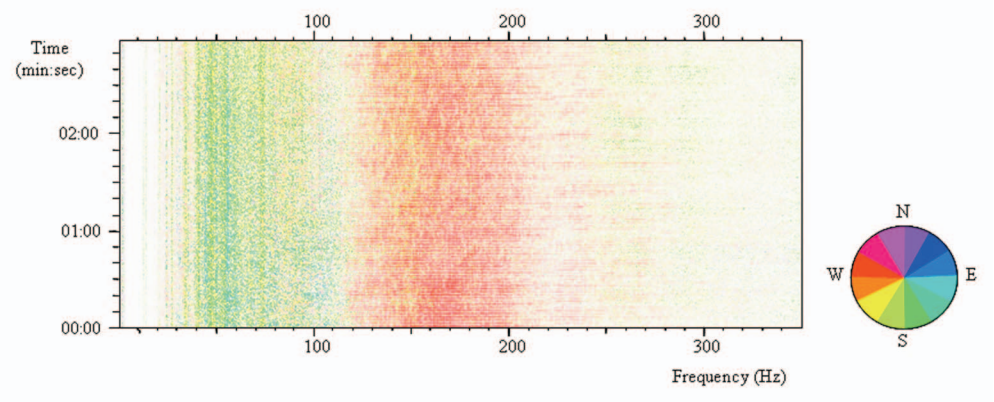


FIG. 5. DIFAR gram, colors added to LOFAR gram display direction of individual signals over 12 pie slices (each slice of width 30°). Directions are determined by ratio of components of acoustic intensity from directional sensors.

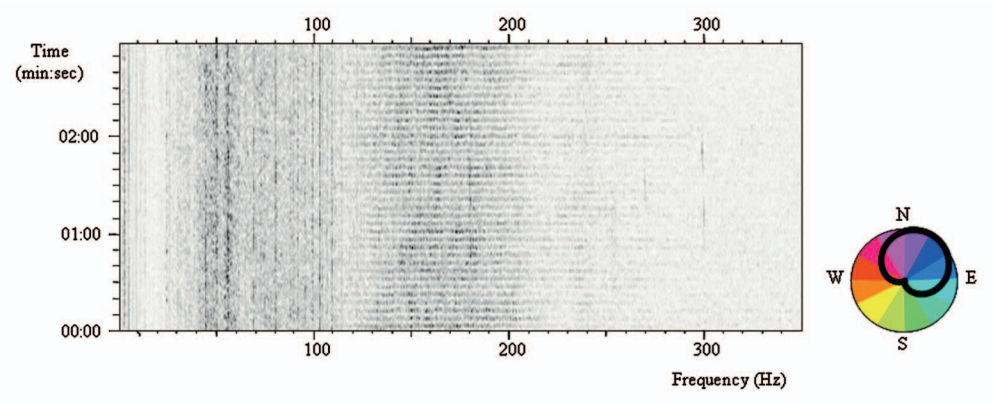


FIG. 6. DIFAR gram from directional sensors with cardioid steered to 30°. Note the null is then steered to 210°, greatly reducing the influence of the "yellow" portion of the directional pie.

B. Vector sensor array beamforming for single plane wave

In this section, we shall compare the predicted performance of an array of vector sensors when cardioids or second order hippioids are employed on each sensor element. The signals measured on each sensor are defined analytically in terms of the response from a single, incident plane wave. Specifically, we shall assume a plane wave of frequency f is incident upon an array of vector sensors, where the spacing between elements is half a wavelength, i.e., $d=\lambda/2=c/(2f)$ and $c=1500$ m/s. The direction of incidence is $(\theta,\phi)=(334°,0°)$, and we shall only examine results in the $\phi=0$ plane.

In Figs. 8 and 9, the response of the beamformer as a function of aperture length is examined for both the standard cardioid and the second order hippioid. These results are compared to that of an array of pressure-only sensors. In each case, the data are plotted on a dB scale normalized with the peak at 0 and the minimum value forced to -100 dB. This provides the upper 100 dB dynamic range of the beam pattern.

Figure 8 displays the results for a single, incident plane wave. In the first plot (upper left, Fig. 8), we observe the typical patterns associated with a single sensor. Specifically, the pressure-only curve (blue) is omnidirectional, the cardioid curve (red) shows its typical shape, and the second order hippioid (green) is consistent with its expected shape. Note the typically 20 dB or more reduction in sidelobe levels from the hippioid as compared to the cardioid.

Furthermore, we also note that as the aperture length is increased, the cardioid processing does not entirely eliminate the right-left ambiguity. An absolute right-left discrimination is only obtained at broadside, where the cardioid peak and null are both aligned with the right-left ambiguous directions. (The use of other linear, noncardioid shapes can improve this performance by varying the relative scale of the pressure term in Eq. (3), thereby adjusting the relative orientation between the null and the peak. The peak resolution remains poor, however.)

For 16 or more elements (lower plots), we see that this is where the use of the hippioid has real advantages over the cardioid. The cardioid processing is able to reduce the ambiguous direction by ~14 dB, whereas the hippioid is able to reduce this signal by ~23 dB (an additional ~9 dB of ambiguity discrimination). For 16 or more elements, we also observe that the angular resolution is dictated by the array aperture. In other words, although there is right-left ambiguity discrimination, the vector sensors do not improve the resolution of the true signal arrival angle.

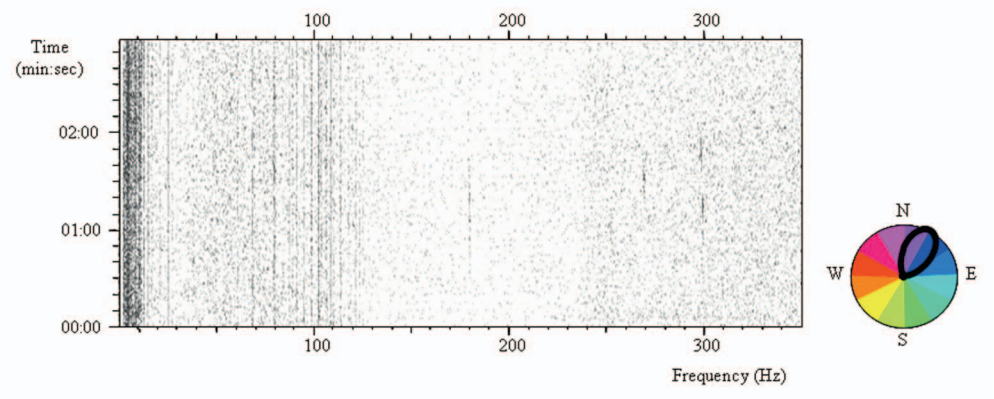


FIG. 7. DIFAR gram from directional sensors with second order hippioid steered to 30°. The broader null of the hippioid greatly reduces the influence of the yellow, "red," and "blue-green" portions of the directional pie. This then reveals other signals of potential interest that were originally obscured by the broad sidelobe of the cardioid.

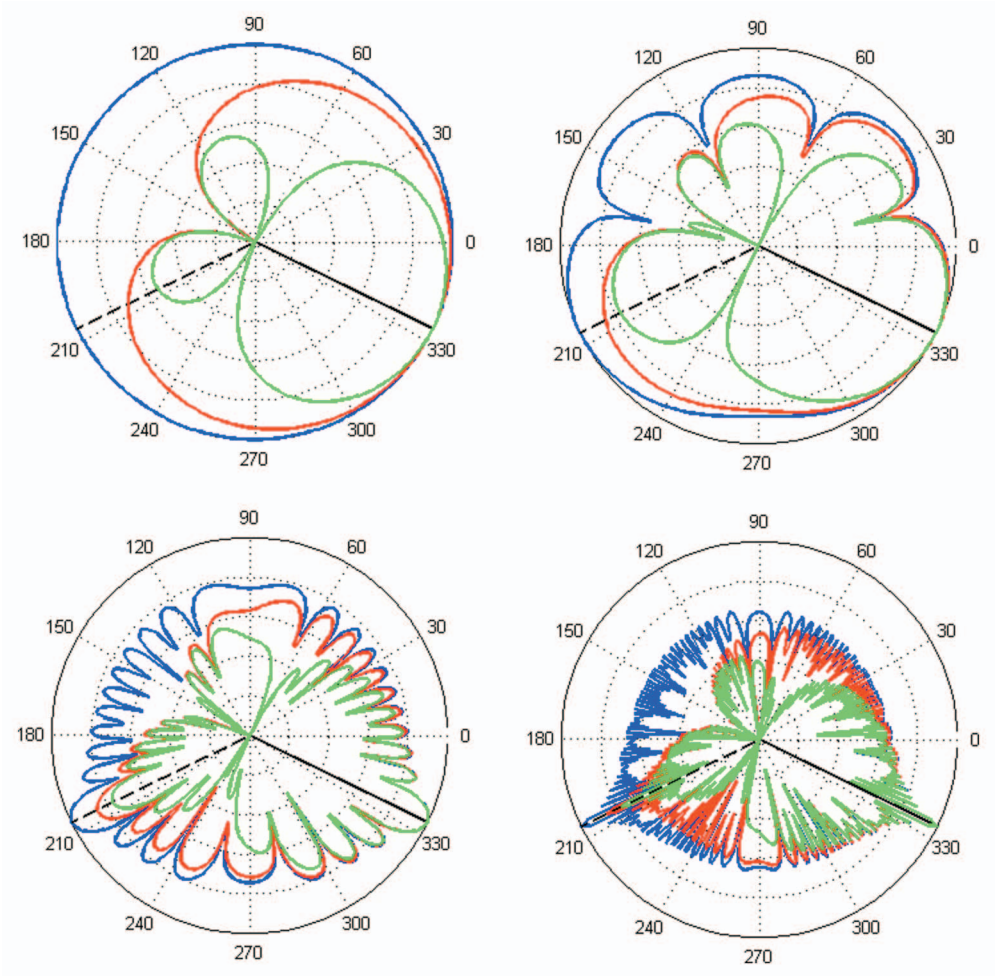


FIG. 8. Array beam patterns for single, plane wave incident at 334° with varying aperture: single element (upper left), four elements (upper right), 16 elements (lower left), and 64 elements (lower right). Blue curves represent solution for pressure-only sensors, red curves correspond to cardioid steering of vector sensors, and green curves are solutions for second order hippioid steering. Solid black lines indicate arrival directions, dashed lines indicate right-left ambiguous directions. Beam patterns normalized to peak of 0 dB with axes contours of 20 dB.

C. Vector sensor array beamforming for multiple plane waves

In Fig. 9, a similar calculation is performed but now we introduce three identical signals (same continuous wave frequency) along different incident directions. Specifically, $(\theta_1, \phi_1) = (334°, 0°)$, $(\theta_2, \phi_2) = (71°, 0°)$, and $(\theta_3, \phi_3) = (157.5°, 0°)$, with relative weighting (1:1:0.5). As before, a single, pressure-only sensor will produce an omni-directional beam pattern (blue curve, upper left). As previously demonstrated by Smith, *et al.*[2] single vector sensors cannot resolve multiple, interfering (coherent) signals to produce direction estimates. This is seen to be the case for both the cardioid and hippioid steering patterns.

The beam patterns obtained from an array of four elements are also unable to clearly distinguish arrival angle information. A larger aperture is required, such as that created by 16 and 64 elements (lower plots). In both cases, the pressure-only sensors have no right-left ambiguity discrimination, as expected. For the first signal, $\theta_1 = 334°$, the cardioid is found to provide ~13 dB of discrimination. For the second arrival at $\theta_2 = 71°$, the cardioid is only able to provide about 1 dB of discrimination. Such problems become apparent as we approach end fire, where the broad frontal lobe of the cardioid does little to help distinguish between left and right directions. For the third arrival at $\theta_3 = 157.5°$, the cardioid gives ~14 dB of discrimination.

For such cases with multiple, interfering arrivals, the cardioid processing (because of its linear form) provides the same relative output in the true directions of incidence. In other words, after normalizing the beam pattern responses, the peaks of the cardioid beamformer in the direction of incidence match the peaks of the pressure-only beamformer. The same is not found to be the case with the hippioids, presumably due to the nonlinear form of the weightings applied. For the first and third arrivals, the true-to-ambiguous ratio provides little or no gains. Only for the second arrival is there a nearly 15 dB gain in discrimination.

IV. SUMMARY

In this paper, a new set of vector sensor steering patterns, referred to as hippioids, were introduced. For single arrivals (within a frequency bin), the response of the second order hippioid was found to greatly reduce the sidelobe influence of the standard cardioid, and improve the angular resolution significantly.

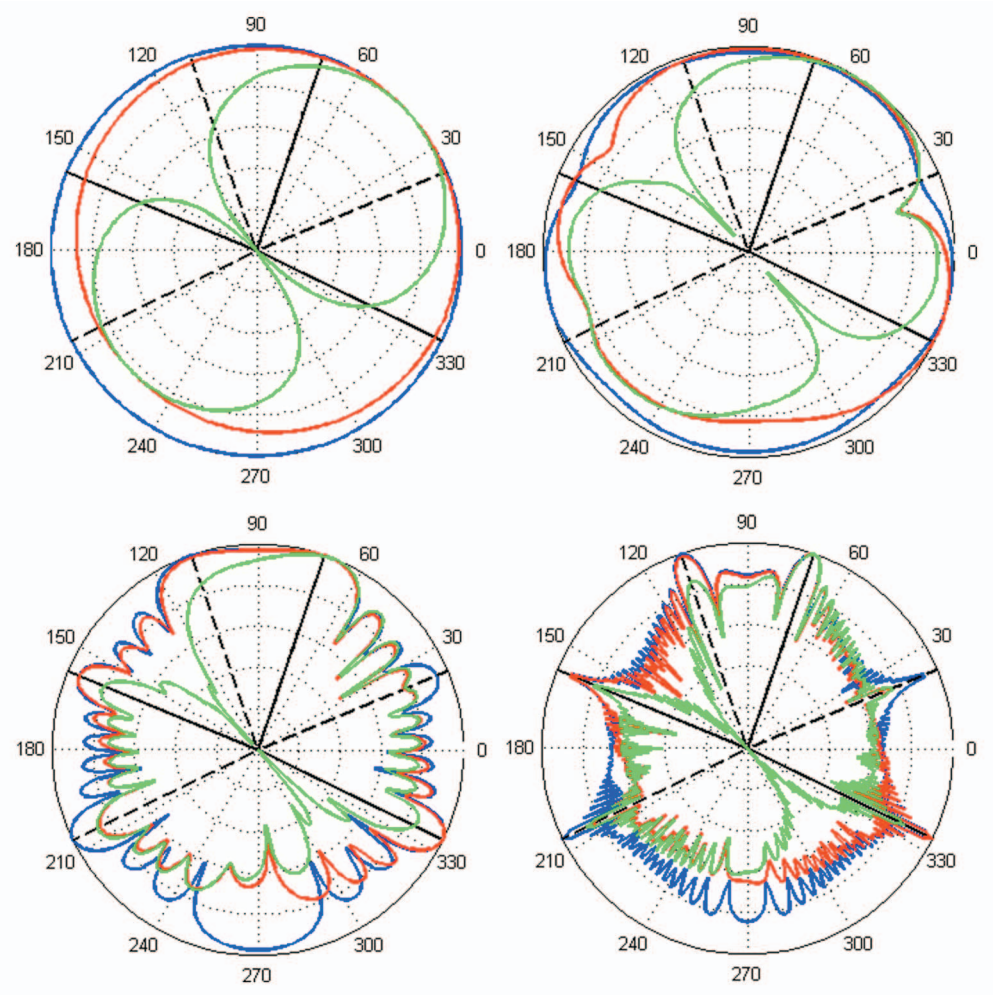


FIG. 9. Array beam patterns for three plane waves incident at 334°, 71°, and 157.5° (relative amplitudes 1:1:0.5) with varying aperture: single element (upper left), four elements (upper right), 16 elements (lower left), and 64 elements (lower right). Blue curves correspond to pressure-only sensors, red curves correspond to cardioid steering of vector sensors, and green curves are solutions for second order hippioid steering. Solid lines indicate arrival directions, dashed lines indicate right-left ambiguous directions. Beam patterns normalized to peak of 0 dB with axes contours of 20 dB.

As a practical application, a comparison was made between DIFARgrams employing cardioid steering and hippioid (second order) steering. The new steering pattern was found to substantially improve performance by reducing the broad sidelobes of the cardioid. This allowed previously unobserved tonals and transients to be clearly identified.

A brief analysis was performed on the influence of these new steering patterns on array beampatterns of varying aperture. Both single arrival and multiple (in-band, interfering) arrivals were considered. For the case of a single, plane-wave arrival, the performance gains of the hippioid were obvious. Away from broadside, there was significant additional right-left ambiguity discrimination, upwards of 10 dB or more.

When multiple (interfering) arrivals were considered, however, many of the improvements of the hippioid seemed to be reduced. It was observed that the response in the direction of energy arrival did not remain at the same level, as it did for the pressure-only processing and the vector sensor processing with cardioids. This is presumably due to the nonlinear nature of the hippioid processing.

Future analysis will examine this presumption more closely. And while such new steering patterns may still be useful for their simplicity, a more optimal set of steering patterns will be defined by adaptive processing. Such processing will also be part of future studies by the authors.

ACKNOWLEDGMENTS

The authors wish to acknowledge the helpful discussions with Dr. Denhua Huang at the Naval Undersea Warfare Center, Division Newport. This work was supported by the US Office of Naval Research (Code 321OA), the Naval Postgraduate School's Sabbatical Program, and the Netherlands Defence Academy.

[1]B. A. Cray and A. H. Nuttall, "Directivity factors for linear arrays of velocity sensors," J. Acoust. Soc. Am. **110**, 324–331 (2001).
[2]K. B. Smith, R. T. Richards, and P. Duckett, "Comparative beamforming studies employing acoustic vector sensor data," J. Acoust. Soc. Am. **119**, 3446 (2006).
[3]http://www.2dcurves.com/quartic/quartich.html (last viewed March 8, 2007).
[4]http://www.mathworld.wolfram.com/Hippopede.html (last viewed March

8, 2007).
[5]R. J. Urick, *Principles of Underwater Sound* (McGraw-Hill, New York), pp. 57–63 (1975).
[6]A. V. van Leijen, "DIFAR and beamforming," Coda Report (2001).
[7]C. R. Greene, Jr., M. W. McLennan, R. G. Norman, T. L. McDonald, R. S. Jakubczak, and W. J. Richardson, "Directional frequency and recording (DIFAR) sensors in seafloor recorders to locate calling bowhead whales during their fall migration," J. Acoust. Soc. Am. **116**, 799–813 (2004).

Overexposure effects of a 1-kHz tone on the distortion product otoacoustic emission in humans

Karen Reuter,[a] Rodrigo Ordoñez, and Dorte Hammershøi
Acoustics, Department of Electronic Systems, Aalborg University, Denmark

(Received 18 August 2006; revised 2 May 2007; accepted 2 May 2007)

The effects of overexposure on the properties of distortion product otoacoustic emissions (DPOAEs) are investigated. In total, 39 normal-hearing humans were monaurally exposed to a 1-kHz tone lasting for 3 min at an equivalent threshold sound-pressure level of 105.5 dB. The effects of overexposure were studied in two experiments (1) on the broadband DPOAE and (2) on the DPOAE fine structure, measured using a higher frequency resolution in a narrower frequency range. The obtained DPOAE shifts were compared to temporary threshold shift (TTS) obtained after a similar exposure. Similarities between DPOAE shifts and TTS were found in the affected frequency range and the time course of recovery. The amount of TTS was higher in the early recovery time (1–4-min postexposure), but similar to the DPOAE shift (even in absolute terms) at later recovery times (5–20-min postexposure). The DPOAE fine structure was not systematically changed after the exposure. © *2007 Acoustical Society of America.* [DOI: 10.1121/1.2743163]

PACS number(s): 43.64.Jb, 43.66.Ed [BLM] Pages: 378–386

I. INTRODUCTION

Overexposure to high-intensity sounds can cause changes in the state of hearing, i.e., as it can be recognized as reduced sensitivity to low-level sounds. The degree of the effects depends on the severity of the exposure, the time between the exposure and assessment of hearing, as well as on the susceptibility of the exposed subject. The changes are reversible, as long as the level and duration of the overexposure do not become excessive. During exposure to high-intensity sounds, changes in the metabolic process of the inner ear occur (Kemp, 1986; Kirk *et al.*, 1997). Particularly vulnerable to overexposure are the outer hair cells (OHCs), which are responsible for the active cochlear amplification process.

Reduced sensitivity of the hearing is typically monitored by measuring behavioral pure-tone audiometry before and after overexposures. The difference between the absolute thresholds prior to the exposure and the absolute thresholds after the exposure is called temporary threshold shift (TTS). One characteristic of TTS is that the maximum TTS for a narrow-band exposure (centered at frequencies above 250 Hz) is found approximately 1/2 octave above the exposure frequency (McFadden and Plattsmier, 1983; Melnick, 1991; Ordoñez, 2005). This 1/2-*octave-shift* is generally believed to relate to the difference in excitation pattern during the high-level situation (exposure) and the low-level situation (threshold determination), as the maximum excitation point for high stimulus levels shifts toward the base of the basilar membrane (de Boer and Nuttall, 1997, 2000; Johnstone *et al.*, 1986; Rhode and Recio, 2000; Ruggero *et al.*, 1997). Another characteristic of TTS is a peak on the recovery curve around 2 min, apparent after short intense exposures to low-frequency stimuli <2 kHz (Botsford, 1971; Hirsh and Ward, 1952; Ordoñez, 2005). The reason for this 2-min *bounce* is not clear. It may reflect the interaction between different recovery mechanisms (Patuzzi, 1998) as well as temporary changes in the sensitivity of the active process related to the OHCs (Kemp, 1986).

Temporary changes in the state of hearing caused by overexposure can be seen not only in the TTS, but have also shown to alter otoacoustic emissions (OAEs) (Delb *et al.*, 1999; Engdahl and Kemp, 1996; Kværner *et al.*, 1995; Lonsbury-Martin *et al.*, 1987; Martin *et al.*, 1987; Schmiedt, 1986; Sutton *et al.*, 1994; Vinck *et al.*, 1999). OAEs are sounds generated as a by-product of the active inner ear and can be measured as low-level sounds in the ear canal. Since the active process in the inner ear involves active mechanical feedback from the OHCs, they are believed to be a sensitive measure to detect early noise-induced hearing-loss. Some studies confirm this hypothesis, i.e., Attias *et al.* (1995); Desai *et al.* (1999); LePage and Murray (1993); LePage *et al.* (1993) found that noise-exposed subjects have lower OAEs than subjects with similar hearing thresholds but no significant noise exposure. Desai *et al.* (1999); LePage and Murray (1993, 1998); LePage *et al.* (1993); Lucertini *et al.* (2002); Laplsley Miller *et al.* (2006) found a decrease of OAE amplitudes in noise-exposed subjects, while no change in the behavioral threshold could be seen.

A promising type of emission is distortion product otoacoustic emission (DPOAE), which explores properties of the cochlea in a frequency specific manner. Controlled studies of temporary effects on the OAEs may help in understanding the relation between OAE and hearing threshold, and may be used to find OAE characteristics that indicate early changes in the state of hearing. A few publications exist which study overexposure effects on DPOAEs in humans (Engdahl and Kemp, 1996; Schmiedt, 1986; Sutton *et al.*, 1994). Generally, similarities of OAE shifts to TTS could be observed. The highest DPOAE shift has been found approximately half an octave above the frequency of the exposure for humans

[a]Electronic mail: kr@es.aau.dk

 0001-4966/2007/122(1)/378/9/$23.00

(Engdahl and Kemp, 1996) and rabbits (Martin *et al.*, 1987). Also similarities to the time course of recovery could be observed (Sutton *et al.*, 1994); the DPOAE level was improved before it deteriorated, highly reminiscent of the 2-min bounce.

DPOAE is the response of the inner ear to two pure-tone stimuli (the primaries f_1 and f_2). They cause a series of distortion products, the most prominent being at the frequency $2f_1-f_2$. It is widely accepted that DPOAE are generated by at least two sources, as reviewed by Shaffer *et al.* (2003). According to the two-source model [e.g., Dhar *et al.* (2002); Mauermann *et al.* (1999a)], the DPOAE is composed of two components: (1) the distortion component generated at the f_2 place and (2) the reflection component generated at the $2f_1-f_2$ place. The DPOAE measured in the ear canal is a vector sum of contributions arising from the f_2 and $2f_1-f_2$ region. The two components interfere with each other constructively and destructively and, as a consequence, large level differences from one frequency to another exist in the measured DPOAE, rendering a characteristic fine structure. Recently, some investigations described a negative correlation between hearing loss and distinctiveness of DPOAE fine structure (Engdahl and Kemp, 1996; Mauermann *et al.*, 1999b; Rao *et al.*, 1996). Engdahl and Kemp (1996) demonstrated changes in the DPOAE fine structure following exposure to a narrowband noise centered at 2000 Hz. There was a significant decrease in the maximum to minimum ratio of the fine structure starting at 2-min postexposure that tended to recover over the 32-min postexposure. These results are based on data of two subjects. Reuter and Hammershøi (2006) determined DPOAE fine structures in 50 normal-hearing humans and found a high intersubject variability in the characteristics of DPOAE fine structures, i.e., some subjects had fine structures with ripple heights of up to 32 dB, whereas other subjects had only little fine structure. Intersubject differences in the overexposure effects are therefore also anticipated.

A. Scope of this paper

The present study investigates the properties of DPOAE following overexposure for more human subjects with the purpose of generalizing on previous findings. Which DPOAE frequency range is most affected is investigated by measuring DPOAE over a relatively wide frequency range before and after exposure. DPOAE shifts are compared to TTS data obtained from Ordoñez (2005). In an additional experiment DPOAE fine structures are obtained before and after the same exposure.

II. MATERIALS AND METHODS

A. Subjects

In total, 39 human subjects participated in the experiments of this study. All subjects were aged between 20 and 29 years and had hearing levels better than 25 dB HL in 1/2-octave intervals from 250 Hz to 4 kHz. DPOAEs were recorded in one ear only, which was randomly chosen. Only subjects without an abnormal noise exposure history 24 h prior to the experiment were included.

B. General measurement procedures

During the entire test the subjects were seated in a double-walled, sound-isolated audiometry chamber at Acoustics, Aalborg University. The room complies with the background noise requirements stated in ISO 8253-1:1989. The hearing of the subjects was screened with a custom-built audiometer (Lydolf, 1999), using Sennheiser HDA 200 headphones and the ascending method complying with the norms for automatic audiometry (ISO 8253-1:1989). The system was calibrated using the B&K type 4153 artificial ear complying with IEC 60318-3:1998.

The $2f_1-f_2$ DPOAEs were measured with the ILO96 Research system from Otodynamics, using $f_2/f_1=1.22$ and fixed primary levels of $L_1/L_2=65/45$ dB. The choice of primary levels is based on several considerations. Liebel *et al.* (1996) reported that the detection of overexposure effects on the cochlea using DPOAE with $L_1/L_2=70/70$ dB was inadequate and resulted in low sensitivity. At stimulation levels of $L_1/L_2=70/70$ dB the OHCs contribute only very little to the displacement of the basilar membrane (Johnstone *et al.*, 1986); therefore, changes of DPOAE after exposure cannot be expected for high-level primaries. Several researchers have reported that maximum DPOAE levels can be obtained when lowering L_2 compared to L_1 (Gaskill and Brown, 1990; Hauser and Probst, 1991; Whitehead *et al.*, 1995a, b). The largest changes in amplitude after acoustic trauma could be detected when the amplitudes of the primary tones were considerably different and relatively low. Sutton *et al.* (1994) found an improved sensitivity regarding acoustic trauma for L_1-$L_2=25$ dB; the best results were obtained for $L_1/L_2=55/30$ dB. Prior to the present study a pilot study was performed, in which different primary combinations were tested. The level combination of $L_1/L_2=55/30$ dB resulted in very low-level DPOAEs, and therefore it was decided to use higher-level primaries of $L_1/L_2=65/45$ dB, for which moderate pre-exposure DPOAE levels could be measured in normal-hearing subjects. The level combination of 65/45 dB was chosen based on results from Whitehead *et al.* (1995b), who measured DPOAEs for various primary level combinations at different frequencies (1.39, 2.79, and 5.57 kHz). The primary level combination 65/45 dB showed relatively high-level DPOAEs for the tested frequencies.

The probe was generally not removed between the postexposure DPOAE measurements. The probe fit was assessed using the ear canal volume estimation of the ILO96. In a few cases changes in the ear canal volume were observed, and the probe was reinserted.

C. Exposure

A 1-kHz tone was presented monaurally at an equivalent threshold sound-pressure level (ETSPL) of 105.5 dB for 3 min. The exposure was delivered by Sennheiser HDA 200 headphones at 100 dB above threshold (according to ISO 389-8:2004), corresponding to a free-field related equivalent continuous A-weighted sound-pressure level of $L_{FF,H,Aeq}=108.0$ dB (according to ISO 11904-1:2002). The exposure corresponds to an exposure level normalized to an 8-h working day of $L_{Ex,8\ h}=86$ dB (according to ISO 1999:1990). All

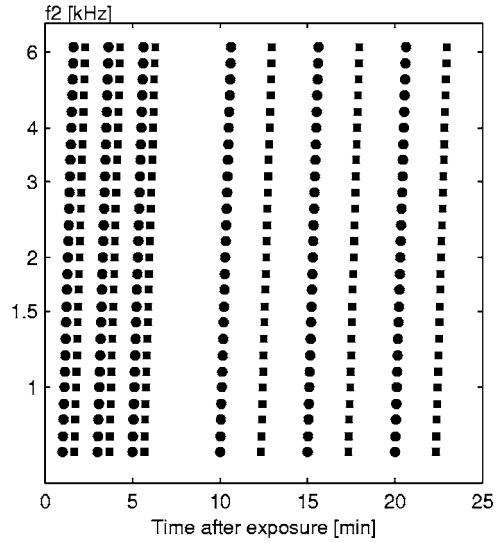


FIG. 1. Distribution of postexposure DPOAE measurements as a function of time and primary frequency f_2, group 1 subjects (circles), and group 2 subjects (squares).

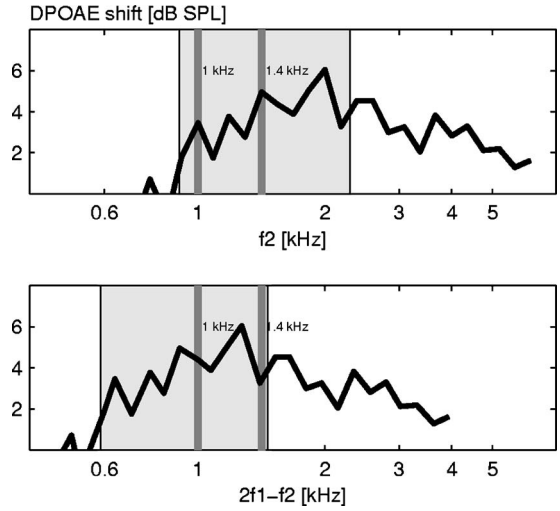


FIG. 2. Frequency range of DPOAE fine structure (gray shaded area), as a function of f_2 (top panel) and $2f_1-f_2$ (bottom panel). Black line: DPOAE shift obtained from the broadband DPOAE experiment (first determination time, group 1). The gray bars indicate the exposure frequency of 1 kHz and the corresponding 1/2 octave shift.

subjects listened shortly to the exposure before the experiment and were instructed to remove the headphones if they felt any discomfort; none did. Some subjects could hear a sound right after the end of the exposure, but none heard phantom sounds at the end of the experiment.

D. Broadband DPOAE

A total of 32 subjects participated in the broadband DPOAE experiment. The subjects were divided into two groups of 16 subjects each, which differ in the DPOAE recording starting time. Each group consists of eight male and eight female subjects, group 1 having a mean age of 23.4±2.3 years and group 2 a mean age of 23.3±1.8 years.

DPOAEs were measured in the frequency range of 708 Hz$\leqslant f_2 \leqslant$6165 Hz, with a frequency resolution of eight DPOAEs per octave. This is a compromise between frequency resolution and measurement duration. The primary tones were swept discretely through 26 frequencies, starting at the lowest frequency. Each pair of primary tones was averaged in the time domain (16 subaverages), corresponding to approximately 1.5-s measurement time for one pair of frequencies.

1. Repeated pre-exposure measurements

Eight repeated DPOAE measurements were taken before the exposure. For four DPOAEs the probe was removed and replaced before each measurement; for the remaining four DPOAEs the measurement probe stayed in position. In this way the test-retest repeatability (standard deviation between repeated measurements) is obtained between pre- and postexposure measurements (with probe reinsertion) and between postexposure measurements (without probe reinsertion).

2. Postexposure measurements

The time-frequency distribution of DPOAE measurements can be seen in Fig. 1. For each group the first three measurements were taken every 2 min starting at 1 min for group 1 and 1 min and 40 s for group 2. The last three measurements were taken every 5 min starting at 10 min for group 1 and at 12 min and 20 s for group 2. This gave a total of 12 observations from 1 to 24 min for the combination of both groups. The duration of one DPOAE sweep was approximately 40 s.

E. DPOAE fine structure

Eight male and eight female subjects (24±2.3 years) participated in the experiment. Nine out of these 16 subjects participated in the broadband DPOAE experiment also.

The DPOAE fine structure was measured in the frequency range from 903 Hz$\leqslant f_2 \leqslant$2295 Hz, using the frequency resolution "micro" (17 primary tones within 200-Hz intervals). The frequency range was chosen in a way that both the primary frequency f_2 and the DP frequency $2f_1-f_2$ covered the frequency region of the exposure frequency and 1/2 octave above (Fig. 2). To cover the full frequency range, seven DPOAE sweeps in different frequency ranges were measured. For the analysis, the seven measurements of the different frequency regions were concatenated to one DPOAE measurement covering the measured frequency range (Fig. 3). Each pair of primary tones was averaged in the time domain (32 subaverages for the pre-exposure measurements and 16 subaverages for the postexposure measurements), corresponding to approximately 3- and 1.5-s measurement time for one pair of frequencies, respectively.

1. Postexposure measurements

The DPOAE fine structures were measured according to the same measurement protocol for all 16 subjects, i.e., all subjects followed the same time-frequency order of DPOAE measurements. The first DPOAE sweep was started 1 min after the end of the exposure. Sweeps were measured every minute. To cover the full measured frequency range, seven

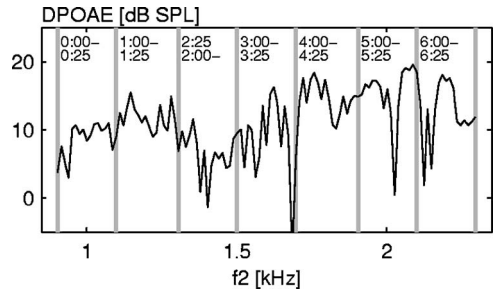


FIG. 3. The measurement of DPOAE fine structure covering the full frequency range is performed by measuring seven sweeps, which are concatenated afterwards. The measurement duration of one DPOAE fine structure is illustrated.

sweeps were measured with a duration of approximately 7 min. The duration of one fine structure measurement can be seen in Fig. 3. After the completion of one fine structure measurement, the measurement was repeated three times, covering an observation time of 29 min.

F. Calculation of DPOAE shift

It was expected that DPOAE levels would decrease after the exposure; thus, the magnitude of the DPOAE shift is taken as the difference between the pre-exposure level and the postexposure level. In this manner a positive shift denotes a decrease in amplitude of the DPOAE. All pre-exposure DPOAE levels are calculated as the average of repeated measurements. Eight repeated measurements were taken, four with and four without reinsertion. The four measurements obtained without reinsertion are averaged and counted as one measurement in order not to have a higher weighting on one probe fit. Consequently, pre-exposure levels are calculated as an average of five reinsertion measurements.

The ILO96 system does not automatically reject DPOAE values if they are below the noise floor, but estimates the noise floor level as the average of the ten Fourier components nearest but not including the $2f_1-f_2$ frequency component. For the calculation of DPOAE shift in the present study, data are not included if they are below the noise floor both before and after the exposure, i.e., if a DPOAE level is above the noise floor before the exposure, but below the noise floor after the exposure (and vice versa), both values are accepted as valid. The noise floor is in the analysis defined as the average across all measured postexposure noise floors. More information on the choice of this data rejection criterion is given in Reuter *et al.* (2007).

III. RESULTS

A. Repeated pre-exposure DPOAE

For the pre-exposure measurements, the standard deviation (SD) of four repeated measurements with probe reinsertion is similar to the SD of four repeated measurements without probe reinsertion at the lowest frequencies and consistently (approximately 0.5 dB) higher at frequencies above $f_2=1$ kHz. The SDs of repeated measurements are highest at the lowest frequencies (5 dB) and decrease with increasing frequency to 1–2 dB.

TABLE I. Balanced three-way-ANOVA p-values, $^{*}p<0.05$, and $^{**}p<0.01$ for broadband DPOAE measured before the exposure. DPOAE levels are compared across frequency=26 levels, probe fit (four repeated measurements with reinsertion and four repeated measurements without reinsertion=8 levels), and groups (two groups with 16 subjects each=2 levels).

	p-value
Frequency	0^{**}
Probe fit	0.1929
Group	0.9519
Freq * Probe fit	1
Freq * Group	0^{**}
Probe fit * Group	0.6851
Freq * Probe fit * Group	1

Mean pre-exposure DPOAE levels across the 16 subjects of each group are similar. Group 1 has a mean DPOAE level, averaged across frequencies, of 3.3±4.6 dB, and group 2 a mean DPOAE level of 3.4±4.9 dB. The DPgram shapes are similar for the two groups. DPOAE levels are relatively flat in the lower frequency range and increased in the higher frequencies with a dip around 3 kHz. Similar DPgram shapes have been found by Lonsbury-Martin *et al.* (1990); Vinck *et al.* (1996); Zhao and Stephens (1999).

A three-way ANOVA was applied to analyze the pre-exposure DPOAE levels according to probe placement and group differences. The data were compared across frequency (=26 levels), probe fit (4 with and 4 without reinsertion=8 levels), and the two participant groups (=2 levels). The results can be seen in Table I. No statistically significant difference could be found between measurements with and without reinsertion. The pre-exposure DPOAE levels were not statistically different between the two groups. The ANOVA does show a significant effect of frequency, because the DPOAE levels vary across frequency. The interaction between frequency and groups reflects a difference of the two groups at some frequencies.

B. Broadband DPOAE shifts

Figure 4 shows DPOAE data obtained before and after the 3-min exposures for three subjects of group 1. The top panels show the pre-exposure DPOAE levels and the first postexposure DPOAE levels. Also shown is the average postexposure noise floor, which is used for the rejection criterion of data below the noise floor. The bottom panels show the calculated DPOAE shift for each subject at the first determination time. As described in Sec. II F, data are not included in the calculation if both pre- and postexposure measurements are below the noise floor. This is the case for some of the frequencies for the subject shown on the right panel. The intersubject variability of DPOAE shift is considerable. The subject shown on the left panel has relatively little DPOAE shift. For the subject shown on the midpanel there is an obvious decrease in DPOAE level after the exposure over the entire measured frequency range. DPOAE shifts are up to

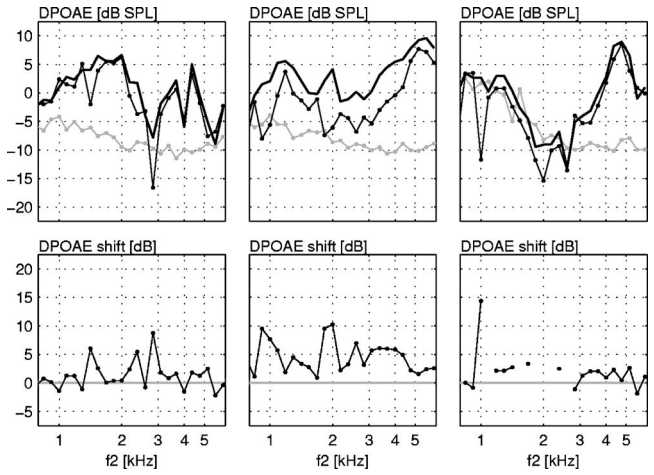


FIG. 4. DPOAE data for three subjects of group 1. Top panels: Measured pre-exposure DPOAE levels (thin, black line), postexposure (1–2 min) DPOAE levels (marked, black line), and average postexposure noise floor (marked, gray line). Bottom panels: DPOAE shift at first determination (marked, black line) time and reference zero (thick, gray line).

10 dB. The subject shown on the right panel has relatively low-level DPOAEs in the midfrequency range and a high noise floor in the low-frequency range. At 1 kHz a very low-level DPOAE is measured after the exposure. Since the pre-exposure DPOAE at that frequency is above the noise floor, a DPOAE shift is calculated. This high amount of DPOAE shift might indicate a DPOAE minimum at that frequency (see Sec. III D).

DPOAE shifts averaged across the 16 subjects of each group are shown in Fig. 5. The data from the subjects of group 1 are presented in the left panels and the data from the subjects of group 2 in the right panels of the figure. For each group the DPOAE shift is shown at all six determination times from top to bottom of the figure. The error bars are SDs of the DPOAE shifts between subjects. The gray shaded area at the zero line represents the measurement uncertainty (SD of pre-exposure DPOAEs with reinsertion) for each group. At the first determination times, mean DPOAE shifts of up to 5 dB can be seen for both groups. The DPOAE shift occurs in a relatively broad frequency range from approximately 1.4–3 kHz, corresponding to 0.5–1.5 octaves above the exposure frequency. Intersubject variations in the DPOAE shift (as indicated by the error bars) are relatively high for the first measurements after the exposure and decrease with time. At the last determination times, DPOAE shifts have normalized to zero. For the data of each group, a two-way ANOVA was performed across measurement time (one pre-exposure and six postexposure measurements=7 levels) and frequency (=26 levels); see Table II. The DPOAE shift is significant up to 10 min for group 1 and up to 4 min for group 2 subjects. From 10-min postexposure there is no statistically significant difference between pre- and postexposure DPOAE levels. A significant difference can be seen for frequency, because not all frequencies are affected equally by the exposure.

In order to study the time course of recovery of DPOAE shift, DPOAE shifts are plotted as a function of time in Fig. 6. DPOAE shifts are averaged across subjects and are shown

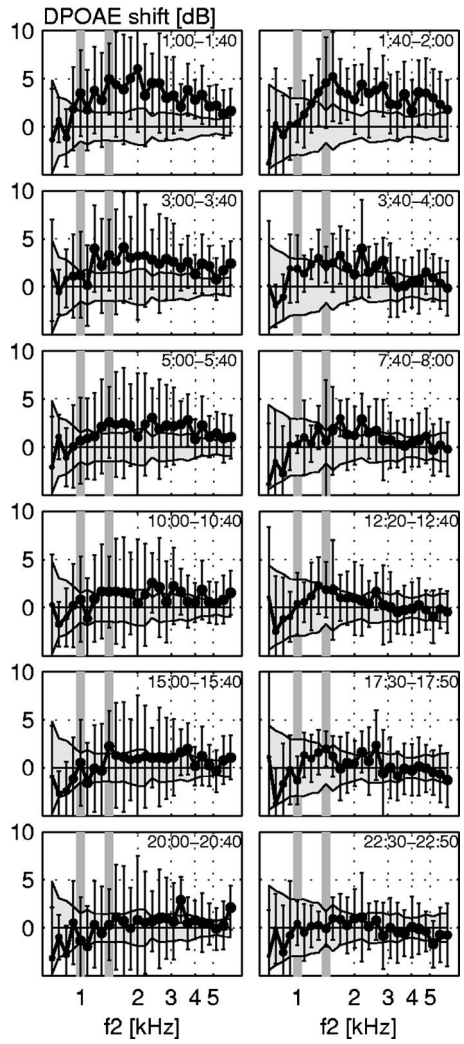


FIG. 5. Mean DPOAE shifts (marked, black lines) and SDs (error bars) across subjects. The marker size corresponds to the number of valid data (see Sec. II F). Left column: group 1 subjects, right column: group 2 subjects. The exposure frequency (1 kHz) and the 1/2-octave shift (1.4 kHz) are indicated. The gray shaded area represents the test-retest repeatability (SD between repeated measurements). The upper right corner indicates the postexposure measurement time.

at single frequencies (black squares) and averaged over the frequency range $916\ \text{Hz} \leqslant f_2 \leqslant 2380\ \text{Hz}$ (gray line). The DPOAE shift recovery can be well modeled with a superposition of only two exponentials, since no 2-min bounce is seen.

C. Comparison to TTS

In a TTS experiment the relationship between the probe signal and the affected area of the basilar membrane is direct. This means that a raised threshold at a certain frequency reflects a change in the normal behavior at the place along the basilar membrane that responds to that specific frequency. For high intensity levels the maximum basilar-membrane excitation is shifted basally to higher frequencies, which is believed to be the reason why the maximum TTS is seen at 1/2 octave above the exposure frequency. In the case of DPOAEs this is not so clear. The DPOAE frequency (in this case $2f_1-f_2$) is a result of the combination of the vibra-

TABLE II. Balanced two-way-ANOVA p-values, $^{*}p<0.05$ and $^{**}p<0.01$ for postexposure DPOAE levels compared to pre-exposure DPOAE levels (=2 levels) and frequency (=26 levels).

Postexposure time	Group	Frequency	Pre/Post	Frequency * Pre/Post
1:00–1:40	1	0^{**}	0^{**}	0.9999
1:40–2:20	2	0.0028^{**}	0^{**}	1
3:00–3:40	1	0^{**}	0^{**}	0.9999
3:40–4:20	2	0.0041^{**}	0.0002^{**}	0.9997
5:00–5:40	1	0^{**}	0^{**}	0.9997
5:40–6:20	2	0.0075^{**}	0.0823	0.9999
10:00–10:40	1	0^{**}	0.0097^{**}	1
12:20–13:00	2	0.0078^{**}	0.4359	1
15:00–15:40	1	0^{**}	0.2332	1
17:20–18:00	2	0.0274^{*}	0.7658	1
20:00–20:40	1	0^{**}	0.5495	1
22:20–23:00	2	0.016^{*}	0.5858	1

tion patterns induced in the basilar membrane by the primaries (f_1 and f_2) and the nonlinear amplification attributed to the OHC. Thus, the significance of a decreased DPOAE may not necessarily mean that the affected area of the basilar membrane corresponds to the place where the DPOAE frequency would generate a vibration peak.

DPOAE shifts of the present study were compared with TTS data, which were induced by a 500-Hz tone presented at $L_{EX,8h}=85.6$ dB (Ordoñez, 2005). The best match between TTS and DPOAE profile is determined from a correlation analysis, where frequencies were normalized with the exposure frequencies, and where DPOAE data were presented as a function of f_2, $2f_1-f_2$, and the geometric mean of the primaries f_1 and f_2. The comparison results in a very high correlation (up to 0.9), but the best match is not unequivocal, as it depends on the determination time and on the group of subjects. For the first set of measurements of group 1 (1 to 1:40 min) there is a greater similarity between TTS and DPOAE when plotted as a function of the geometric mean. For later determination times the $2f_1-f_2$ frequency gives a closer fit. For group 2 the results are reversed: The $2f_1-f_2$ component gives a better fit for the first set of measurements (1:40 to 2:20 min). For later measurements the best fit is found for the geometric mean (2nd, 3rd, and 6th determination) or f_2 (4th and 5th determination).

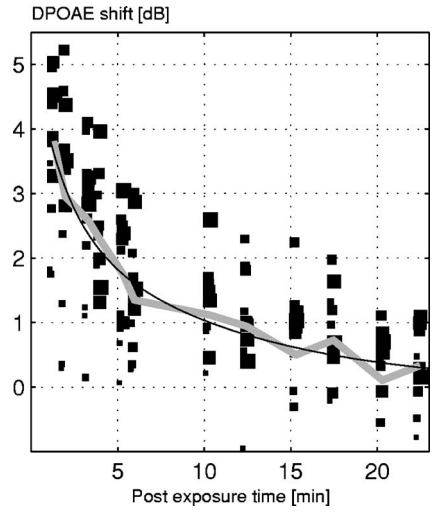


FIG. 6. DPOAE shift averaged across subjects at certain frequencies (black squares; the size increases with increasing frequency) and averaged over the frequency range 916 Hz$\leqslant f_2 \leqslant$2380 Hz (gray line) as a function of time. The data of both groups are concatenated, i.e., data from 12 determination times are shown. An exponential regression (black line) is fitted to the data ($y=4.06\cdot e^{-t/0.70}+3.14\cdot e^{-t/10.24}$).

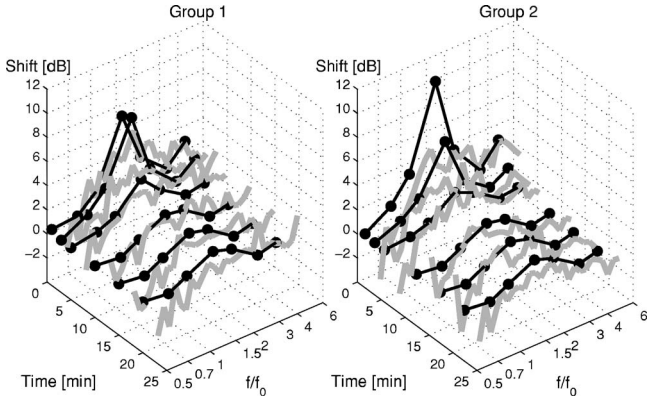


FIG. 7. Comparison of DPOAE shift (gray lines) to TTS data (Ordoñez, 2005) (marked, black lines) in dependence of time and frequency for group 1 (left panel) and group 2 (right panel). DPOAE data are averaged over 16 subjects of each group and plotted in dependence on the geometric mean. The frequency axis is normalized to the exposure frequency.

Figure 7 shows DPOAE shifts and TTS data (Ordoñez, 2005) at various postexposure times for the two groups. Both the frequency and time dependence of the shifts can be seen in the figure. The frequency range and amplitude of TTS and DPOAE shift is similar except for the first two measurements after the exposure. At the early recovery the amount of TTS is higher and pronounced 0.5 octave above the exposure. The DPOAE shift is pronounced over a broader frequency range, 0.5–1.5 octaves above the exposure frequency (in dependence on the geometric mean). At later recovery times, DPOAE shift and TTS are similar, even in absolute terms.

D. DPOAE fine structure

Figure 8 shows the DPOAE fine structures measured before and at 1–7-min postexposure for all subjects. For some subjects (e.g., test subjects 20, 24, and 37), the preexposure fine structure has a pronounced periodic pattern, while for other subjects (e.g., test subjects 25, 29, 35, 43, and 45) this pattern is less pronounced. The purpose of this experiment was to investigate whether the overexposure causes a systematic change in the periodic pattern. For some (e.g., test subjects 23, 24, 31 in the midfrequency range, 33, and 37) the levels at fine structure minima are reduced more than the levels of fine structure maxima, and in this way result in a more pronounced fine structure. The DPOAE levels at minima are often below the noise floor after the exposure. For other subjects (e.g., test subjects 19, 20, 31 at the higher frequency range, 43, and 44) the fine structure is less pronounced after the exposure. The results indicate that the DPOAE fine structure is highly individual, and does not support the idea that an unequivocal relation exists between the fine structure characteristics and the state of hearing.

The statistical differences between the pre- and postexposure DPOAE fine structures are listed in Table III. The

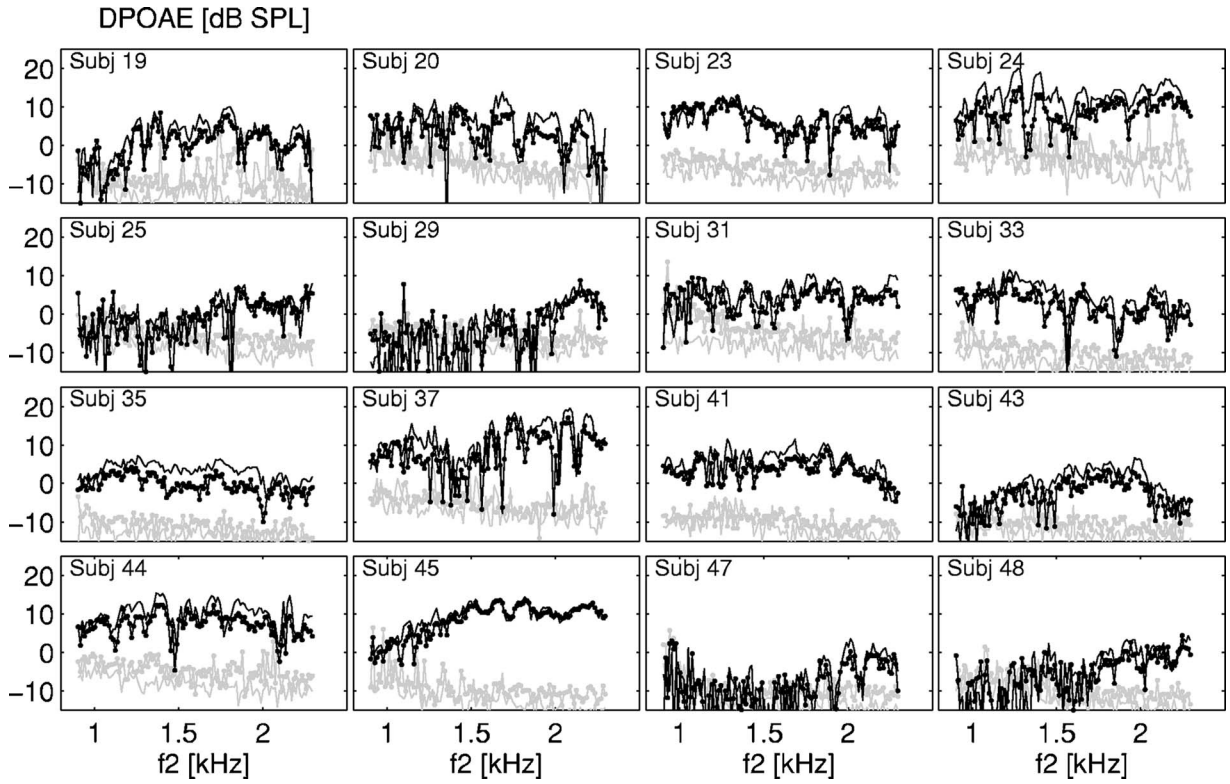


FIG. 8. Individual pre-exposure DPOAE fine structures (thin, black line), postexposure (1–8 min) DPOAE fine structures (marked, black line), pre-exposure noise floors (gray line), and postexposure noise floors (marked, gray line).

TABLE III. Balanced two-way-ANOVA p-values, $^{*}p<0.05$ and $^{**}p<0.01$, comparison of pre-exposure DPOAE fine structure with postexposure fine structure at each determination time. Comparison of frequency (approximately 17 frequencies, depending on the length of sweep) and pre- and postexposure (=2 levels).

		Postexposure time [min]			
f_2 [Hz]		1–7	8–14	15–21	22–28
	Freq	0.2674	0.3733	0.1793	0.101
903–1086	Pre/Post	0.9228	0.3108	0.4109	0.3107
	Freq * Pre/Post	0.948	0.9761	0.9504	0.9799
	Freq	0.4459	0.5571	0.9396	0.3113
1099–1294	Pre/Post	0^{**}	0.0658	0.1106	0.5938
	Freq * Pre/Post	0.9951	0.9998	0.9995	0.9998
	Freq	0.6157	0.7969	0.3791	0.7327
1306–1489	Pre/Post	0^{**}	0.0025^{**}	0.2186	0.1893
	Freq * Pre/Post	0.9991	0.9999	0.997	0.9999
	Freq	0.3693	0.3028	0.619	0.7397
1501–1685	Pre/Post	0^{**}	0.0004^{**}	0.0499^{*}	0.0226^{**}
	Freq * Pre/Post	1	1	1	1
	Freq	0.9682	0.8675	0.9631	0.9631
1697–1892	Pre/Post	0.0001^{**}	0.0013^{**}	0.0322^{*}	0.0196^{*}
	Freq * Pre/Post	1	1	1	1
	Freq	0.8995	0.7185	0.3974	0.4632
1904–2087	Pre/Post	0^{**}	0.0035^{**}	0.0015^{**}	0.0184^{*}
	Freq * Pre/Post	0.9727	0.9856	1	0.9995
	Freq	0.9641	0.8341	0.987	0.9467
2100–2295	Pre/Post	0.0007^{**}	0.0295^{*}	0.0283^{*}	0.0508
	Freq * Pre/Post	0.9999	1	1	1

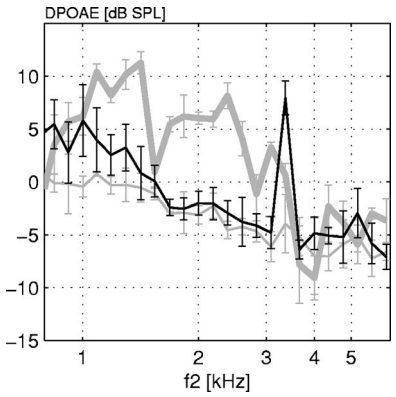


FIG. 9. Pre-exposure DPOAE level (thick gray line), pre-exposure noise floor (thin gray line), and postexposure noise floor (black line) for one subject. The postexposure noise floor is averaged over six postexposure measurements.

ANOVA has been performed for each sweep and at all determination times. An effect can be seen for the parameter pre- versus postexposure, where the statistical differences are generally highest in the early recovery time and lower in the later determination times.

IV. DISCUSSION

In the present study high intersubject differences could be observed in the amount of DPOAE shift after the exposure. Also, in TTS experiments the amount of shift differs from subject to subject (Melnick, 1991; Ordoñez, 2005; Quaranta *et al.*, 1998). Differences in the physical exposure will also systematically contribute to the interindividual differences in TTS and DPOAE shifts. In the present study a pure tone of 1 kHz was used, which is known to vary less than approximately 2 dB between subjects [e.g., Shaw (1966)]. Thus, the high interindividual variation in shifts can only be explained by differences in susceptibility.

The comparison of DPOAE shifts to TTS shows that the maximum shift of both measures occurs in a frequency range above the exposure frequency. This maximum shift towards higher frequencies supports the idea of a spread in excitation toward the basal end of the cochlea with increases in exposure level. While the maximum TTS occurs in a relatively narrow frequency range at approximately 1/2 octave above the exposure frequency (depending on the frequency resolution used for the determination of hearing thresholds), the DPOAE shift occurs in a relatively broad frequency range covering 0.5–1.5 octaves above the exposure frequency. The broad DPOAE shift might indicate that the overexposure causes an exhaustion of OHCs over a broad region on the basilar membrane. This exhaustion might be detectable by OAE; in this case OAE would be more sensitive than TTS to detect these small changes.

Figure 9 shows the pre-exposure DPOAE level, the pre-exposure noise floor, and the mean postexposure noise floor for one subject. For this subject a clear increase of the noise floor was observed after the exposure above 3 kHz. The measurement system used in this study estimates the noise floor from the ten Fourier components nearest but not including the $2f_1-f_2$ component. This means that there possibly is an increased activity in the neighbor bands after the exposure that has not been there before the exposure. Generally, a slight increase of noise floor with increasing postexposure time was found in both experiments of the present study. This may either be due to true activities in neighbor bands, but it may also be due to minor changes in the fit of the probe combined with more noise at the end of the experiment, where the subject may no longer sit perfectly still.

In the present study the data of the DPOAE fine structure experiment are shown as raw data, and conclusions are drawn by visually inspecting the data. By determining frequency positions and levels at fine structure maxima and minima, the fine structure pattern can be described by the parameters ripple spacing, ripple height, and ripple prevalence, as suggested by Reuter and Hammershøi (2006). Whether the fine structure changes systematically, e.g., flattens or becomes more pronounced after the exposure, can be determined from an analysis of ripple heights at each determination time, but the analysis is, however, not trivial in the present case for two reasons: (1) The calculation of ripple parameters depends on the noise floor of the system. Since different measurement average times were used for the DPOAE fine structures obtained before and after the exposure, the noise floor differs systematically between these two measures. Additionally, an increase of noise floor with time during postexposure has been observed in both experiments. (2) The measurement time of the DPOAE fine structure in the present study lasts 7 min and is rather long. The data were always collected in a sweep starting at the lowest frequency and ending at the highest frequency; therefore, the first effects of the exposure will always be measured for the lower frequencies, while the higher frequencies would always be measured at later recovery times.

Both flattened and more pronounced fine structures could be observed after the exposure, which contradicts the hypothesis that fine structure should systematically indicate the state of hearing. The exposure in the present study was a narrow-band exposure, which is expected to affect hair cells in a narrow region of the basilar membrane. According to the two-source model, DPOAE fine structure is caused by constructive and destructive interference of the distortion and the reflection component. The most pronounced fine structure is expected when the two components have similar amplitudes. If one component has a higher amplitude than the other, then a reduction of the higher-amplitude component will result in an increase of fine structure, and a reduction of the lower-amplitude component will result in a decrease of DPOAE fine structure. According to this theory, one might expect a systematic modification of fine structure after an equal narrow-band exposure, if both components have similar relations to each other for all subjects. The results of the present study indicate, that the relative component contributions of the pre-exposure DPOAEs might differ from subject to subject.

ACKNOWLEDGMENTS

This work was financed by the William Demant Foundation (Oticon) and The Danish Research Council for Tech-

nology and Production Sciences. The authors would like to thank all subjects for participating in this experiment.

Attias, J., Furst, M., Furman, V., Reshef, I., Horowitz, G., and Bresloff, I. (**1995**). "Noise-induced otoacoustic emission loss with or without hearing loss," Ear Hear. **16**, 612–618.

Botsford, J. H. (**1971**). "Theory of temporary threshold shift," J. Acoust. Soc. Am. **49**, 440–446.

de Boer, E., and Nuttall, A. L. (**1997**). "The mechanical waveform of the basilar membrane. I. Frequency modulation (glides) in impulse responses and cross-correlation functions," J. Acoust. Soc. Am. **101**, 3583–3592.

de Boer, E., and Nuttall, A. L. (**2000**). "The mechanical waveform of the basilar membrane. III. Intensity effects," J. Acoust. Soc. Am. **107**, 1497–1507.

Delb, W., Hoppe, U., Liebel, J., and Iro, H. (**1999**). "Determination of acute noise effects using distortion product otoacoustic emissions," Scand. Audiol. **28**, 67–76.

Desai, A., Reed, D., Chayne, A., Richards, S., and Prasher, D. (**1999**). "Absence of otoacoustic emissions in subjects with normal audiometric thresholds implies exposure to noise," Noise Health **2**, 58–65.

Dhar, S., Talmadge, C. L., Long, G., and Tubis, A. (**2002**). "Multiple internal reflections in the cochlea and their effect on DPOAE fine structure," J. Acoust. Soc. Am. **112**(6), 2882–2897.

Engdahl, B., and Kemp, D. T. (**1996**). "The effect of noise exposure on the details of distortion product otoacoustic emissions in humans," J. Acoust. Soc. Am. **99**, 1573–1587.

Gaskill, S. A., and Brown, A. M. (**1990**). "The behavior of the acoustic distortion product, $2f_1-f_2$, from the human ear and its relation to auditory sensitivity," J. Acoust. Soc. Am. **88**, 821–839.

Hauser, R., and Probst, R. (**1991**). "The influence of systematic primary-tone level variation L_2-L_1 on the acoustic distortion product emission $2f_1-f_2$ in normal human ears," J. Acoust. Soc. Am. **89**, 280–286.

Hirsh, I. J., and Ward, W. D. (**1952**). "Recovery of the auditory threshold after strong acoustic stimulation," J. Acoust. Soc. Am. **24**, 131–141.

Johnstone, B. M., Patuzzi, R., and Yates, G. K. (**1986**). "Basilar membrane measurements and the travelling wave," Hear. Res. **22**, 147–153.

Kemp, D. T. (**1986**). "Otoacoustic emissions, travelling waves and cochlear mechanisms," Hear. Res. **22**, 95–104.

Kirk, D. L., Moleirinho, A., and Patuzzi, R. B. (**1997**). "Microphonic and DPOAE measurements suggest a micromechanical mechanism for the 'bounce' phenomenon following low-frequency tones," Hear. Res. **112**, 69–86.

Kværner, K. J., Engdahl, B., Arnesen, A. R., and Mair, I. W. S. (**1995**). "Temporary threshold shift and otoacoustic emissions after industrial noise exposure," Scand. Audiol. **24**, 137–141.

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., and Murray, N. M. (**1993**). "Click-evoked otoacoustic emissions: Comparing emissions strength with pure tone audiometric thresholds," Aust. J. Audiol. **15**, 9–22.

LePage, E. L., and Murray, N. M. (**1998**). "Latent cochlear damage in personal stereo users: a study based on click-evoked otoacoustic emissions," Med. J. Aust. **169**, 588–592.

LePage, E. L., Murray, N. M., Tran, K., and Harrap, M. J. (**1993**). "The ear as an acoustical generator: Otoacoustic emissions and their diagnostic potential," Acoust. Aust. **21**, 86–90.

Liebel, J., Delb, W., Andes, C., and Koch, A. (**1996**). "Die Erfassung von Lärmschäden bei Besuchern einer Diskothek mit Hilfe der TEOAE und DPOAE" ("The collection of noise damage with visitors of a discotheque with the help of the TEOAE and DPOAE"), Laryngorhinootologie **75**, 259–264.

Lonsbury-Martin, B. L., Martin, G. K., Probst, R., and Coats, A. C. (**1987**). "Acoustic distortion products in the rabbit ear canal. I. Basic features and physiological vulnerability," Hear. Res. **28**, 173–189.

Lonsbury-Martin, B. L., Harris, F. P., Stagner, B. B., Hawkins, M. D., and Martin, G. K. (**1990**). "Distortion product emissions in humans I: basic properties in normally hearing subjects," Ann. Otol. Rhinol. Laryngol. Suppl. **147**, 3–14.

Lucertini, M., Moleti, A., and Sisto, R. (**2002**). "On the detection of early cochlear damage by otoacoustic emission analysis," J. Acoust. Soc. Am. **111**, 972–978.

Lydolf, M. (**1999**). "The thresholds of hearing and contours of equal loudness," Ph.D. dissertation, Aalborg University.

Mauermann, M., Uppenkamp, S., van Hengel, P. W. J., and Kollmeier, B. (**1999a**). "Evidence for the distortion product frequency place as a source of distortion product otoacoustic emission (DPOAE) fine structure in humans. I. Fine structure and higher-order DPOAE as a function of the frequency ratio f_2/f_1," J. Acoust. Soc. Am. **106**, 3473–3483.

Mauermann, M., Uppenkamp, S., van Hengel, P. W. J., and Kollmeier, B. (**1999b**). "Evidence for the distortion product frequency place as a source of distortion product otoacoustic emission (DPOAE) fine structure in humans. II. Fine structure for different shapes of cochlear hearing loss," J. Acoust. Soc. Am. **106**, 3484–3491.

Martin, G. K., Lonsbury-Martin, B. L., Probst, R., Scheinin, S. A., and Coats, A. C. (**1987**). "Acoustic distortion products in the rabbit ear canal. II. Sites of origin revealed by suppression contours and pure-tone exposures," Hear. Res. **28**, 191–208.

McFadden, D., and Plattsmier, H. S. (**1983**). "Frequency patterns of tts for different exposure intensities," J. Acoust. Soc. Am. **74**, 1178–1184.

Melnick, W. (**1991**). "Human temporary threshold shifts (tts) and damage risk," J. Acoust. Soc. Am. **90**, 147–154.

Ordoñez, R. (2005). "Temporary changes in human hearing caused by intense sounds," Ph.D. dissertation, Aalborg University.

Patuzzi, R. (**1998**). "Exponential onset and recovery of temporal threshold shift after loud sound: Evidence for long-term inactivation on mechano-electrical transduction channels," Hear. Res. **125**, 17–38.

Quaranta, A., Portalatini, P., and Henderson, D. (**1998**). "Temporary and permanent threshold shift: An overview," Scand. Audiol. Suppl. **27**, 75–86.

Rao, A., Long, G. R., Narayan, S., and Dhar, S. (**1996**). "Changes in the temporal characteristics of TEOAE and the fine structure of DPOAEs with aspirin consumption," 19th ARO Midwinter Research Meeting, Abstract p. 27.

Reuter, K., and Hammershøi, D. (**2006**). "Distortion product otoacoustic emission fine structure analysis of 50 normal-hearing humans," J. Acoust. Soc. Am. **120**, 270–279.

Reuter, K., Ordoñez, R., de Toro, M. A. A., and Hammershøi, D. (**2007**). "Rejection criteria for DPOAE measurements below the noise floor," (to be presented at the 19th International Congress on Acoustics, ICA 2007, Madrid).

Rhode, W. S., and Recio, A. (**2000**). "Study of mechanical motion in the basal region of the chinchilla cochlea," J. Acoust. Soc. Am. **107**, 3317–3332.

Ruggero, M. A., Rich, N. C., Recio, A., Narayan, S. S., and Robles, L. (**1997**). "Basilar-membrane responses to tones at the base of the chinchilla cochlea," J. Acoust. Soc. Am. **101**, 2151–2163.

Schmiedt, R. A. (**1986**). "Acoustic distortion in the ear canal. I. Cubic difference tones: Effects of acute noise injury," J. Acoust. Soc. Am. **79**, 1481–1490.

Shaffer, L., Withnell, R., Dhar, S., Lilly, D., Goodman, S., and Harmon, K. (**2003**). "Sources and mechanisms of DPOAE generation: Implications for the prediction of auditory sensitivity," Ear Hear. **24**, 367–379.

Shaw, E. A. G. (**1966**). "Earcanal pressure generated by circumaural and supraaural earphones," J. Acoust. Soc. Am. **39**, 471–480.

Sutton, L. A., Lonsbury-Martin, B. L., Martin, G. K., and Whitehead, M. L. (**1994**). "Sensitivity of distortion-product otoacoustic emissions in humans to tonal overexposure: Time course of recovery and effects of lowering L_2," Hear. Res. **75**, 161–174.

Vinck, B. M., Devel, E., Xu, Z. M., and Vancauwenberge, P. B. (**1996**). "Distortion product otoacoustic emissions: A normative study," Audiology **35**, 231–245.

Vinck, B. M., Van Cauwenberge, P. B., Leroy, L., and Corthals, P. (**1999**). "Sensitivity of transient evoked and distortion product otoacoustic emissions to the direct effects of noise on the human cochlea," Audiology **38**, 44–52.

Whitehead, M. L., McCoy, M. J., Lonsbury-Martin, B. L., Martin, G. K. (**1995a**). "Dependence of distortion-product otoacoustic emissions on primary levels in normal and impaired ears. I. Effects of decreasing L_2 below L_1," J. Acoust. Soc. Am. **97**, 2346–2358.

Whitehead, M. L., Stagner, B. B., McCoy, M. J., Lonsbury-Martin, B. L., and Martin, G. K. (**1995b**). "Dependence of distortion-product otoacoustic emissions on primary levels in normal and impaired ears. II. Asymmetry in L_1, L_2 space," J. Acoust. Soc. Am. **97**, 2359–2377.

Zhao, F., and Stephens, D. (**1999**). "Test-retest variability of distortion-product otoacoustic emissions in human ears with normal hearing," Scand. Audiol. **28**, 171–178.

Otoacoustic emission sensitivity to low levels of noise-induced hearing loss

Renata Sisto[a)]
Dipartimento Igiene del Lavoro, ISPESL, Via Fontana Candida, 1, 00040 Monte Porzio Catone, Roma, Italy

Silvia Chelotti, Lido Moriconi, and Stefania Pellegrini
Prevention Department, AUSL, Viareggio, Italy

Angela Citroni and Valeria Monechi
Prevention Department, AUSL, Firenze, Italy

Rosa Gaeta, Iole Pinto, and Nicola Stacchini
Prevention Department, AUSL, Siena, Italy

Arturo Moleti[b)]
Dipartimento di Fisica, Università di Roma "Tor Vergata," Via della Ricerca Scientifica, 1, 00133 Roma, Italy

(Received 15 January 2007; revised 13 April 2007; accepted 16 April 2007)

With the aim of investigating the capability of otoacoustic emission (OAE) in the detection of low levels of noise-induced hearing loss, audiometric and otoacoustic data of young workers (age: 18–35) exposed to different levels of industrial noise have been recorded. These subjects are participating in a long-term longitudinal study, in which audiometric, exposure (both professional and extra-professional), and OAE data (transient evoked and distortion product) will be collected for a period of several years. All measurements have been performed, during routine occupational health surveillance, with a standard clinical apparatus and acquisition procedure, which can be easily used in the occupational safety practice. The first study was focused on the correlation between transient evoked OAE signal-to-noise ratio and distortion product (DPOAE) OAE level and the audiometric threshold, investigating the causes of the rather large intersubject variability of the OAE levels. The data analysis has shown that, if both OAE data and audiometric data are averaged over a sufficiently large bandwidth, the correlation between DPOAE levels and audiometric hearing threshold is sufficient to design OAE-based diagnostic tests with good sensitivity and specificity also in a very mild hearing loss range, between 10 and 20 dB. © *2007 Acoustical Society of America.* [DOI: 10.1121/1.2737668]

PACS number(s): 43.64.Jb, 43.64.Wn [BLM] Pages: 387–401

I. INTRODUCTION

Otoacoustic emissions (OAEs) are acoustic signals generated in the inner ear and recorded in the ear canal. They may either be spontaneous (SOAEs) or an evoked response (EOAEs) to acoustical stimulation (Probst *et al.*, 1991). OAEs are usually classified according to the recording technique, distinguishing between transient evoked OAEs (TEOAEs), evoked by a click, a tone-burst, a chirp, or a MLS sequence, distortion product OAEs (DPOAEs), evoked by two tones f_1 and f_2 at the frequency of their distortion products (the largest and most commonly studied is the cubic distortion product at frequency $2f_1$-f_2) due to the cochlear nonlinearity, and stimulus-frequency OAEs (SFOAEs), evoked at the same frequency of the single-tone stimulus.

The different OAE techniques provide a great deal of information about the most delicate mechanism of the cochlear function, namely the active feedback mediated by the outer hair cells (OHCs), which are typically the first part of the auditory system to be significantly affected by the exposure to important ototoxic agents, such as noise and drugs. Therefore, OAEs have been considered as a very promising tool to detect mild hearing loss in exposed subjects. The large intersubject variability of the OAE levels makes it difficult to design OAE-based tests capable of directly evaluating the hearing threshold in a single subject. On the other hand, the increasingly accurate knowledge of the cochlear amplifier and of the OAE generation mechanisms provided by recent advances in physiological acoustics permits one to understand the causes of this variability, and to minimize its effects by designing appropriate data acquisition and analysis procedures. For example, the DPOAE response variability is partly due to the vector sum of two separate interfering sources, as demonstrated by several studies of the DPOAE spectral fine structure (e.g., Mauermann *et al.*, 1999).

Several studies have demonstrated the sensitivity to noise exposure of the SOAE levels (e.g., Furst *et al.*, 1992),

[a)]Electronic mail: renata.sisto@ispesl.it
[b)]Electronic mail: arturo.moleti@roma2.infn.it

and of the EOAE levels (e.g., Attias *et al.*, 1995), and also the OAE sensitivity to ototoxic drugs (e.g., Long and Tubis, 1988; Furst *et al.*, 1995).

It has generally been observed that hearing loss above a given level (about 40 dB) over all the hearing range is correlated to almost total TEOAE absence, and that when hearing loss is lower than 20 dB over the whole audiogram, measurable TEOAEs are generally present (e.g., Bonfils *et al.*, 1988; Prieve *et al.*, 1993). This observation is consistent with the loss of amplification associated with the destruction of the OHC active feedback mechanism and with the fact that passive cochlear models do not predict measurable OAE levels. The above-mentioned evidence led to the development of pass-fail neonatal hearing screening tests, based on the measure of TEOAE signal-to-noise ratio (SNR) in selected frequency bands, which have found wide clinical application. For very high HL levels, the TEOAE techniques are not very useful, because the SNR is always too poor to allow discrimination between different levels of severe hearing loss.

The frequency specificity of the OAE response permits one to get more accurate spectral information on the hearing function. For example, no SOAEs are present in the frequency bands where the HL level is higher than 20 dB, while they may be found in the same hearing-impaired ears, just outside the impaired frequency range (Sisto *et al.*, 2001). This frequency specific SOAE sensitivity to mild hearing loss is interesting, because it confirms the relation between the SOAE absence and the increased hearing threshold, which are both due to the damage of the OHC active mechanism. Unfortunately, SOAE presence in a given frequency band is not necessary at all for normal hearing, so these studies have had scarce clinical output.

In the mild hearing loss range, the evoked OAE levels are quantitatively correlated with the sensorineural HL level. Several cross-section statistical studies (e.g., Attias *et al.*, 1995; Lucertini *et al.*, 2002) established indeed that frequency specific indicators based on the TEOAE and DPOAE spectral levels can effectively discriminate between populations of hearing impaired and normal hearing ears, and also between "normal hearing" populations with different levels of noise exposure. This last finding must be interpreted, in the sense that the audiometric threshold distributions of the exposed and nonexposed "normal" populations actually differed, but the maximum HL level was for all subjects below the conventional clinical limit of 20 dB HL, generally adopted for defining the onset of hearing impairment. Within this operational definition of hearing impairment, a correlation between hearing threshold level and DPOAE level had also been reported by Dorn *et al.* (1998), who showed a clear correlation between maximum hearing threshold in the normal range (from −5 to 20 dB) and the DPOAE level averaged across frequencies, which varied from approximately 6 to −3 dB in that threshold range. These observations generally demonstrate OAE sensitivity to very low levels of noise-induced hearing loss, with a maximum in the 1–4 kHz frequency range. Moleti *et al.* (2002) also evaluated the different sensitivity of linear and nonlinear TEOAE acquisition techniques, in a mild hearing loss range. The effectiveness of TEOAEs elicited by MLS stimuli for the diagnosis of hearing loss has not been fully established yet (Hine and Thornton, 2005).

Other studies also suggested a correlation between hearing threshold and TEOAE latency, estimated as a function of frequency by time-frequency analysis of the TEOAE responses (Sisto and Moleti, 2002; Jedrzejczak *et al.*, 2005). Increased latency was generally observed in hearing impaired ears, analyzing the response to intense (80 dB) clicks or tone bursts. This result is not easily interpreted, because in principle the OAE latency should be shorter in impaired ears. In fact, if hearing impairment is associated with reduced frequency resolution of the cochlear filter, lower values of the quality factor of the cochlear resonance should imply shorter TEOAE latency (Moleti and Sisto, 2003). TEOAE latency measurements could also be affected by large systematic errors, associated with data selection criteria and the presence of SOAEs, as recently shown by Moleti *et al.* (2005).

The results of the above-presented cross-section studies are certainly promising, but do not directly imply the possibility of developing a reliable OAE diagnostic tool for evaluating the hearing threshold in a single subject. During the 1990s, this problem was extensively studied (e.g., Gorga *et al.*, 1993b), analyzing the individual correlation between TEOAE and DPOAE levels and audiometric threshold. Good correlation has been found between OAE levels at mid-high frequencies and HL, for HL$>$20 dB, even if a very large intersubject variability of the OAE levels in ears with the same hearing threshold was observed. Gorga and co-workers used the ROC plots to quantitatively evaluate the performance of dichotomous OAE-based tests of the hearing threshold level, finding that both TEOAEs (Prieve *et al.*, 1993; Gorga *et al.*, 1993b) and DPOAEs (Gorga *et al.*, 1993a, b, 1997) can provide good screening tests for detecting hearing loss levels in the 20–30 dB HL range. They also demonstrated that the best test performances are obtained with DPOAEs elicited by primary levels of 65 and 55 dB (Stover *et al.*, 1996a). Later, they also found that the predictive power of a DPOAE test is significantly increased if more primary frequencies around the audiometric frequency are simultaneously taken into account in a multivariate approach (Dorn, 1999; Gorga *et al.*, 1999, 2005), and, to a lesser extent, if both the $2f_1$-f_2 and the $2f_2$-f_1 distortion product levels are considered (Gorga *et al.*, 2000). The TEOAE test performances were also improved by using multivariate statistical analysis (Hussain *et al.*, 1998). Hall and Lutman (1999)) also compared the sensitivity of OAE techniques and audiometry, studying the individual relation between OAE levels and hearing threshold, and taking into account the different test-retest fluctuation levels. Their results show good OAE sensitivity, such that they conclude that OAE techniques could be preferable to standard audiometry for the monitor of populations of subjects exposed to noise, but they also show very large intersubject variability. This makes it difficult to use OAEs even for a screening test of the hearing threshold in a single subject with no longitudinal monitor.

Longitudinal studies try to investigate directly the effect of noise exposure on the OAE levels. Recently, Lapsley-Miller *et al.* (2006), demonstrated OAE sensitivity to aircraft

noise exposure, but poor correlation of the OAE levels with shifts of the audiometric threshold. This observation needs to be interpreted, but implies anyway that studying the correlation between OAE levels and hearing threshold in exposed subjects does not provide a direct measure of the OAE capability of detecting the early effects of exposure.

An useful OAE classification, proposed by Shera and Guinan (1999), is based on the present knowledge of the OAE generation mechanisms. According to their taxonomy, OAEs arise from two different mechanisms, nonlinear distortion and linear reflection. Almost all the OAE acquisition techniques are sensitive to both these components, with different weights, dependent on the level of the evoking stimulus and on the hearing functionality of the subject. Another fundamental distinction is made between wave-fixed and place-fixed generation mechanism. For wave-fixed mechanisms, the OAE of a given frequency is generated at a cochlear place that is a function of the stimulus frequencies, while place fixed mechanisms assume that emissions come from discrete cochlear irregularities, whose position is fixed, so the OAE source position does not change for small shifts of the stimulus frequency. In the Shera and Guinan (1999) scheme, nonlinear distortion generation is wave fixed and linear reflection is place fixed, which implies a significantly different behavior of the predicted phase-frequency relation. It is rather widely accepted that DPOAE signals at the frequency $f_{dp}=2f_1-f_2$ are the superposition of two waves (Stover *et al.*, 1996b; Shera and Guinan, 1999; Kalluri and Shera, 2001). The first wave is generated by nonlinear distortion through the interaction of the primary waves near the cochlear place that is resonant at the f_2 frequency, while the second is supposed to be due to linear reflection of the fraction of the nonlinearly generated wave that travels in the apical direction, near the cochlear place that is resonant at the frequency f_{dp}. Konrad-Martin *et al.* (2001) used a suppressor tone at the f_{dp} frequency to isolate the nonlinear distortion contribution, and they also found partial correlation between the slope of the audiogram and the weight of the second DP source in hearing impaired subjects (Konrad-Martin *et al.*, 2002). The wave-fixed nature of the first source implies that its phase-gradient delay (the derivative of the emission phase with respect to frequency) should be null, while the place-fixed nature of the second source implies fast rotation of the OAE phase as a function of frequency. The presence of these two components manifests itself in the well-known DPOAE spectral fine structure, generally assumed to be due to the interference between a component with constant phase and another with rapidly rotating phase. The two sources can be separated (Kalluri and Shera, 2001; Mauermann and Kollmeier, 2004) by time windowing of the Fourier Transform (FT) of the fine structure spectrum. The different weight of the two components as a function of the stimulus level has been estimated using this technique, showing that the place-fixed contribution is more important at low stimulus levels, as expected for a linear mechanism. For TEOAEs evoked by wide band stimuli such as clicks, an intermodulation distortion component is present along with a linear reflection component (Yates and Withnell, 1999), and the presence of SOAEs has been shown to influence strongly the overall TEAOE response (Sisto *et al.*, 2001).

If all the above-presented information coming from cochlear models is neglected, the interference between different OAE sources with different phase-frequency behavior may result in an increase of the experimental uncertainty in the OAE measurements, with consequent loss of correlation with the audiometric threshold level. A deep understanding of the cochlear OAE generation mechanisms is therefore a necessary prerequisite for any attempt to design a reliable OAE-based indicator of the hearing threshold level. For example, it has been recently shown (Mauermann and Kollmeier, 2004) that the large uncertainty in the hearing threshold estimates of the procedure proposed by Boege and Janssen (2002), which is based on a logarithmic fit to the low level end of the DPOAE I-O curves, is mainly due to the presence of the DPOAE fine structure. In fact, the test frequency may randomly fall either on a minimum or on a maximum of the fine structure, and due to the slow (but not negligible) frequency shift of this structure with increasing stimulus level, this condition is not maintained over all the I-O curve. The time-windowing technique may solve this problem (Mauermann and Kollmeier, 2004), but it requires a rather long measurement time, because the whole spectral fine structure must be measured at several stimulus levels.

Another difficulty arises from the presence of SOAEs. As mentioned before, although the presence of SOAEs is typically correlated with low hearing thresholds (Schloth, 1983; Sisto *et al.*, 2001) in a frequency specific sense, the SOAE presence is not a necessary condition for normal hearing in a given frequency band. On the other hand, the SOAE presence strongly affects the EOAE responses in its frequency range. For example, the TEOAE response is often dominated by the contribution of a few long-lasting resonant emissions at the SOAE frequencies (Sisto *et al.*, 2001) and this results in an additional source of intersubject variability of the TEOAE levels in a given frequency band, and in a different dependence of the expected TEOAE level on the hearing loss level. For DPOAEs and SFOAEs, the presence of a SOAE near one of the emission places may even cause suppression phenomena, resulting in a low OAE response in ears with very low threshold around that frequency.

In this work, we will try to find a reliable OAE indicator of early cochlear damage, starting from standard TEOAE and DPOAE data, recorded with simple and fast experimental procedures, which can be easily followed in a typical workplace environment.

With respect to previous studies two differences are noteworthy:

(1) Much care has been put into the selection of the subjects, to include only subjects whose hearing loss is presumably due to noise exposure. All selected subjects reported no exposure to other ototoxic agents, no pathology of the inner or middle ear, or any other pathology that could imply a hearing threshold shift, and they were all younger than 35 years, to limit the effect of age. This is different from what has usually been done in important previous studies, based on populations of patients visit-

ing the clinic for audiological tests (e.g., Gorga *et al.*, 1993b). In those studies, only conductive problems of the outer and middle ear were typically excluded, accepting subjects affected by several different cochlear pathologies. This study focuses on noise-induced hearing loss only, whose relation with OAE levels could be different than that caused by other pathologies, due to the peculiar relation between the OAE generation mechanisms and the OHC active filter. Therefore, these results provide specific information, which could help in solving the occupational health problem of early detection of noise-induced hearing loss in exposed workers.

(2) The average hearing loss levels of the subjects of this study are very low. Even for the hearing-impaired class of subjects, most of the statistical sample is concentrated between 0 and 20 dB of hearing loss, thus the study focuses on the sensitivity of OAEs to very early stages of hearing impairment, which is the most important issue in the occupational safety field, while in the audiological standard practice the quantitative evaluation of higher levels of hearing threshold shift could be more important.

First we will present new data of a population of young workers exposed to industrial noise that confirm the results of previous cross-section studies (e.g., Lucertini *et al.*, 2002), as regards the sensitivity of TEOAEs and DPOAEs to mild hearing loss, then we will focus on the intersubject variation of the OAE levels and look for a composite OAE indicator capable of providing an effective screening test for mild hearing impairment, with a reasonably low rate of errors of Type I and II. The TEOAE SNR and the DPOAE level will be used as primary candidate indicators, according to the results of previous studies (Gorga *et al.*, 1993b; Stover *et al.*, 1996a). The design of a composite OAE-based indicator will be based only on the information coming from basic OAE physiology, and not on advanced statistical procedures. This choice may be questioned, because slightly better test performances could have probably been achieved by more sophisticated statistical approaches, as demonstrated both for TEOAEs (Hussain *et al.*, 1998) and DPOAEs (Dorn *et al.*, 1999; Gorga *et al.*, 1999, 2005). At the present stage, our goal was to show which direction to follow, rather than looking for an empirical fine-tuning of the paradigm. Further implications concerning the possibility of using OAE techniques to monitor hearing function in populations of workers exposed to noise will also be briefly discussed.

The analysis of the TEOAE latency, which has also been evaluated for all the subjects as a function of frequency, deserves a separate study, because its theoretical interpretation is more complex and might be controversial.

II. METHODS

Otoacoustic emissions and audiometric threshold have been measured in both ears of 217 young workers (age: 18–35 years), exposed to different levels of industrial noise. All these subjects are participating in a long-term longitudinal study, in which audiometric, OAE, and exposure (professional and extra-professional, including the use of hearing protector devices) data will be collected for a period of several years. All subjects were interviewed, and accepted for the study only when they reported no exposure to other ototoxic agents such as drugs, and no previous pathology that could have caused a hearing threshold shift. In the longitudinal study, all the exposure information will be used to evaluate the sensitivity of OAEs to different kinds of professional and extra-professional noise exposure. In the present study, a cross-section analysis has been performed on this population, to assess the relation between audiometric and OAE measurements in a mild hearing loss range.

Otoscopy and tympanometry were performed before the OAE and audiometric measurements, to exclude conductive problems from the outer and middle ear. TEOAEs, DPOAEs, and audiometric thresholds have been recorded in an audiometric booth.

The audiometric threshold was measured in 5 dB steps up to −10 dB at the standard frequencies 250, 500, 1000, 2000, 3000, 4000, 6000, 8000 Hz, starting from a level of 40 dB, testing first at 1 kHz, then at increasingly higher frequencies, and finally at the two lowest frequencies.

TEOAEs and DPOAEs were recorded using the commercial clinical apparatus ILO-96 (Otodynamics, Ltd.). TEOAEs were acquired in the standard "derived nonlinear" mode, with an 80 dB click stimulus. In this acquisition mode, two alternate averages of 260 low-noise wave forms are recorded, each wave form being the sum of three responses to identical click stimuli plus the response to a click of opposite polarity and triple amplitude. The low-noise threshold was adaptively set by the tester during acquisition, according to the visual shape of the data histogram. This acquisition mode effectively suppresses the linear ringing artifact exploiting the nonlinearity of the cochlear response. The difference between the two alternate averages provides a rough estimate of the noise level. The data are bandpass filtered and multiplied by a window that zeros the first 2.5 ms and applies a linear ramp from 2.5 to 5 ms, before performing spectral analysis with 1/3 octave resolution. In these TEOAE recordings, the typical noise floor was −12 dB.

DPOAEs at the frequency $2f_1$-f_2 were recorded using a primary frequency ratio $f_2/f_1=1.22$, with 1/3 octave frequency resolution. The acquisition was stopped by the tester after stabilization of the averaged DP gram. The typical noise floor varied between −8 and −15 dB, with the highest noise level at 1 kHz and the lowest at 3–4 kHz. The primary levels were set at $L_1=65$ dB and $L_2=55$ dB in a subset of 164 ears. In two other subsets different primary levels were used ($L_1=75$ dB and $L_2=70$ dB, and $L_1=L_2=70$ dB), to check their relative effectiveness for the diagnosis of hearing impairment. The first choice of primary levels was significantly more effective than the other two, as regards the DPOAE capability of discrimination between audiometric classes, confirming the findings of Stover *et al.* (1996a). Therefore, both for brevity and to avoid confusion, in this study we will present only the data of the subjects tested for DPOAEs using 65–55 dB primary levels. Kummer *et al.* (1998) showed that their $L_1=0.4L_2+39$ dB paradigm permits one to obtain good discrimination between normal and impaired ears, particularly at low stimulus levels, using the DP level and

growth rate as indicators. Recently, much effort has been put into the study of the optimal DPOAE stimuli, defined as those eliciting the maximum DP level. Johnson *et al.* (2006) demonstrated that, for each frequency, there is an optimal choice of primary levels and frequency ratio, capable of eliciting higher level DPOAEs than with the rule proposed by Kummer *et al.* (1998). In this study we used the 65–55 dB paradigm to allow comparison with previous important cross-section studies (Gorga *et al.*, 1997; Dorn *et al.*, 1998) in which this paradigm had been used. A data selection rule was further applied to exclude records in which the actually used stimulus was different from the target stimulus by more than 2 dB, to reduce the OAE level variability due to fluctuations of the stimulus level. After this selection, a subset of 160 ears with TEOAE, 65–55 dB DPOAE, and audiometric recordings matching all the criteria was identified.

These 160 ears were first divided into three classes according to the maximum hearing threshold level: NORM, with threshold $\leqslant$10 dB at all audiometric frequencies, MHL, with threshold $\leqslant$20 dB at all audiometric frequencies and $>$10 dB at least at one frequency, and HL, with threshold $>$20 dB at least at one frequency. In this study, a different audiometric classification has also been used, in which the ears were classified according to the average value of the hearing threshold in the 1–3 kHz frequency range, defining three new classes (AVN if the average threshold is $\leqslant$5 dB, AVM if it is $>$5 and $\leqslant$10 dB, and AVH if it is $>$10 dB). The above-mentioned limits have been arbitrarily chosen to get sample sizes roughly similar to those of the previous three classes. This classification proved to work better than that based on the maximum hearing threshold, as regards the OAE discrimination capability between different classes. Exposure data, which have not been considered here, could be used to get a better classification of the subjects, to assess the different OAE sensitivity to hearing impairment due to different kinds of noise. In this study we chose to show the main results in the simplest way, also to preserve the statistical size of the classes.

The average TEOAE SNRs and the DPOAE levels in each frequency band were chosen as the basic parameters to be correlated with the three classes of hearing loss in this preliminary cross-section study. This choice was motivated by previous studies, which had initially shown (Gorga *et al.*, 1993b) that OAE SNR performs slightly better than other quantities (OAE level or reproducibility) as regards their correlation with hearing threshold, while later the same authors argued that the best indicator was the DPOAE level (Gorga *et al.*, 2000).

A composite OAE indicator was investigated, capable of providing a good correlation with the individual hearing threshold. The effectiveness of a dichotomous test for the early detection of mild hearing loss levels based on this indicator was estimated by the ROC curve method (see, e.g., Gorga *et al.*, 1993b). The ROC curve is a plot describing the quality of a test reporting the hit rate (fraction of hearing impaired ears correctly evaluated as positive by the test) against the rate of false alarms (fraction of the normal hearing ears erroneously evaluated as positive). The curve is obtained by varying the test threshold (the level of some OAE parameter under which the null hypothesis (normal hearing) is negated, i.e., the test is positive) within a sufficiently large range. By increasing the value of the test threshold both the hit rate and the false alarm rate increase, for a good test the first one increasing much more rapidly than the second one. A useless test is represented by a ROC curve that cuts in two halves the plot along its diagonal, a perfect test leads to a right-angle curve that runs along the left and upper sides of the plot. The area below the curve is a quantitative measure of the power of the test, varying from 0.5 for a random output test to unity for a perfectly deterministic test. The procedure may be repeated for different values of the other audiometric threshold level chosen to define hearing impairment, to evaluate the sensitivity and specificity of the dichotomous test to different levels of hearing impairment.

III. RESULTS AND DISCUSSION

A. Discrimination between audiometric classes using TEOAE SNR and DPOAE level

The average audiometric profiles of the three audiometric classes, HL (20 ears), MHL (63 ears), and NORM (77 ears) are shown in the top panel of Fig. 1. The averages of the distributions of the individual thresholds within each class at each audiometric frequency are plotted. Error bars represent one standard error, to provide a visual hint of the significance of the average threshold difference between classes. The average threshold of the HL class is above 20 dB at 4 kHz only, and the difference between continuous classes is typically smaller than 10 dB. Despite these small differences, a statistical comparison of the average TEOAE and DPOAE spectra of the three classes shows significant differences, as shown in the bottom panels of Fig. 1 and in Table I. Indeed, a significant statistical OAE level difference may be easily found, not only between the NORM and HL classes, but also between the NORM and the MHL classes, separated by a small average audiometric threshold difference, and that would both fall in the normal hearing category applying the clinical standard criterion for hearing impairment (HL$>$20 dB).

The average TEOAE SNR is shown in the mid panel of Fig. 1 for the three classes, with error bars corresponding to the standard error. The statistical significance of these differences is synthesized in Table I by the value of the student's t-test probability, labeling as "not significant" (n.s.) the p values above the conventional significance threshold $p=0.05$. The same probability is reported for the distributions of the audiometric thresholds, for comparison. From Table I it is clear that the TEOAE capability of statistical discrimination between the three classes is good below 3 kHz.

In the bottom panel of Fig. 1 we show the average DPOAE level of the three classes NORM, MHL, and HL, in 1/3 octave bands. As is commonly done in clinical DPOAE studies, the frequency axis, here and in the following, reports the value of f_2, the highest of the primary frequencies. As remarked earlier, this is just the frequency corresponding to the cochlear place where the first component of the DP is generated by nonlinear distortion. A second linear reflection component is assumed to come from the cochlear place reso-

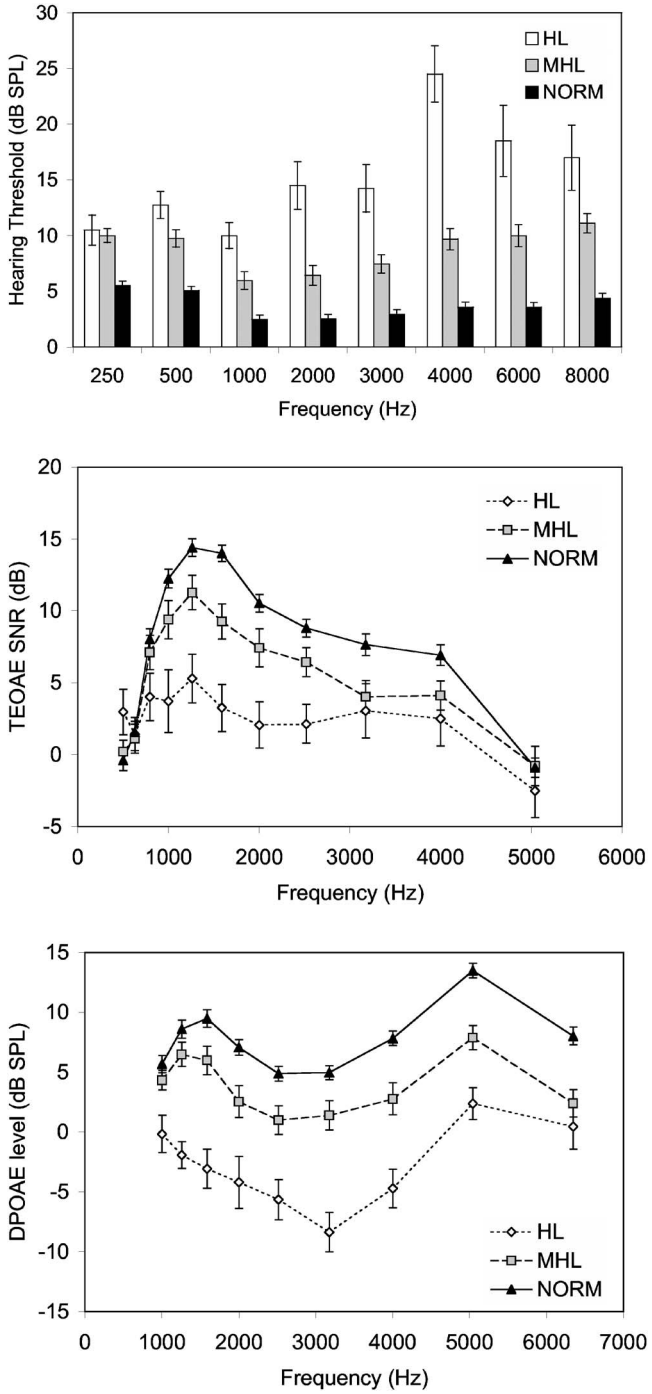


FIG. 1. Top panel: Average audiometric threshold profile of the three classes of ears HL (20 ears), MHL (63), and NORM (77). Error bars correspond to 1 standard error. This classification, which is based on the worst hearing threshold at any frequency, implies that most of the difference between the classes is at $f \geqslant 3$ kHz; middle panel: Average TEOAE SNR in 1/3 octave bands for the three audiometric classes HL, MHL, and NORM. Good discrimination capability of TEOAEs is evident in the frequency range below 3 kHz. Bottom panel: Average DPOAE level in 1/3 octave bands and audiometric threshold for the three audiometric classes NORM, MHL, and HL. The discrimination capability of DPOAEs is evident in the whole frequency range, even if most of the audiometric difference between the classes is concentrated at the highest frequencies.

nant at the f_{DP} frequency, thus, in principle, the response at frequency f_{DP} should not necessarily be correlated only to the audiometric threshold at the corresponding f_2 frequency.

The good discrimination capability of 65–55 dB DPOAEs among classes NORM, MHL, and HL over the whole frequency range is also shown in Table I, where the statistical significance of the DPOAE distribution difference between the classes is evaluated by the student's t-test. From Table I it may be appreciated that the DPOAE level discrimination capability between these classes is comparable to that of audiometry.

The other two DPOAE stimulus paradigms (75–70 and 70–70 dB) proved less effective than the 65–55 dB paradigm. For brevity, and to avoid confusion, only the data from the subjects analyzed with the 65–55 dB stimuli will be discussed here. We just mention that the 75–70 dB paradigm performed reasonably well at $f > 3$ kHz only, while the 70–70 dB paradigm showed poor sensitivity at all frequencies.

From Fig. 1 and Table I, it is possible to note that also the TEOAE and DPOAE discrimination capability is very good in the midfrequency range, despite the fact that the classes are separated according to the highest audiometric threshold, so most of the audiometric difference is at high frequency, as shown in the top panel of Fig. 1. The trend shown in the top of Fig. 1 is typical of mild noise-induced hearing loss in young workers. Most of the difference between the classes is concentrated in the high frequency tail of the audiogram. Thus the clinical definition of hearing impairment (based on the maximum value of the threshold) tends to separate the classes according to audiometric threshold differences at $f \geqslant 3$ kHz. On the other hand, the OAE sensitivity to hearing loss is expected and observed to be frequency specific, and it is typically highest in a lower frequency range, particularly for TEOAEs. For this reason the alternative classification based on the average threshold level in the 1–3 kHz range was introduced. In this alternative classification, the AVN class includes all ears (96) whose average audiometric threshold between 1 and 3 kHz is $\leqslant 5$ dB, the AVM those (42 ears) between 5 and 10 dB, and the AVH class those (22 ears) with average threshold >10 dB.

For each audiometric frequency, the average of the audiometric threshold distribution of the three classes AVN, AVM, and AVH is shown in the top panel of Fig. 2. This classification leads to a more balanced distribution of the audiometric differences between contiguous classes, with the highest differences (always lower than 10 dB) concentrated in the 1–4 kHz range. The classification based on the average threshold is less sensitive to the audiometric measurement uncertainties than that based on the maximum threshold level, as regards both random errors and quantization errors, which are particularly large if a standard stimulus step of 5 dB is used, as done in this study. This new classification could be useful to assess the sensitivity of composite indicators based on the OAE levels averaged over a wide frequency band, where the sensitivity is highest, as we will do in Sec. III B.

As shown in Fig. 2 and Table II, the class discrimination capability between the two extreme classes of TEOAE SNR and DPOAE level in the 1–3 kHz range is slightly better if the ears are classified according to the average hearing threshold in this range. This is not surprising, because, as shown in the top panels of Figs. 1 and 2, the difference

TABLE I. Statistical significance of the difference between the three classes HL, MHL, and NORM, expressed by the p value of the student's t-test, for the 1/3 octave TEOAE SNR, the DPOAE level, and the audiometric threshold.[a]

Frequency (kHz)	1	1.25	1.6	2	2.5	3.2	4	5	6.4
TEOAE (80 dB)									
$p_{\text{NORM-HL}}$	10^{-3}	5×10^{-5}	3×10^{-6}	9×10^{-5}	4×10^{-4}	n.s.	5×10^{-2}	n.s.	n.s.
$p_{\text{NORM-MHL}}$	2×10^{-2}	6×10^{-3}	3×10^{-5}	10^{-2}	2×10^{-2}	4×10^{-3}	2×10^{-2}	n.s.	n.s.
$p_{\text{MHL-HL}}$	3×10^{-2}	5×10^{-3}	3×10^{-3}	9×10^{-3}	2×10^{-2}	n.s.	n.s.	n.s.	n.s.
DPOAE (65–55)									
$p_{\text{NORM-HL}}$	10^{-2}	10^{-4}	8×10^{-5}	4×10^{-5}	2×10^{-7}	5×10^{-6}	5×10^{-6}	7×10^{-5}	3×10^{-3}
$p_{\text{NORM-MHL}}$	n.s.	n.s.	5×10^{-3}	6×10^{-4}	10^{-3}	6×10^{-4}	3×10^{-5}	10^{-5}	6×10^{-4}
$p_{\text{MHL-HL}}$	5×10^{-2}	10^{-3}	3×10^{-3}	8×10^{-3}	3×10^{-4}	3×10^{-4}	2×10^{-3}	3×10^{-2}	n.s.
Frequency (kHz)	1			2		3	4		6
Audiometry									
$p_{\text{NORM-HL}}$	8×10^{-6}			4×10^{-5}		10^{-4}	3×10^{-7}		4×10^{-4}
$p_{\text{NORM-MHL}}$	10^{-5}			8×10^{-6}		3×10^{-7}	3×10^{-11}		8×10^{-11}
$p_{\text{MHL-HL}}$	7×10^{-3}			2×10^{-3}		10^{-2}	3×10^{-5}		3×10^{-2}

[a]n.s.=not significant.

between the average audiometric thresholds in the 1–3 kHz range is also slightly higher among the classes AVN, AVM, and AVH. To account for this, the average TEOAE and DPOAE differences between the first and the third class and the corresponding average audiometric threshold difference are plotted in Fig. 3 as a function of frequency. This comparison provides only partial information about the sensitivity of these techniques, because the different test-retest uncertainty of the three techniques should be taken into account, as done by Hall and Lutman (1999). However, Fig. 3 shows that the sensitivity of TEOAEs is good only below 2 kHz, while the DPOAE perform very well also at higher frequencies. This may be partly due to the fact that, above 4 kHz, the TEOAE SNR is rather low, so the data are not fully reliable. Moreover, the acquisition window cuts off a significant part of the response at frequency higher than 4 kHz, whose latency is typically lower than 4 ms (e.g., Sisto and Moleti, 2002). The two panels, which are obtained using the two different classifications (NORM, MHL, HL; AVN, AVM, AVH), show that the second classification permits one to better exploit the frequency specificity of the DPOAE measurements. Indeed, the DPOAE sensitivity to hearing loss in the 4–6 kHz range is lower if the ears are classified according to the maximum threshold level, which is typically at these frequencies. This may be an indication of the fact that the frequency specificity of DPOAEs is a complicated issue, as we have already pointed out, because the DPOAE elicited by f_1 and f_2 is also sensitive to hearing loss at frequencies lower than f_2, due to the existence of a DPOAE second source at the f_{DP} place, which is apical to the f_2 place.

B. OAE-based indicators of mild hearing threshold shift

Despite these promising results, confirming and extending those reported by Lucertini *et al.* (2002), the intersubject variability of the TEOAE SNR at a given level of the audiometric hearing threshold is very high, and makes it very difficult to design a test for hearing loss based on the 1/3 octave band TEOAE SNR level. This is shown in Fig. 4, where TEOAE SNR in 1/3 octave bands is plotted against the hearing threshold at the same frequency, for 1, 2, 3, and 4 kHz, for all ears. Some decreasing trend may be observed, but the correlation is poor, due to the vertical dispersion of the data. The particularly low correlation at 3 to 4 kHz may be due to the fact that the TEOAE SNR was rather low at high frequency. It is therefore possible that the diagnostic potential of TEOAEs at high frequency was underestimated by our measurements due to the noise floor of the measurements.

The individual correlation between the DPOAE level in a given 1/3 octave band and the corresponding audiometric threshold at the same frequency is reported in Fig. 5, as previously done for TEOAEs in Fig. 4. The correlation with the audiometric threshold is significantly better than that of TEOAEs, but still unsatisfactory, due to the large vertical spread of the DP levels in ears with the same audiometric threshold. The best correlation is observed at 2 kHz, in agreement with the estimates of Tables I and II, and of Fig. 3.

All the above-presented results on the TEOAE SNR and the DPOAE level confirm the strengths and the weaknesses of OAEs as a hearing diagnostic tool, previously demonstrated by Lucertini *et al.* (2002) and several other cross-section studies (e.g., Attias *et al.*, 1995; Hall and Lutman, 1999). They demonstrate indeed only the OAE capability of discriminating between populations of several tens of ears. The real possibility of evaluating the hearing functionality of a single ear with a single OAE test is seriously limited by the large dispersion of the OAE levels within each population, whose effect on the standard error is limited in the cumulative analysis by averaging over a large number of ears. This problem is evident in Figs. 4 and 5, where the scatter plots of the individual relation between hearing threshold and OAE measurements are shown for the TEOAE SNR and for the DPOAE level. These plots are useful to evaluate the different sensitivity of the OAE parameters in the frequency bands, which is dependent on the slope of the best fit curves to the data and on the data dispersion along the ordinate (OAE) axis. The slopes in Figs. 4 and 5 are generally less steep than those reported by Hall and Lutman (1999), and similar to those reported by Gorga *et al.* (1993b), while the indication

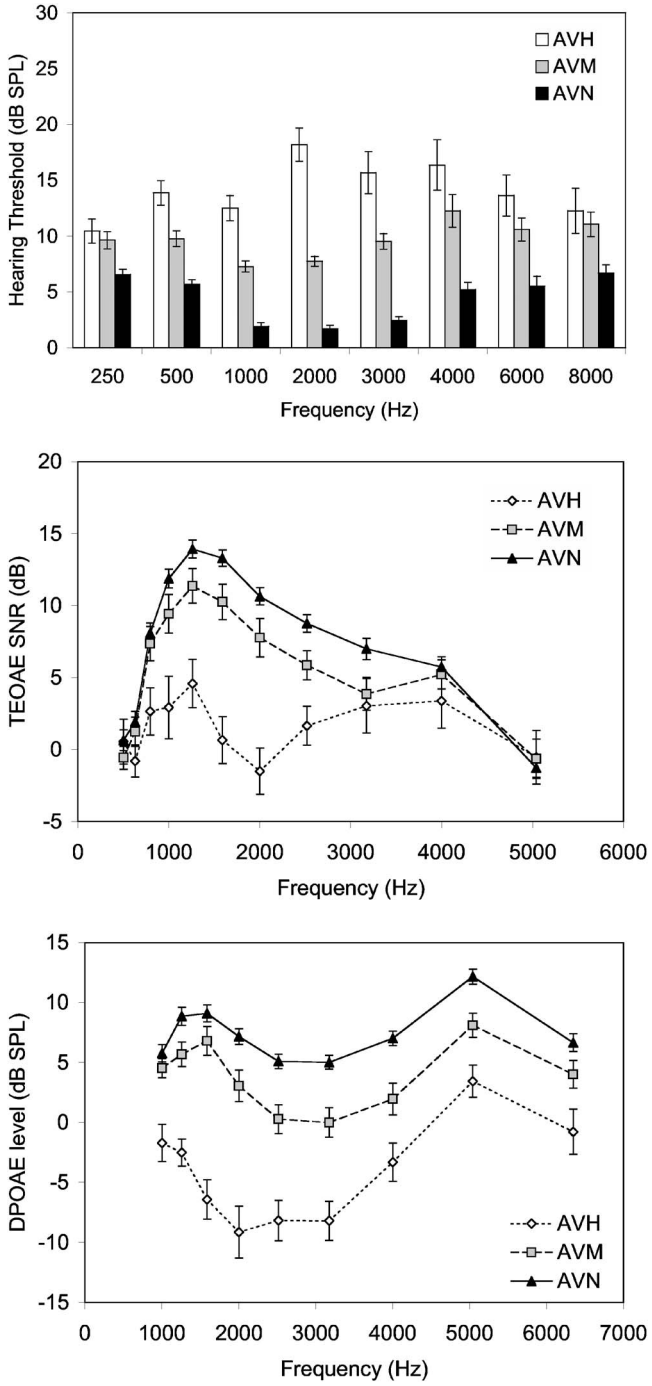


FIG. 2. Average audiometric threshold, TEOAE SNR and DPOAE level of the three classes, as in Fig. 1, for the classes of ears AVH (22 ears), AVM (42), and AVN (96). Error bars represent 1 standard error. This classification is based on the average hearing threshold level in the 1–3 kHz bands, where the OAE techniques are very sensitive. The OAE discrimination capability among these three classes is higher than in Fig. 1, particularly at the lower frequencies.

that DPOAE slopes are generally steeper than those of TEOAEs and the large vertical data dispersion are confirmed.

The frequency specificity of the OAE response and its sensitivity to the first stages of hearing loss could suggest using the response at a given frequency as a sensitive test of the hearing threshold at that specific frequency. In Fig. 6 the power of a clinic decision test for hearing threshold higher than 10 dB at a given frequency, based on the TEOAE SNR (top panel) or on the DPOAE level (bottom), is expressed by the area below the corresponding ROC curves. The ROC curves of Fig. 6 show that the vertical dispersion shown in Figs. 4 and 5 limits the OAE diagnostic potential: only in the 2 kHz DPOAE case reasonably good performances may be obtained. These results may be compared with previous cross-section studies: Gorga *et al.* (1993a, b) and Prieve *et al.* (1993) showed ROC curves for DPOAE and TEOAE tests, for different choices of the hearing threshold cut-off. They found that the OAE tests were most effective for threshold shifts between 20 and 30 dB, with areas under the ROC curve above 0.9, particularly at frequencies higher than 2 kHz. The results of the present study show that good test performances could be extended to the subclinical hearing threshold range, but the vertical dispersion of the OAE data needs to be reduced.

The causes of the TEOAE and DPOAE level variability must be understood to fully exploit the possibilities of the OAE techniques. In the DPOAE case, the interference between the two DP sources may be responsible for this dispersion. In fact, especially in normal hearing or mild hearing loss ears, a large amplitude fine structure has been observed in the high-resolution DP grams (Mauermann *et al.*, 1999). The amplitude of the DP level fine structure is often larger than 10 dB, tends to disappear in cases of sudden temporary hearing loss to appear again after recovery, and, for this reason, the very presence of this fine structure has also been proposed as a sensitive indicator for normal hearing threshold (Mauermann *et al.*, 1999). The fine structure becomes a source of uncertainty when DP measurements at a single frequency are considered as an indicator of hearing loss in the same frequency range. In fact, the 1/3 octave DP gram cannot resolve the structure, but each test frequency may randomly coincide with a maximum or a minimum of the DP fine structure, with very different results, especially in the case of mild hearing loss ears, which still show a large amplitude fine structure. Without exploring more complicated and time-consuming techniques, which could overcome this problem exploiting the different phase behavior of the two sources (Mauermann and Kollmeier, 2004), or exploring a narrow frequency range around each test frequency, we decided to reduce the effect of this phenomenon by simply taking a fast DP gram with a moderate frequency resolution (1/3 octave) and averaging the DPOAE levels over a suitably large frequency band, to include at least four or five independent DP levels. A similar approach had been proposed by Gorga *et al.* (2000), who used three adjacent points of a 1/2-octave DP gram in a multivariate analysis. This procedure guarantees a substantial reduction of the uncertainty associated with the presence of fine structure, because each point has the same probability of hitting a maximum or a minimum, at the cost of accepting a poorer frequency resolution. It also reduces the systematic error associated with the presence of SOAEs near one of the DPOAE frequencies, which may cause suppression of the DP response in normal-hearing ears.

In the case of the TEOAE SNR (the same considerations also apply to reproducibility or response) the uncertainty

TABLE II. Statistical significance of the difference between the three classes AVN, AVM, and AVH, expressed by the p value of the student's t-test, for the 1/3 octave TEOAE SNR, the DPOAE level, and the audiometric threshold.[a]

Frequency (kHz)	1	1.25	1.6	2	2.5	3.2	4	5	6.4
TEOAE (80 dB)									
$p_{\text{AVN-AVH}}$	6×10^{-4}	10^{-5}	10^{-6}	2×10^{-7}	2×10^{-4}	n.s.	n.s.	n.s.	n.s.
$p_{\text{AVN-AVM}}$	n.s.	4×10^{-2}	5×10^{-3}	10^{-2}	7×10^{-3}	3×10^{-2}	n.s.	n.s.	n.s.
$p_{\text{AVM-AVH}}$	10^{-2}	10^{-3}	8×10^{-5}	2×10^{-5}	2×10^{-2}	n.s.	n.s.	n.s.	n.s.
DPOAE (65–55)									
$p_{\text{AVN-AVH}}$	6×10^{-4}	4×10^{-6}	9×10^{-7}	2×10^{-8}	10^{-9}	4×10^{-7}	10^{-4}	9×10^{-4}	6×10^{-3}
$p_{\text{AVN-AVM}}$	n.s.	10^{-2}	n.s.	6×10^{-4}	4×10^{-5}	5×10^{-5}	2×10^{-4}	4×10^{-3}	n.s.
$p_{\text{AVM-AVH}}$	6×10^{-3}	5×10^{-4}	2×10^{-6}	2×10^{-6}	10^{-5}	5×10^{-4}	4×10^{-2}	n.s.	n.s.
Frequency (kHz)	1			2		3	4		6
Audiometry									
$p_{\text{AVN-AVH}}$	3×10^{-9}			2×10^{-10}		5×10^{-7}	8×10^{-5}		4×10^{-4}
$p_{\text{AVN-AVM}}$	6×10^{-14}			2×10^{-17}		6×10^{-13}	4×10^{-5}		3×10^{-4}
$p_{\text{AVM-AVH}}$	2×10^{-4}			5×10^{-7}		5×10^{-3}	n.s.		n.s.

[a]n.s.=not significant.

may partly be due also to differences in the spectral shape of the actual stimulus. The coupling of the probe to the ear canal crucially affects the stimulus wave form, and, consequently, its spectral content in a given band may vary significantly even if the stimulus level is kept constant. This extrinsic source of uncertainty may be limited by the skill and experience of the operator, but a too strong request of optimizing the coupling could lead to increasing the test time significantly, which is not always possible in the occupational safety practice. It is not easy to account for the spectral shape of the stimulus in the data analysis phase, due to the nonlinear nature of the cochlear response. An intrinsic source of intersubject variability of the TEOAE response at a given frequency in subjects with the same hearing threshold is also the presence of resonant TEOAEs, sometimes coincident with measurable SOAEs, whose contribution dominates the TEOAE response (Sisto *et al.*, 2001). These peculiar features occur in most normal-hearing ears at random individual frequencies, and the response in the corresponding frequency bands is higher. Of course, there is a correlation of the presence of SOAEs (and, more generally, high-Q resonances) with the minima of the hearing threshold, but the relation with the audiogram is altered by the fact that the audiogram itself is a collection of narrow band audiometric tests, and that the standard audiometric frequencies do not generally coincide with the frequencies of these peculiar cochlear resonances. Therefore, the same reasoning applied to DPOAEs suggests averaging the TEOAE responses over suitably large frequency bands, in order to limit the fluctuations of the TEOAE SNR. As in the DPOAE case, this suggestion is different from the obvious concept of increasing the SNR by averaging, because the above-mentioned fluctuations are not entirely due to instrumental random noise, as confirmed by the high values of the OAE SNR in normal hearing or mild hearing loss subjects in the 1–4 kHz range. All the above-noted considerations led us to look for a composite OAE-based indicator of the hearing threshold level in a mild hearing loss range, based on band-averaged DPOAE levels and/or TEOAE SNR. On the other hand, the audiogram itself is also a collection of sparse narrow band measurements, and the presence of an unresolved fine structure in the audiometric threshold (e.g., Mauermann *et al.*, 2004) suggests, for the

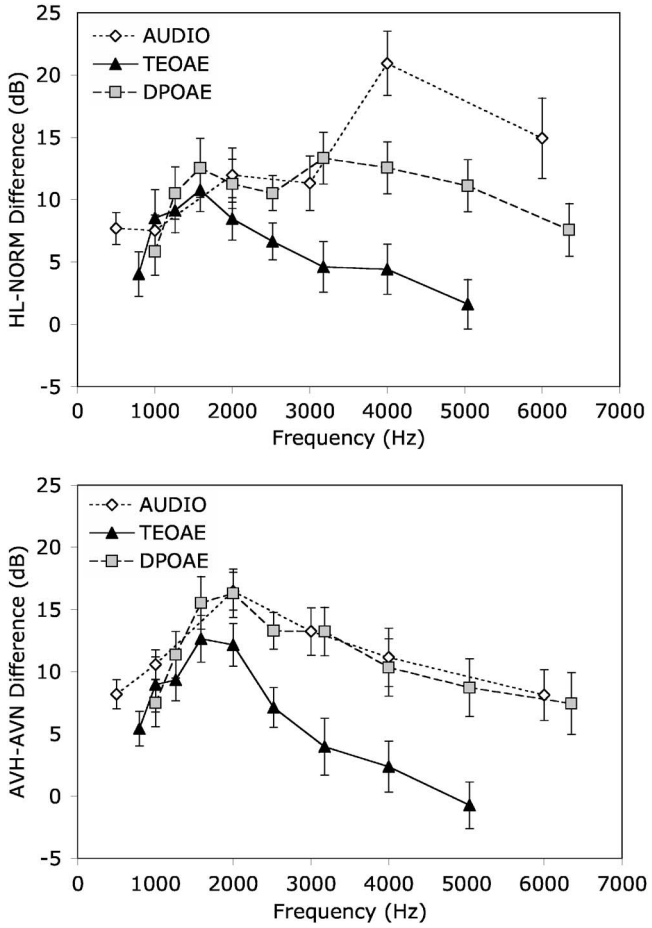


FIG. 3. Absolute value of the difference between the classes HL and NORM (top) and between the classes AVH and AVN (bottom) for the 1/3 octave band average TEOAE SNR and DPOAE level, and for the audiometric threshold. This plot shows that the TEOAE sensitivity to hearing threshold is comparable to that of audiometry only at low frequency ($f<3$ kHz), while DPOAE sensitivity is extended over the whole frequency range above 1 kHz. Using the classification based on the average threshold level in the 1–3 kHz band, the sensitivity of DPOAEs is very similar to that of audiometry.

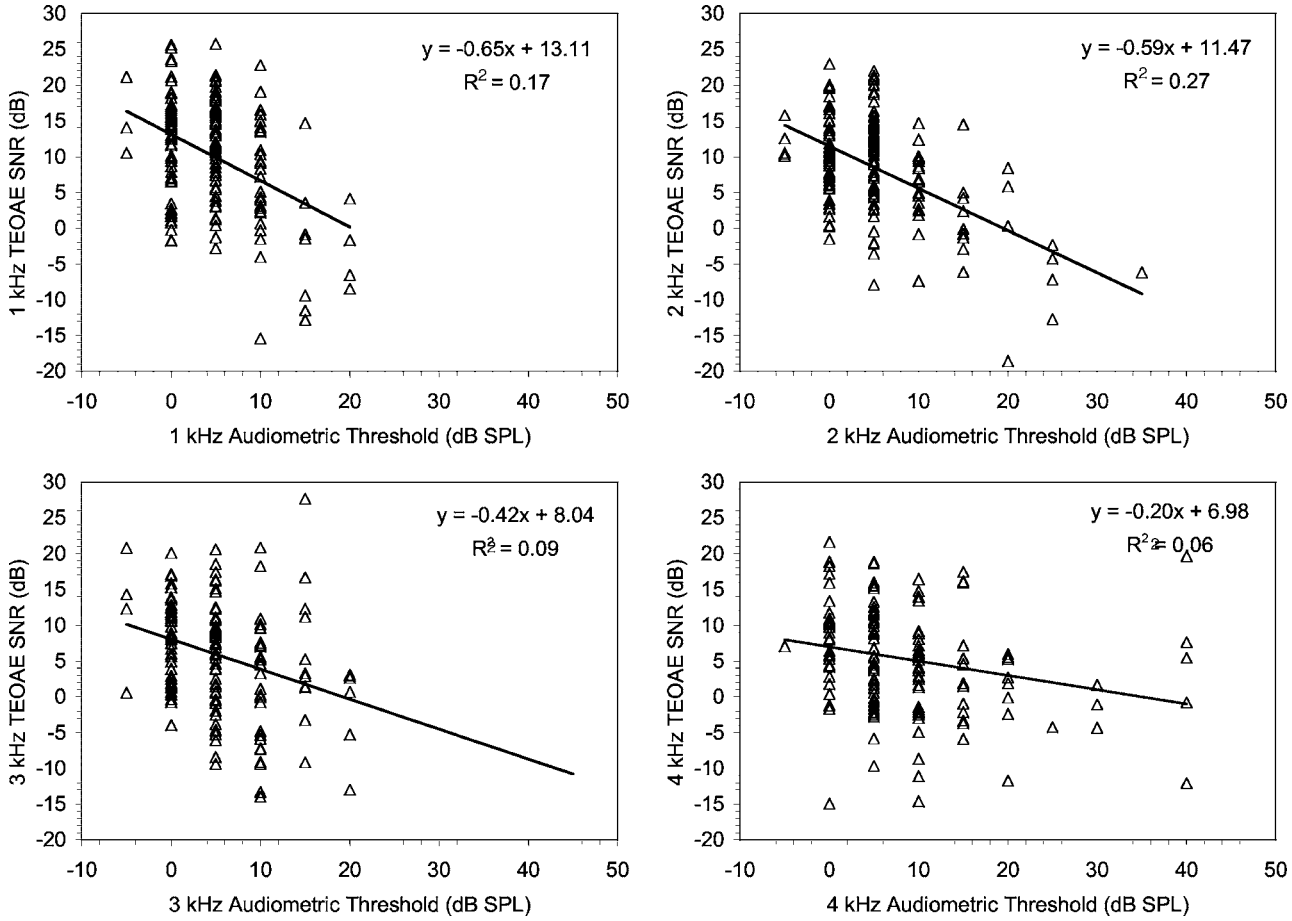


FIG. 4. Correlation between audiometric threshold and 1/3 octave band TEOAE SNR. Despite the rather good results shown in Figs. 1 and 2, relative to averages over tens of ears, the individual correlation between TEOAE SNR and audiometric threshold is generally rather poor.

same reasons, to average also a few audiogram points before attempting to find a correlation with frequency-averaged OAE levels, particularly for low levels of hearing loss.

From Fig. 2 and Table II, we observe that the DPOAE levels between 1 and 3.2 kHz show a good discrimination capability between contiguous audiometric populations classified according to the average 1–3 kHz threshold level. As discussed earlier in this paper, the second DPOAE source is at the cochlear place resonant at $f_{DP}=2f_1-f_2$, thus it would be possible that the DPOAE frequency band, which is expressed in terms of f_2, should be slightly shifted with respect to the audiometric band. In this study, we have chosen to neglect this possibility, for simplicity, and allow a direct comparison with previous studies, in which it is generally assumed that the DP level is associated with the threshold at frequency f_2.

The correlation between the DPOAE level averaged between 1 and 3.2 kHz and the average 1–3 kHz hearing threshold is shown in Fig. 7. As expected, averaging reduces both the statistical fluctuations and those due to the DPOAE an audiometric fine structure, and the correlation between these frequency-averaged quantities (42%) is better than that shown in Fig. 5. A dichotomous test for average hearing threshold higher than 10 dB would have a reasonably small rate of errors of Type I and II, as shown by the corresponding ROC curve (area=0.93) in Fig. 8.

Starting from this result, one could investigate whether the inclusion of TEOAE SNR data from the same ears could lead to the design of a more effective composite DP-TEOAE indicator for mild hearing loss. In Fig. 9 we show the correlation between the TEOAE SNR averaged in the 1–3.2 kHz range and the average 1–3 kHz audiometric threshold. The result is far from being satisfactory, but significantly better than that of Fig. 4. The corresponding ROC curve expressing the power of a test of mild average 1–3 kHz hearing loss (average threshold higher than 10 dB) based on the 1–3.2 kHz TEOAE SNR is shown in Fig. 10. Figure 10 shows that such a test alone would have a poorer screening capability (area=0.86), even if it should be mentioned that the average hearing threshold shift of 10 dB is also quite small, so the test-retest fluctuations of audiometry itself could partly affect this result.

A composite DPOAE-TEOAE indicator could be in principle more effective than that based on a single OAE technique if the two OAE parameters provided complementary information on the hearing functionality. The correlation between TEOAE SNR and DPOAE level for these ears resulted to be 62%. The correlation is not particularly high, indicating that the information is only partially coincident, so we checked if the two OAE quantities could complement each other in a more effective composite indicator.

The simplest composite TEOAE-DPOAE indicator is

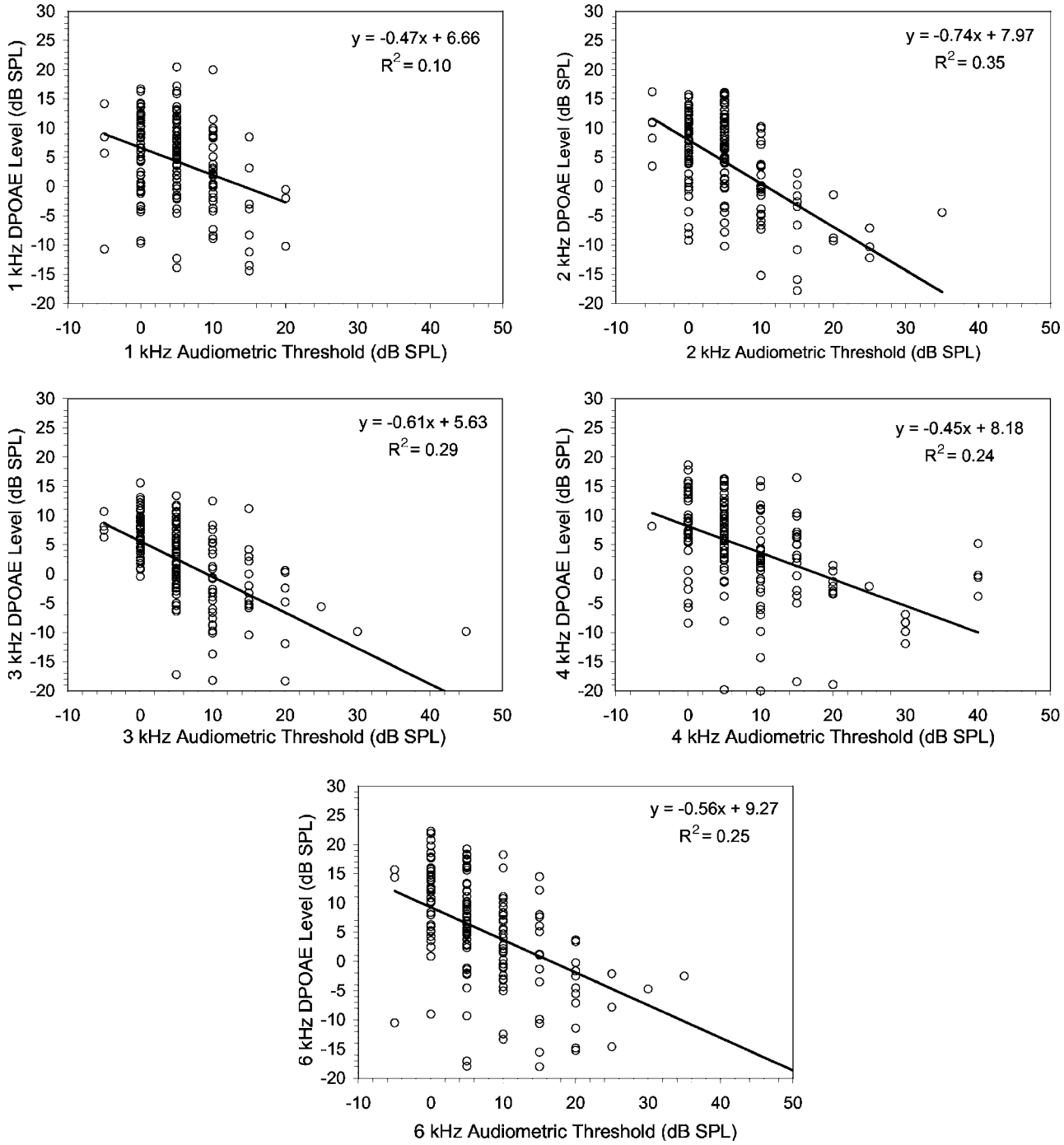
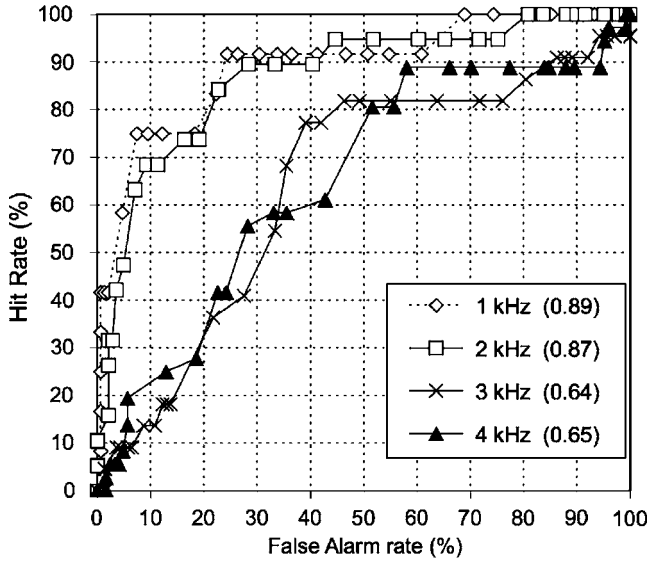


FIG. 5. Correlation between audiometric threshold and 1/3 octave band DPOAE level. The DPOAE level is better correlated to the audiometric threshold than the TEOAE SNR (Fig. 4), but the vertical dispersion of the data is still very large.

the average of the two quantities. The correlation between this composite indicator and the hearing threshold is slightly lower (40%) than that of the single DPOAE average level. It is evident from the corresponding ROC curve of Fig. 11 that this composite indicator provides no improvement (area =0.91) with respect to the result obtained using the DPOAE data only. This negative result suggests that the lower sensitivity of the TEOAE test makes it useless to include the TEOAE SNR in a composite indicator, at least in the very simple way proposed here.

Returning to the DPOAE-based indicator, we show in Fig. 12 a family of ROC curves relative to tests aimed at discriminating different levels of the average hearing threshold. Each curve refers to a test in which the "impaired" population includes all ears whose 1–3 kHz average audiometric thresholds exceed a threshold of, respectively, 3, 5, 8, 10, and 12 dB. The area below the curve, which gives a measure of the performance of the test, is also reported. It is clear that the test becomes effective for average hearing threshold levels well below the 20 dB limit, which is conventionally used for defining hearing impairment.

These results are sufficiently encouraging, but the following objection could be raised: the 1–3 kHz range was chosen because it had showed the best correlation with the



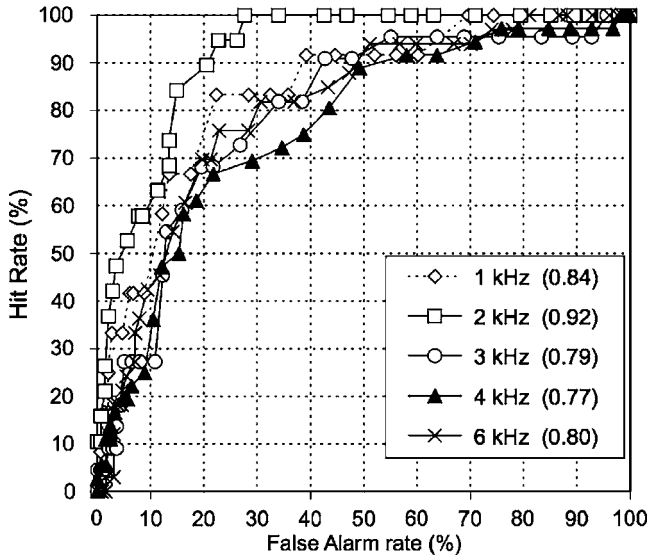


FIG. 6. ROC curves for a mild hearing loss test (average hearing threshold in the 1–3 kHz range higher than 10 dB) based on the 1/3 octave TEOAE SNR (top panel) and DPOAE level (bottom panel). The area under the curve is reported for each frequency as a measure of the power of the test.

hearing threshold, but early noise-induced hearing loss is more often observed at 4 kHz. Also in the data that we have used in this study, the 4 kHz frequency is that showing the highest average hearing threshold shift in the impaired ears. Thus, a test sensitive to the damage in the 1–3 kHz range only could not be as sensitive to the early stages of hearing impairment as suggested by Fig. 12. One could argue that it would be better to include the 4 kHz audiometric threshold level in the average audiometric threshold and look for a correlation with the corresponding band DPOAE average level. This is done in Fig. 13, where the correlation between these two average quantities is shown, and Fig. 14, where the ROC curves for a test for average 1–4 kHz hearing threshold higher than thresholds varying from 5 to 20 dB. The correlation in Fig. 13 is similar to that of Fig. 7, but the ROC curves show that for a threshold of 10 dB the performance is significantly worse (area=0.87) than that of the 1–3 kHz test. Rather good test performances seem to be possible

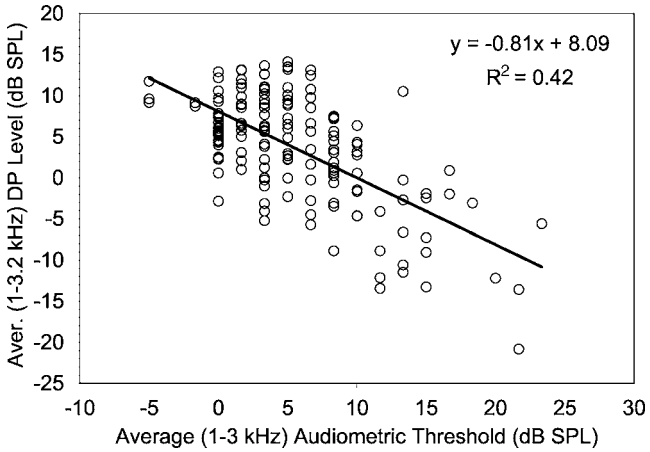


FIG. 7. Correlation between the DPOAE level, averaged in the 1–3.2 kHz range, and the average 1–3 kHz audiometric threshold. Averaging improves the correlation, losing frequency specificity.

only at higher levels of the audiometric test threshold (15–20 dB), but in the present database the number of ears with average threshold above 15 dB is too low to use the area under the ROC curve as a reliable indicator of the power of the test.

We wish to remark that all the above-mentioned results refer to a "single hit" test, which is applied to a subject for the first time, without any other information about the subject's previous OAE levels. Part of the variability of the OAE levels, implying loss of correlation with the audiometric threshold level, comes from the accidental coincidence of the test frequencies with SOAEs or with minima or maxima of the DPOAE fine structure. In this case, our proposal of averaging over a large bandwidth reduces the problem to some extent. Part of the OAE level intersubject variability is instead associated with intrinsic systematic differences between subjects with the same hearing threshold level, and in this case averaging cannot help. On the other hand, this last

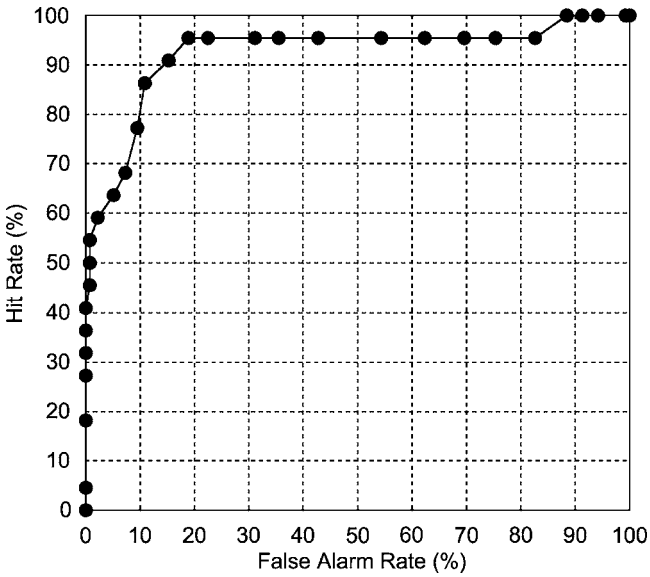


FIG. 8. ROC curve for a test of mild 1–3 kHz hearing loss (average threshold higher than 10 dB) based on the average 1–3.2 kHz DPOAE level. The area under the curve is 0.93.

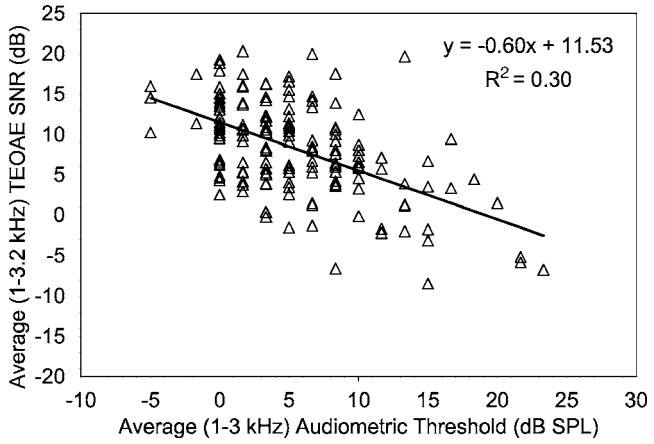


FIG. 9. Correlation between the TEOAE SNR averaged in the 1–3.2 kHz range and the average 1–3 kHz audiometric threshold. Also for TEOAE SNR, averaging over a wide frequency band improves the correlation significantly, but the performance of TEOAEs is worse than that of DPOAEs, as regards both the slope and the correlation.

problem would not affect the diagnostic potential of an OAE test for the detection of changes in the hearing functionality of a single subject during his lifetime exposure to noise. For this reason, the good results obtained in this cross-section study are also a strong motivation for the longitudinal study that is being conducted on the same subjects, because they indicate that the diagnostic power of OAE-based tests could be particularly important for the monitor of populations exposed to noise.

IV. CONCLUSIONS

This study confirms previous findings demonstrating the sensitivity and specificity of OAEs for the detection of hearing loss, with a particular emphasis on very mild hearing loss. Unique features of this study are (1) that the subjects have been carefully selected to ensure that hearing loss was mostly (if not only) due to exposure to noise, and (2) that the hearing loss levels of the subjects were typically very low, due to the young age and short occupational exposure, and, in many cases, to the use of hearing protector devices. The difficulties in the development of an OAE-based mild hearing loss test, coming from the large intersubject variability of the OAE levels, have been analyzed, and a simple solution has been proposed. Both the standard audiometric and DPOAE measurements are a collection of sparse narrowband measurements, which cannot resolve the existing spectral fine structure, while high-resolution techniques are too time-consuming to be used in the occupational safety prac-

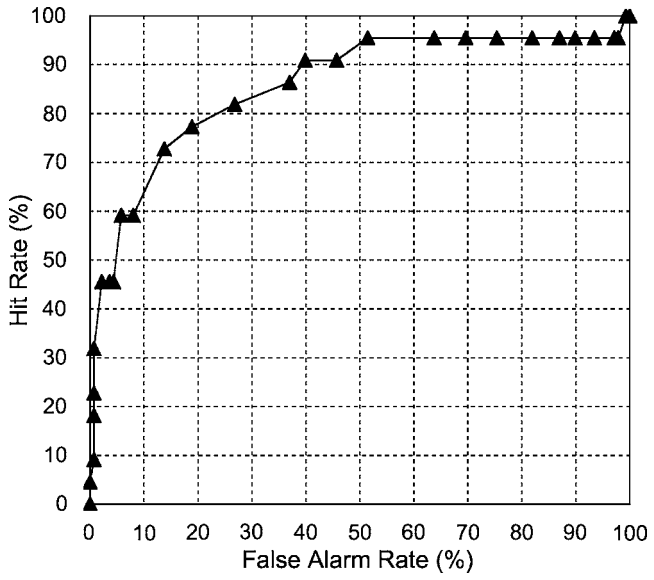


FIG. 10. ROC curve for a test of 1–3 kHz mild hearing loss (average threshold higher than 10 dB) based on the 1–3.2 kHz TEOAE SNR. The area under the curve is 0.86.

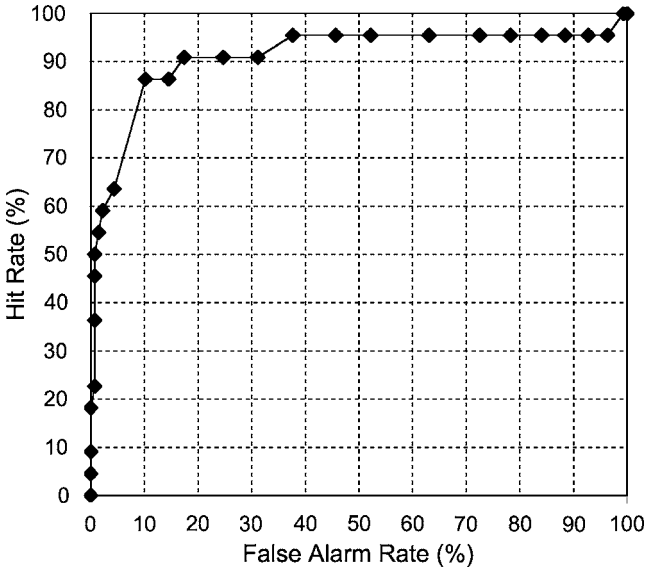


FIG. 11. ROC curve for a test of 1–3 kHz mild hearing loss (average threshold >10 dB) based on the average between the 1–3.2 kHz DPOAE level and the 1–3.2 kHz TEOAE SNR. There is no improvement with respect to the result of Fig. 7, obtained with the DPOAE data only (the area under the curve is 0.91).

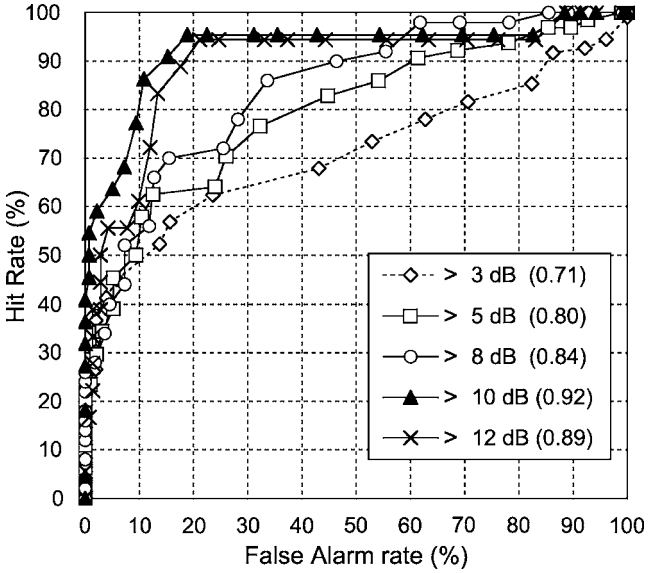


FIG. 12. ROC curves for a test of mild 1–3 kHz hearing loss, based on the average 1–3.2 kHz DPOAE level, for different levels of the audiometric threshold (3, 5, 8, 10, 12 dB) used as the "impairment" test threshold. Good test performances are obtained starting from a 10 dB threshold.

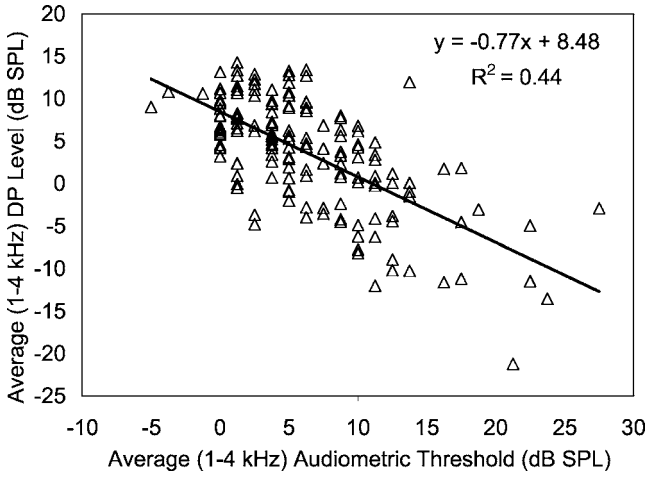


FIG. 13. Correlation between the 1–4 kHz audiometric hearing threshold and the 1–4 kHz DPOAE level.

tice. This observation, and the sensitivity of the DPOAE and TEOAE levels to the presence of SOAEs and resonant TEOAEs, suggested that averaging over a rather large bandwidth is necessary to get a good correlation between OAEs and audiometry. The preliminary data analysis, and some basic knowledge of the OAE generation mechanisms, helped understanding which OAE frequency band should be used for a test. The sensitivity and specificity of TEOAE- and DPOAE-based tests have been evaluated using OAE and audiometric quantities averaged over a large bandwidth. In the case of DPOAEs, a slight frequency shift could be introduced between the audiogram and the DP gram expressed in terms of f_2, to account for the fact that only the first DPOAE source is assumed to come from the cochlear place associated with the perception of the frequency f_2. This possibility has not been analyzed in the present study, because more data would be necessary to show a statistical difference between two slightly shifted wide band indicators. Good test performance was found using the DPOAE data only, and inclusion of the information from TEOAEs added no predictive power to the test. Such a test would be sensitive to average hearing loss levels in the 1–3 kHz range as low as 10 dB. This result may have relevant implications in the occupational safety practice, because the standard DPOAE technique is very fast, and the diagnosis of these very low hearing threshold levels could be useful for the early detection of cochlear damage in workers exposed to noise.

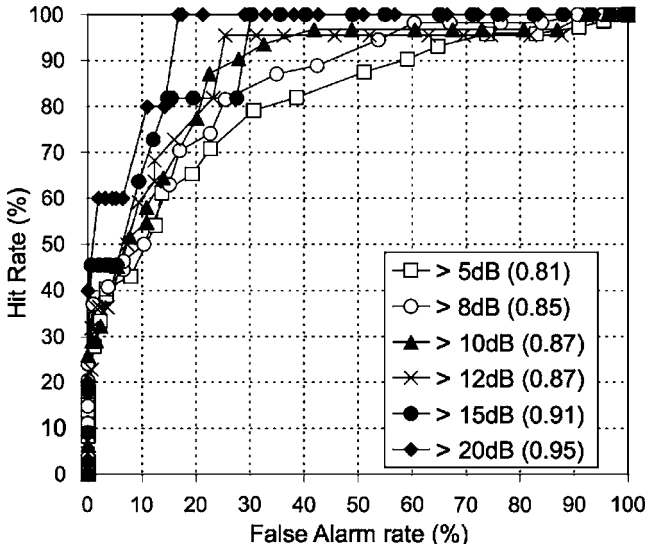


FIG. 14. ROC curves for a test of mild 1–4 kHz hearing loss, based on the average 1–4 kHz DPOAE level, for different levels of the audiometric threshold (5, 8, 10, 12, 15, 20 dB) used as the "impairment" test threshold. Good test performances seem to be obtained starting from a 15 dB threshold, but the subjects with average threshold above 15 dB are too few to allow quantitative analyses.

Attias, J., Furst, M., Furman, V., Reshef, I., Horowitz, G., and Bresloff, I. (**1995**). "Noise-induced otoacoustic emission loss with or without hearing loss," Ear Hear. **16**, 612–618.

Boege, P., and Janssen, T. (**2002**). "Pure-tone threshold estimation from extrapolated distortion product otoacoustic emission I/O-functions in normal and cochlear hearing loss ears," J. Acoust. Soc. Am. **111**, 1810–1818.

Bonfils, P., Piron, J. P., Uziel, A., and Pujol, R. (**1988**). "A correlative study of evoked otoacoustic emission properties and audiometric thresholds," Arch. Otorhinolaryngol. Suppl. **245**, 53–56.

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.

Dorn, P. A., Piskorski, P., Keefe, D. H., Neely, S. T., and Gorga, M. P. (**1998**). "On the existence of an age/threshold/frequency interaction in distortion product otoacoustic emissions," J. Acoust. Soc. Am. **104**, 964–971.

Furst, G., Maurer, J., and Schlegel, J. (**1995**). "Monitoring ototoxic side effects in streptomycin therapy of tubercolosis patients with transitory evoked otoacoustic emissions TEOAE," Pneumologie **49**, 590–595.

Furst, M., Reshef, I., and Attias, J. (**1992**). "Manifestations of intense noise stimulation on spontaneous otoacoustic emission and threshold microstructure: experiment and model," J. Acoust. Soc. Am. **91**, 1003–1013.

Gorga, M. P., Dierking, D. M., Johnson, T. A., Beauchaine, K. L., Garner, C. A., and Neely, S. T. (**2005**). "A validation and potential clinical application of multivariate analyses of distortion-product otoacoustic emission data," Ear Hear. **26**, 593–607.

Gorga, M. P., Neely, S. T., Bergman, B. M., Beauchaine, K. L., Kaminski, J. R., Peters, J., and Jesteadt, W. (**1993**). "Otoacoustic emissions from normal-hearing and hearing-impaired subjects: distortion product responses," J. Acoust. Soc. Am. **93**, 2050–2060.

Gorga, M. P., Neely, S. T., Bergman, B. M., Beauchaine, K. L., Kaminski, J. R., Peters, J., Sculte, L., and Jesteadt, W. (**1993**). "A comparison of transient-evoked and distortion product otoacoustic emissions in normal-hearing and hearing-impaired subjects," J. Acoust. Soc. Am. **94**, 2639–2648.

Gorga, M. P., Neely, S. T., and Dorn, P. A. (**1999**). "Distortion product otoacoustic emission test performance for a priori criteria and for multi-frequency audiometric standards," Ear Hear. **20**, 345–362.

Gorga, M. P., Neely, S. T., Ohlrich, B., Hoover, B., Redner, J., and Peters, J. (**1997**). "From laboratory to clinic: A large scale study of distortion product otoacoustic emissions in ears with normal hearing and ears with hearing loss," Ear Hear. **18**, 440–455.

Gorga, M. P., Nelson, K., Davis, T., Dorn, P. A., and Neely, S. T. (**2000**). "Distortion product otoacoustic emission test performance when both $2f_1$-f_2 and $2f_2-f_1$ are used to predict auditory status," J. Acoust. Soc. Am. **107**, 2128–2135.

Hall, A. J., and Lutman, M. E. (**1999**). "Methods for early identification of noise induced hearing loss," Audiology **38**, 277–280.

Hine, J. E., and Thornton, A. R. D. (**2005**). "Transient evoked otoacoustic emissione recorded using maximum length sequences from patients with sensorineural hearing loss," Hear. Res. **203**, 122–133.

Hussain, D. M., Gorga, M. P., Neely, S. T., Keefe, D. H., and Peters, J. (**1998**). "Transient evoked otoacoustic emissions in patients with normal hearing and in patients with hearing loss," Ear Hear. **19**, 434–449.

Jedrzejczak, W. W., Blinowska, K. J., and Konopka, W. (**2005**). "Time-frequency analysis of transiently evoked otoacoustic emissions of subjects exposed to noise," Hear. Res. **205**, 249–255.

Johnson, T. A., Neely, S. T., Kopun, J. G., and Gorga, M. P. (**2006**). "Reducing reflected contributions to ear-canal distortion product otoacoustic emissions in humans," J. Acoust. Soc. Am. **119**, 3896–3907.

Kalluri, R., and Shera, C. A. (**2001**). "Distortion-product source unmixing:

A test of the two-mechanism model for DPOAE generation," J. Acoust. Soc. Am. **109**, 622–637.
Konrad-Martin, D., Neely, S. T., Keefe, D. H., Dorn, P. A., and Gorga, M. P. (**2001**). "Source of distortion product otoacoustic emissions revealed by suppression experiments and inverse fast Fourier transforms in normal ears," J. Acoust. Soc. Am. **109**, 2862–2879.
Konrad-Martin, D., Neely, S. T., Keefe, D. H., Dorn, P. A., Cyr, E., and Gorga, M. P. (**2002**). "Sources of DPOAEs revealed by suppression experiments, inverse fast Fourier transforms, and SFOAEs in impaired ears," J. Acoust. Soc. Am. **111**, 1800–1809.
Kummer, P., Janssen, T., and Arnold, W. (**1998**). "The level and growth behavior of the 2f1-f2 distortion-product otoacoustic emission and its relationship to auditory sensitivity in normal hearing and cochlear hearing loss," J. Acoust. Soc. Am. **103**, 3431–3444.
Lapsley Miller, J. A., 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.
Long, G. R., and Tubis, A. (**1988**). "Modification of spontaneous and evoked otoacoustic emissions and associated psychoacoustic microstructure by aspirin consumption," J. Acoust. Soc. Am. **84**, 1343–1353.
Lucertini, M., Moleti, A., and Sisto, R. (**2002**). "On the detection of early cochlear damage by otoacoustic emission analysis," J. Acoust. Soc. Am. **111**, 972–978.
Mauermann, M., and Kollmeier, B. (**2004**). "Distortion product otoacoustic emission (DPOAE) input/output functions and the influence of the second DPOAE source," J. Acoust. Soc. Am. **116**, 2199–2212.
Mauermann, M., Long, G. R., and Kollmeier, B. (**2004**). "Fine structure of hearing threshold and loudness perception," J. Acoust. Soc. Am. **116**, 1066–1080.
Mauermann, M., Uppenkamp, S., van Hengel, P. W. J., and Kollmeier, B. (**1999**). "Evidence for the distortion product frequency place as a source of distortion product otoacoustic emission (DPOAE) fine structure in humans. II. Fine structure for different shapes of cochlear hearing loss," J. Acoust. Soc. Am. **106**, 3484–3491.
Moleti, A., and Sisto, R. (**2003**). "Objective estimates of cochlear tuning by otoacoustic emission analysis," J. Acoust. Soc. Am. **113**, 423–429.
Moleti, A., Sisto, R., Tognola, G., Parazzini, M., Ravazzani, P., and Grandori, F., (**2005**). "Otoacoustic emission latency, cochlear tuning, and hearing functionality in neonates," J. Acoust. Soc. Am. **118**, 1576–1584.
Moleti, A., Sisto, R., and Lucertini, M. (**2002**). "Linear and nonlinear transient evoked otoacoustic emission in human exposed to noise," Hear. Res. **174**, 290–295.
Prieve, B. A., Gorga, M. P., Schmidt, A., Neely, S., Peters, J., Schultes, L., and Jesteadt, W. (**1993**). "Analysis of transient-evoked otoacoustic emissions in normal-hearing and hearing-impaired ears," J. Acoust. Soc. Am. **93**, 3308–3319.
Probst, R., Lonsbury-Martin, B. L., and Martin, G. K. (**1991**). "A review of otoacoustic emissions," J. Acoust. Soc. Am. **89**, 2027–2067.
Schloth, E. (**1983**). "Relation between spectral composition of spontaneous oto-acoustic emissions and fine-structure of threshold in quiet," Acustica **53**, 250–256.
Shera, C. A., and Guinan, J. J. Jr. (**1999**). "Evoked otoacoustic emissions arise from two fundamentally different mechanisms: A taxonomy for mammalian OAEs," J. Acoust. Soc. Am. **105**, 782–798.
Sisto, R., and Moleti, A. (**2002**). "On the frequency dependence of the otoacoustic emission latency in hypoacoustic and normal ears," J. Acoust. Soc. Am. **111**, 297–308.
Sisto, R., Moleti, A., and Lucertini, M. (**2001**). "Spontaneous otoacoustic emissions and relaxation dynamics of long decay time OAEs in audiometrically normal and impaired subjects," J. Acoust. Soc. Am. **109**, 638–647.
Stover, L., Gorga, M. P., Neely, S. T., and Montoya, D. (**1996a**). "Toward optimizing the clinical utility of distortion product otoacoustic emission measurements," J. Acoust. Soc. Am. **100**, 956–967.
Stover, L., Neely, S. T., and Gorga, M. P. (**1996b**). "Latency and multiple sources of distortion product otoacoustic emissions," J. Acoust. Soc. Am. **99**, 1016–1024.
Yates, G. K., and Withnell, R. H. (**1999**). "The role of intermodulation distortion in transient-evoked otoacoustic emissions," Hear. Res. **136**, 49–64.

Representation of the vowel /ε/ in normal and impaired auditory nerve fibers: Model predictions of responses in cats

Muhammad S. A. Zilany and Ian C. Bruce[a)]
Department of Electrical and Computer Engineering, McMaster University, Hamilton, Ontario L8S 4K1, Canada

(Received 5 December 2006; revised 16 March 2007; accepted 5 April 2007)

The temporal response of auditory-nerve (AN) fibers to a steady-state vowel is investigated using a computational auditory-periphery model. The model predictions are validated against a wide range of physiological data for both normal and impaired fibers in cats. The model incorporates two parallel filter paths, component 1 (C1) and component 2 (C2), which correspond to the active and passive modes of basilar membrane vibration, respectively, in the cochlea. The outputs of the two filters are subsequently transduced by two separate functions, added together, and then low-pass filtered by the inner hair cell (IHC) membrane, which is followed by the IHC-AN synapse and discharge generator. The C1 response dominates at low and moderate levels and is responsible for synchrony capture and multiformant responses seen in the vowel responses. The C2 response dominates at high levels and contributes to the loss of synchrony capture observed in normal and impaired fibers. The interaction between C1 and C2 responses explains the behavior of AN fibers in the transition region, which is characterized by two important observations in the vowel responses: First, all components of the vowel undergo the C1/C2 transition simultaneously, and second, the responses to the nonformant components of the vowel become substantial. © *2007 Acoustical Society of America.* [DOI: 10.1121/1.2735117]

PACS number(s): 43.64.Sj, 43.64.Bt, 43.64.Pg, 43.64.Wn [WPS] Pages: 402–417

I. INTRODUCTION

Toward the goal of understanding the mechanisms by which speech signals are represented in the brain, auditory-nerve (AN) fiber responses to speech-like sounds, especially for vowels, have been investigated by several researchers over the last few decades (Young and Sachs, 1979; Delgutte, 1980; Sinex and Geisler, 1983; Delgutte and Kiang, 1984; Palmer *et al.*, 1986; Miller *et al.*, 1997; Wong *et al.*, 1998). A number of different cochlear nonlinearities prove important in encoding vowel spectra in both normal and impaired ears at sound pressure levels spanning the dynamic range of hearing. Responses to vowels presented at high sound pressure levels are interesting because they highlight some very striking nonlinearities in the normal cochlea and provide insight into the ability of amplification to restore normal vowel representations in the impaired cochlea (Miller *et al.*, 1997; Wong *et al.*, 1998; Schilling *et al.*, 1998; Miller *et al.*, 1999a, b). Some recent studies (Sachs *et al.*, 2002; Bruce *et al.*, 2003) have modeled AN responses to vowels presented at low to moderate stimulus levels with a sufficient accuracy but fail to account for a number of features of high stimulus level responses observed in the physiological experiments. Holmes *et al.* (2004) modeled vowel responses in guinea pigs and made comparisons with some of the cat data, but the guinea pig vowel data are substantially less extensive than the cat vowel data. In particular, no guinea pig vowel data have been published for high presentation levels or for impaired ears. In this paper, we have investigated the temporal aspects of cat AN fiber coding of a steady-state vowel in normal and impaired ears using a computational cat auditory-periphery model (Zilany and Bruce, 2006), which is suitable for representing responses to simple and complex stimuli such as tones, two-tone, and broadband stimuli across a wide range of sound pressure levels.

An important observation in normal fibers at sound pressure levels corresponding to conversational speech is that the response is phase locked almost exclusively to the formant frequency closest to the fiber's characteristic frequency (CF),[1] a phenomenon referred to as synchrony capture (Young and Sachs, 1979; Deng and Geisler, 1987b; Miller *et al.*, 1997). However, as the sound level increases to around 80 dB SPL or above, fibers start to show synchrony to other stimulus frequency components, i.e., the responses of the fibers with CFs near one formant now also respond to other formants, particularly to the lower-frequency formants and to the distortion products of the formants (Wong *et al.*, 1998). This is presumably due to broadening of the auditory filters with stimulus level and nonlinearity of the cochlea. At very high sound pressure levels, fibers respond in a broadband fashion, meaning that synchrony to frequencies other than formants also becomes substantial. The degradation in the temporal representation (i.e., the spread of synchrony) is also a characteristic of responses to the vowel in impaired fibers in cats (Miller *et al.*, 1997; Wong *et al.*, 1998; Schilling *et al.*, 1998). Thus, the normal tonotopic representation of the vowel is substantially degraded in impaired and normal fibers at high levels.

The spread of synchrony and the synchrony capture phenomenon itself reflect nonlinear signal processing in the cochlea. Another nonlinear change in AN fiber responses seen at high levels is the component 1 (C1)-component 2 (C2)

a)Electronic mail: ibruce@ieee.org

 0001-4966/2007/122(1)/402/16/$23.00

transition, which is characterized by a sharp change in the phase-level function and occasionally an accompanying dip in the rate-level function. Wong *et al.* (1998) reported that the C1/C2 transition occurs in AN fiber responses to a vowel as well as to tones. However, they behave differently in that in response to a vowel all components undergo the C1/C2 transition at approximately the same vowel level. This observation rules out the possibility of separate processing of the individual components of the vowel (Wong *et al.*, 1998). On the other hand, individual tones cause AN fibers to undergo C1/C2 transition in a fairly narrow range of levels (Liberman and Kiang, 1984; Wong *et al.*, 1998). Consequently, to reach C2 threshold (C1/C2 transition level), the level required by a particular frequency component in vowel could be significantly lower than the level required for a tone at that frequency.

A number of previous attempts had been made to predict AN responses to speech-like stimuli (e.g., Deng and Geisler, 1987a; Geisler, 1989; Jenison *et al.*, 1991; Sachs *et al.*, 2002; Bruce *et al.*, 2003; Holmes *et al.*, 2004). The cat vowel responses we attempt to describe in this paper were previously studied by Bruce *et al.* (2003) using a model that incorporates many of the details of the signal processing in the cochlea and has scaling functions in the outer hair cell (OHC) and inner hair cell (IHC) sections of the model to facilitate simulation of impairment in the cochlea. The model by Bruce *et al.* (2003) gives good qualitative and quantitative descriptions of the vowel responses at low to moderate levels only. This is expected, because this model includes only a single mode of vibration to the IHC that produces the C1 responses only. To address high level effects such as the C1/C2 transition, peak splitting, and the shift in the best frequency (BF)[1] with level, Zilany and Bruce (2006) significantly extended the model of Bruce *et al.* (2003) by including a C2 filter, parallel to the signal-path C1 filter. The selection of the C2 filter followed by a separate transduction function (referred to as the C2 transduction function) has been motivated by some physiological observations in AN fibers [for details see Zilany and Bruce (2006)]. The output of the C2 filter can be regarded as a secondary passive mode of vibration of the basilar membrane in the cochlea, whereas the output of C1 filter resembles the primary active mode of vibration. This model not only better describes AN fiber responses at high levels but also gives insight into the possible physical mechanisms of the C2 response and the C1/C2 interaction. Another improvement made in Zilany and Bruce (2006) that contributes to accurate representations at high levels is that the C1 filter, originally a gammatone filter in Bruce *et al.* (2003), has been replaced by a chirp filter, which is responsible for the level-independent instantaneous frequency glide in the impulse response of AN fibers (Carney *et al.*, 1999; Tan and Carney, 2003). Thus, the model can simulate the shift in BF that occurs at high levels, which is partially responsible for the loss of synchrony capture seen at high levels. The increased accuracy of the model predictions in this paper also results from the fact that we have taken into account the frequency response of the sound delivery system used in the physiological studies. This has been done to closely match the vowel's spectral shape at the tympanic membrane, as was presented during physiological experiments.

By using accurate cochlear models and providing quantitative simulations, we can predict how certain underlying mechanisms may affect the perception of speech. In addition, a more accurate model of cochlear processing will lead to a deeper understanding of how the various physiological changes account for the change in the neural responses associated with hearing impairment. The ability to break down a single hearing loss to its constituent components may enable us to design strategies that could restore the normal neural representation for listeners with hearing loss. In this paper, we have investigated the various AN fiber nonlinearities seen in the cat vowel responses using a computational auditory-periphery model (Zilany and Bruce, 2006). The accuracy of the model predictions are assessed by comparison with the reported physiological data in cats (Miller *et al.*, 1997; Wong *et al.*, 1998). The model responses to other stimuli such as tones and broadband noise are described in Zilany and Bruce (2006). In Sec. II, the vowel stimuli are described and the model is explained briefly, along with a description of how the acoustic trauma from the physiological studies is modeled. The model predictions are described in Sec. III and compared with the published physiological data.

II. METHODS

The representation of the synthesized vowel /ɛ/ in cat AN fibers (Miller *et al.*, 1997; Wong *et al.*, 1998) has been investigated in this paper using an auditory-periphery model developed by Zilany and Bruce (2006). To be consistent with the physiological experiments, the same stimuli and response analysis methods have been utilized here.

A. Stimuli

Following the methods used in the physiological studies, we have filtered the vowel /ɛ/ (Wong *et al.*, 1998) and the stimulus "besh" (Miller *et al.*, 1997) by the head related transfer function (HRTF) of the human head (Wiener and Ross, 1946). Note that the synthesized syllable "besh" has the vowel spectrum identical to that of the isolated vowel, and the responses are investigated only in the steady-state vowel portion. The HRTF-filtered stimulus is then passed through a filter that approximately matches the frequency response of the sound delivery system used in the experiments. This correction for the sound delivery system is necessary as the magnitude of the vowel in the F2 region (delivered to the tympanic membrane) is often several decibels lower than that of the nominal vowel (Wong *et al.*, 1998). The spectral representation of the vowel is shown in Fig. 1, where the thin line shows the example spectral envelope after compensating for the acoustics of the sound delivery system used in the physiological studies.

B. Model of the auditory-periphery

The auditory-periphery model used to simulate the responses to the vowel has been developed by Zilany and Bruce (2006) and is capable of generating realistic response

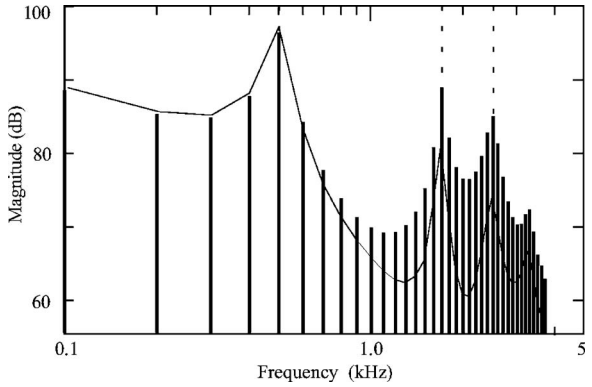


FIG. 1. Line spectrum of the synthesized vowel /ɛ/ after filtering by the HRTF of the human head. The fundamental frequency is 100 Hz, and the first three formants indicated by the vertical dashed lines are at 0.5, 1.7, and 2.5 kHz, respectively. A representative spectral envelope of the vowel, as delivered to the tympanic membrane, is shown by the thin line.

properties of the AN fibers in cats across a wide range of CFs and intensities spanning the dynamic range of hearing. A schematic diagram of the model is given in Fig. 2. Each section of the model provides a phenomenological description of the major functional components of the auditory-periphery, from the middle ear (ME) to the auditory nerve.

The first section models the filtering properties of the ME, which affects the relative levels of the components of wide-band stimuli and hence plays an important role in simulating responses to stimuli such as vowels. The input to the ME is the instantaneous pressure wave form of the stimulus in Pascals sampled at 500 kHz. The ME filter is followed by a signal-path C1 filter, which sets up the baseline tuning for the AN fiber. A feed-forward control path regulates the gain and bandwidth of the C1 filter to account for several level-dependent properties in the cochlea. The C1 filter has been designed in such a way that it can address a range of realistic response properties of the cochlea. The asymmetrical orientation of the poles and zeros of this filter causes inclusion of the instantaneous frequency (IF) glides in the impulse response of the AN, and thus the model is able to simulate the BF shift as a function of sound pressure level (Carney *et al.*, 1999; Tan and Carney, 2003). The tails of the tuning curves, which are important in modeling responses to multi-component stimuli, arise from the careful selection of the poles and zeros of this filter (Zilany and Bruce, 2006). The output of the C1 filter closely resembles the primary active mode of vibration of the basilar membrane (BM) in the cochlea.

A parallel-path C2 filter has been introduced as a second mode of excitation to the IHC and is critical for simulating the transition region effects at high levels. The C2 filter is linear, static, and is the same as the C1 filter with complete OHC impairment, i.e., the tuning of the C2 filter is the same as the broadest possible tuning of the C1 filter. For CFs >600 Hz, this tuning will also be the same as the C1 filter tuning in the normal model of the cochlea at high presentation levels, where the control signal has saturated. However, it is to be noted that for fibers with CFs <600 Hz, the frequency response of the C1 filter at high levels is dynamic in the normal model of the cochlea, changing cycle-by-cycle, and thus differs somewhat from the C2 filter. The two filters (C1 and C2) are followed by two separate transduction functions, referred to as the C1 and C2 transduction functions. The selection of these functions, consistent with Kiang's two factor cancellation hypothesis (Kiang, 1990), is such that at low levels the C2 output is significantly smaller than the C1 output, whereas at high levels the C2 output dominates, and the C1 and C2 responses are out of phase. At levels within the transition region, both outputs are approximately equal in magnitude and tend to cancel each other. In addition, the C2 response is not subject to rectification, unlike the C1 response at high levels, and consequently peak splitting results from the C1/C2 interaction (Kiang, 1990; Zilany and Bruce, 2006). Note that, unlike other models having parallel filter paths with static nonlinearities (e.g., Goldstein, 1990, 1995; Lin and Goldstein, 1995; Meddis *et al.*, 2001) where the interaction between filter paths gives rise to the tips and tails

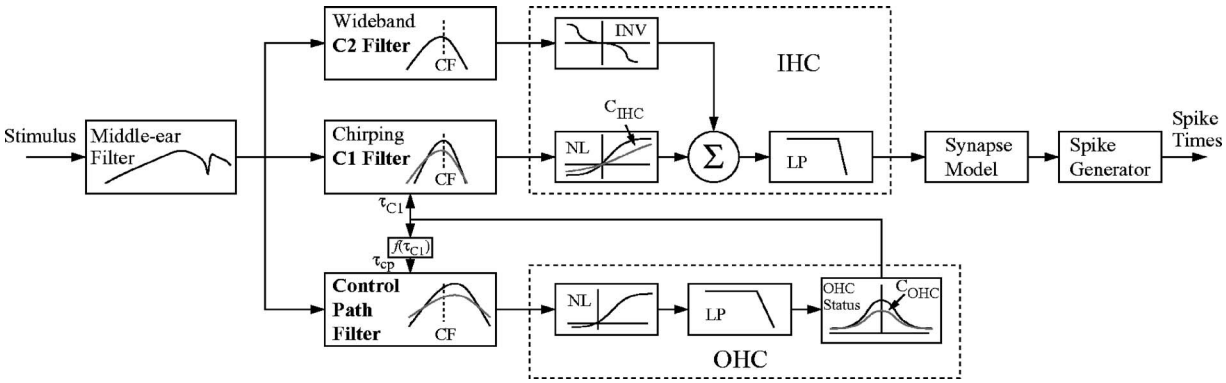


FIG. 2. Schematic diagram of the auditory-periphery model, reprinted from Zilany and Bruce (2006) with permission from the Acoustical Society of America (© 2006). The input to the model is an instantaneous pressure wave form of the stimulus in units of Pa and the output is the spike times in response to that input. The model has a middle-ear filter, a feed-forward control path, a signal-path C1 filter, and a parallel-path C2 filter, an inner hair-cell (IHC) section followed by a synapse model and a discharge generator. Abbreviations: outer hair cell (OHC), low-pass (LP) filter, static nonlinearity (NL), characteristic frequency (CF), inverting nonlinearity (INV). C_{OHC} and C_{IHC} are scaling constants that indicate OHC and IHC status, respectively. The bold and grey lines in the filter functions represent the tunings at low and high sound pressure levels, respectively. The wideband C2 filter shape is fixed and is the same as the broadest possible C1 filter. The bold and grey lines in the stage following the C1 filter (C1 transduction function) indicate the nonlinearity in the IHC in normal and impaired (scaled down according to C_{IHC}) conditions, respectively. The stage following the wideband C2 filter, referred to as the C2 transduction function, is an inverting nonlinearity and is insensitive to trauma.

of the tuning curves, the tips and tails of this model's tuning curves arise from the C1 filter alone. Only at very high levels (80 dB SPL or above) does the C2 tuning contribute substantially, as shown in Fig. 5 of Zilany and Bruce (2006). As described earlier, the C2 filter has approximately the same tuning as the C1 filter at these levels, so the C2 filter has little effect on the shape of the magnitude-frequency response of the model.

The summed output of the two transduction functions is then passed through a seventh-order IHC low-pass filter with a cut-off frequency of 3.8 kHz that describes the fall-off in pure tone synchrony with CF above 1 kHz. The IHC output drives the IHC-AN synapse which provides the instantaneous synaptic release rate as output. The synaptic gain has been adjusted such that the model produces responses of high-spontaneous rate (50 spikes/s before refractoriness), low-threshold fibers. Finally the AN discharge times are produced in the model by a renewal process that includes refractory effects. More details of the model can be found in Zilany and Bruce (2006), and model code is available from the authors on request.

1. Modification of the model

The model developed by Zilany and Bruce (2006) has been validated against a wide range of published physiological data. However, we have found that a slight adjustment in the cochlear amplifier (CA) gain improves the model's prediction of the vowel data of Wong *et al.* (1998). Their data show that the range of sound levels over which the C1/C2 transition occurs for tones is fairly narrow in the CF range 1–3 kHz. To obtain similar behavior from the model, the CA gains in this CF range have to be increased from the previously published values, whereas the gains at higher CFs need no further modification.

The CA gain (for cats) in Zhang *et al.* (2001) was chosen to vary as a function of CF in a fashion similar to the direct measurement of CA gain in guinea pig and chinchilla from BM data (Nuttall and Dolan, 1996; Cooper and Rhode, 1997; Ruggero *et al.*, 1997), which increases from 15 dB at low CFs to 70 dB at high CFs (shown by the dashed line in Fig. 3). Bruce *et al.* (2003) modified this function to better explain the change in Q_{10} values in impaired fibers in cats and the vowel responses in cats at high presentation levels. This was done by applying more CA gain in the CF range 500 Hz to 5 kHz, and also reducing the gain to 52 dB at high CFs [Appendix B of Bruce *et al.* (2003), and shown by the dotted line in Fig. 3]. With further increase in CA gain at lower CFs in this paper, this function agrees more with the estimates of CA gain versus frequency in humans, which is thought to be almost flat across frequency (Oxenham and Plack, 1997; Plack and Oxenham, 2000; Lopez-Poveda *et al.*, 2003). The modified CA gain function is shown by the solid line in Fig. 3, and is given by

$$\text{gain}_{\text{CA}}(\text{CF}) = \max\{15, 52(\tanh(2.2\log_{10}(\text{CF}/600) + 0.15) + 1.0)/2\}, \quad (1)$$

where CF has the units of hertz. However, it is to be noted that this modification does not substantially affect the basic response properties of the model AN fibers reported in Zilany and Bruce (2006).

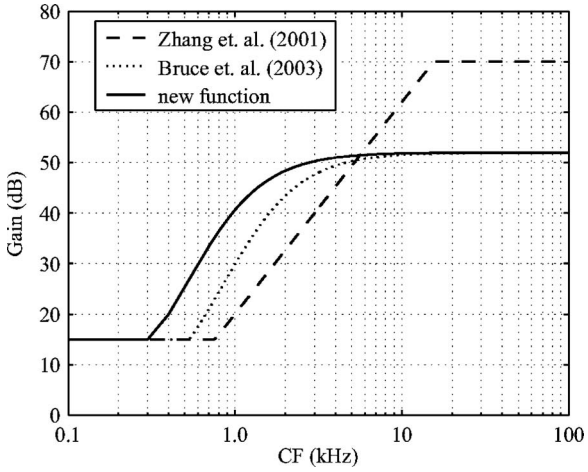


FIG. 3. Cochlear amplifier (CA) gain as a function of CF. The dashed line shows the gain function used in Zhang *et al.* (2001), the dotted line shows the gain used in Bruce *et al.* (2003), and the solid line indicates the gain function used in this paper.

2. Modeling a population of impaired fibers

To predict vowel responses for the population of AN fibers reported in Miller *et al.* (1997), the levels of OHC and IHC impairment, C_{OHC} and C_{IHC}, respectively, have been estimated as a function of frequency (normal or impaired CF[1]) to fit the impaired Q_{10}, impaired CF, and minimum threshold shift of the impaired pool of data. It is known that damage to the OHCs causes both broadening of the tuning and elevation of the threshold along with the shift in CF, whereas impairment in the IHCs primarily produces elevation of the tuning curve without substantial broadening (Liberman and Dodds, 1984). Following the method of Bruce *et al.* (2003), values of C_{OHC} have been chosen as a function of CF such that the model impaired fibers show approximately the same impaired Q_{10} and impaired CF as those of the cat data. Then an appropriate amount of impairment of the IHCs (C_{IHC}) has been applied to obtain the remaining threshold shift observed in the physiological data.

Figure 4(a) shows the impairment in the OHC as a function of CF that gives a close fit to the impaired Q_{10} and impaired CF of the physiological data. Normalized Q_{10} values of the experimental data and model are provided in Fig. 4(b) as a function of impaired CF. The threshold shift relative to the best threshold curve produced due to the impairment in the OHC alone is shown by the dashed line in Fig. 4(d). The remaining threshold shift is accounted for by impairment of the IHCs.

Figure 4(c) shows the functional relationship between C_{IHC} and impaired CF to obtain the remaining threshold shift that is not accounted for by the impairment of the OHCs. The combined threshold shift due to impairment of both OHCs and IHCs is shown by the solid line in Fig. 4(d). The IHC impairment is applied such that the model impaired fibers match the minimum threshold shift of the impaired pool of data. Even severe impairment of the OHCs in the model can only account for around two-thirds of the threshold shift seen

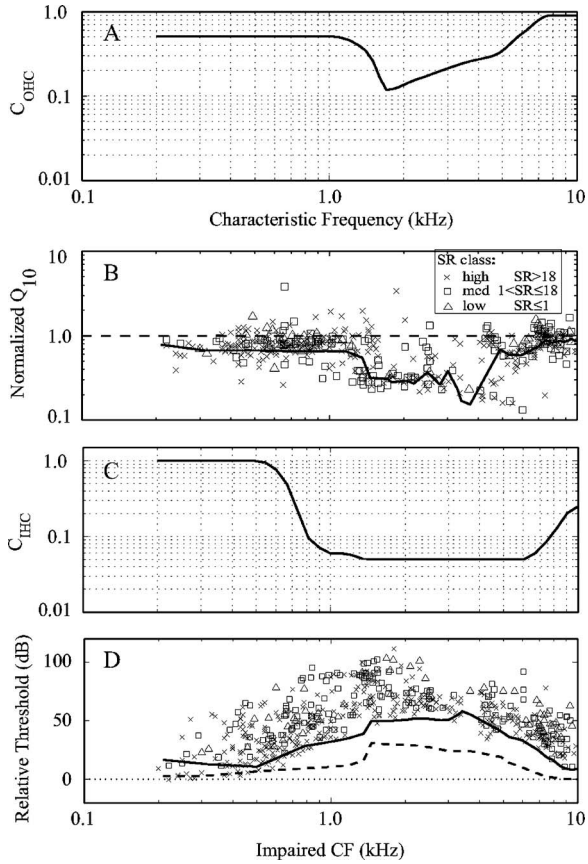


FIG. 4. Modeling of impaired fibers by applying impairment in the OHC (C_{OHC}) and IHC (C_{IHC}). (A) Functional relationship between C_{OHC} and CF (in the model) to fit the impaired Q_{10} and impaired CF of the physiological data. (B) Normalized Q_{10} vs impaired CF. Q_{10} values of the impaired fibers are normalized by their corresponding normal Q_{10} values. Here impaired CF corresponds to the model fiber after applying impairment in the OHC. Gray symbols indicate measured data (Miller *et al.*, 1997) from individual fibers with different spontaneous rates (SRs), and the solid line shows the model predictions for the function shown in (A). The dashed line represents the normalized Q_{10} for the model normal fibers. (C) Impairment in the IHC as a function of impaired CF to account for the remaining threshold shift. (D) Impaired thresholds relative to the respective normal thresholds for the physiological data and model as a function of impaired CF. Gray symbols indicate measured data (Miller *et al.*, 1997) from individual fibers with different SRs. The dashed line represents the model predictions of threshold shifts for impairment in the OHC alone [as in (A)], and the solid line is for impairment of both OHCs and IHCs.

in the physiological data, and thus substantial damage to the IHCs is also necessary to fully model the impaired data (Miller *et al.*, 1997), consistent with histological assessment of OHC and IHC damage in cats with similar acoustic trauma (Liberman and Dodds, 1984). Note that the C2 responses do not contribute to the overall responses until around 80 dB SPL or higher and are trauma insensitive. Consequently, the C2 filter does not contribute to the threshold tuning of normal or impaired fibers except in the case of very severe impairment of C1 IHC transduction, i.e., for very small values of C_{IHC}.

III. RESULTS

Here we provide a detailed comparison of the model predictions with the results of the cat physiological experiments (Miller *et al.*, 1997; Wong *et al.*, 1998) in response to the vowel /ɛ/.

A. Predictions of single normal fiber data at low to moderate levels and impaired fiber data at high levels

Miller *et al.* (1997) recorded the responses of AN fibers to a synthesized stimulus "besh" and evaluated the synchronized responses in the steady-state vowel portion /ɛ/. The synchronized rates of AN fibers are determined by taking the Fourier transform of the poststimulus time histogram (PSTH) normalized to units of spikes/s (Sec. I.D. of Miller *et al.*, 1997).

Shown in Fig. 5 are the synchronized responses of two AN fibers (normal and impaired) with CFs near F2 for three different sound pressure levels; the left column shows the measured responses from cats (Miller *et al.*, 1997) and the right column shows the corresponding model predictions. Normal fiber responses shown in Fig. 5(a) predominantly synchronize to the formant F2, although there is some synchrony to the second harmonic of F2, and also to the adjacent harmonics around F2 at the lowest presentation level. In contrast, for the impaired fiber [Fig. 5(c)] the responses become broadband, and a substantial amount of power is shifted to F1 as the sound level is increased. Note that higher presentation levels are used for the impaired fibers because of the threshold shift with impairment.

The synchronized responses of a model fiber with normal OHC and IHC function are shown in Fig. 5(b). For comparison, the CF, threshold, and Q_{10} are matched to those of the example fiber from Fig. 5(a). The spontaneous rate of the model fiber after the effects of refractoriness is similar to the 33 spikes/s of the example normal fiber. The normal model fiber shows the following response properties. (i) At the lowest presentation level (31 dB SPL), the fiber shows phase locking to a number of harmonics of the vowel around F2, although the peak response is at F2. Synchrony to adjacent harmonics indicates that the fiber is operating in the linear region, i.e., compression/suppression is not in effect at this level. However, the model responses at adjacent harmonics around F2 are less prominent than in the physiological data. This might be due to the model fiber having a lower compression/suppression threshold, such that it starts to exhibit some synchrony capture by F2 at this level. (ii) There are substantial responses present even at higher frequencies near the second harmonic of the F2, which are due to the rectification of the instantaneous discharge rate in the synapse. Again, the model responses are weaker than the measured data. (iii) As the sound level increases, the phase locking to adjacent harmonics around F2 almost disappears, the responses being dominated by F2, which is referred to as synchrony capture. Compressive and suppressive nonlinearity of the BM are the main cause for synchrony capture by F2 at these levels (Bruce *et al.*, 2003). (iv) Synchronized rates do not increase linearly with the presentation levels, partly because of the compressive nonlinearity of the BM responses at these levels and partly because of saturation in the synapse.

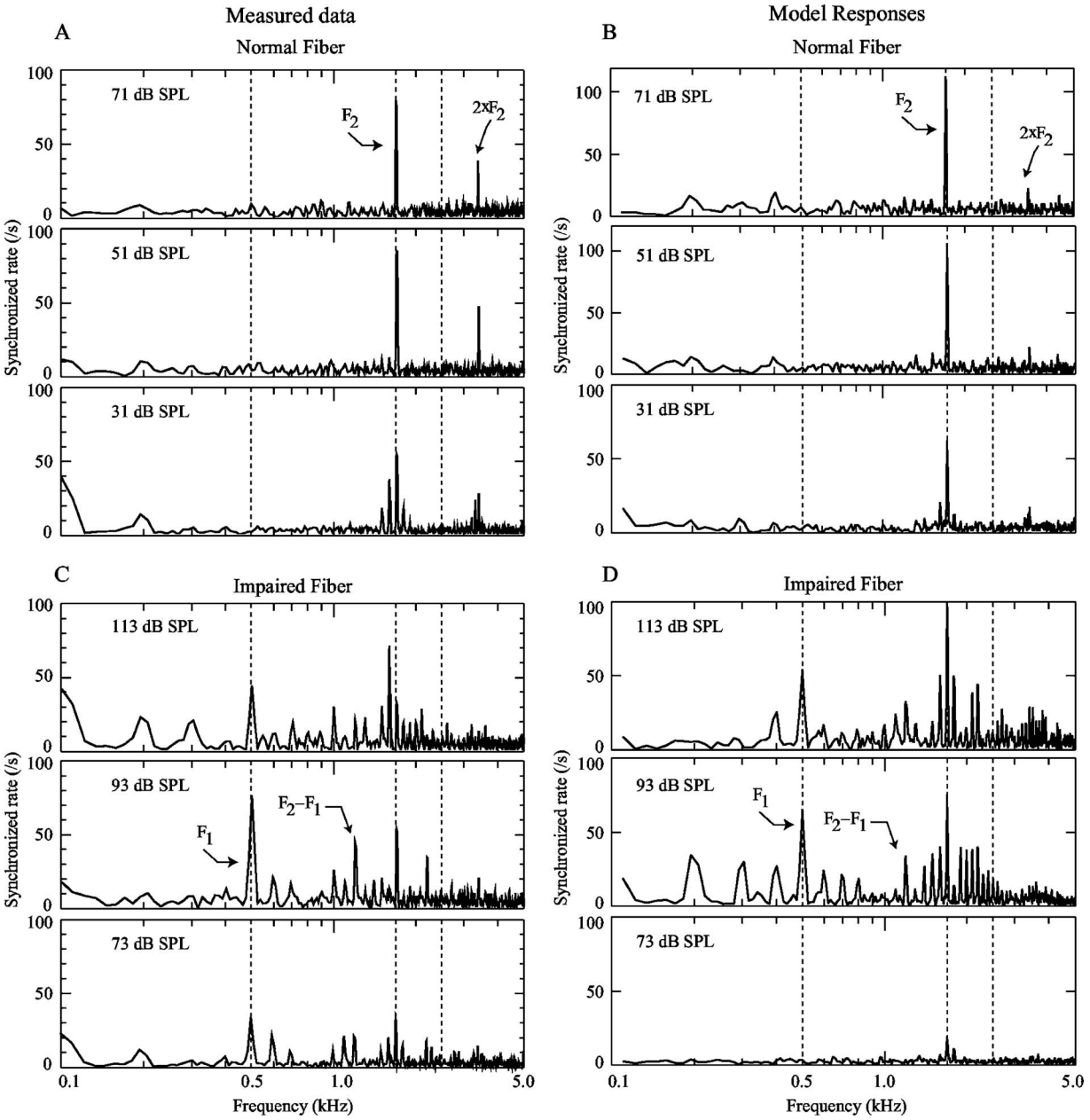


FIG. 5. Synchronized responses of normal [(A), (B)] and impaired [(C), (D)] fibers, all with CFs around F2, to the vowel /ɛ/ for three sound pressure levels as labeled. Left panels show the measured responses, reprinted from Fig. 9 of Miller *et al.* (1997) with permission from the Acoustical Society of America (© 1997), and model predictions are shown in the right panels. (A) Measured data from a normal fiber with CF of 1.7 kHz and Q_{10} of 4.4. (B) Model predictions from a normal model fiber with CF of 1.7 kHz and Q_{10} of 4.5. (C) Measured data from an impaired fiber with CF of 1.6 kHz, a Q_{10} of 3.8, and a threshold shift ~60 dB. (D) Model predictions from an impaired model fiber; no impairment in the OHC ($C_{OHC}=1.0$), but with IHC impairment ($C_{IHC}=0.035$) to a normal model fiber with CF of 1.6 kHz and a Q_{10} of 3.65, produces the tuning properties with an impaired CF at 1.6 kHz, $Q_{10}\approx 3.65$, and a threshold shift of ~55 dB, similar to those from the example fiber in (C).

Shown in Fig. 5(d) are the synchronized responses of a model impaired fiber for three different sound pressure levels. No impairment in the OHC ($C_{OHC}=1.0$), but with IHC impairment ($C_{IHC}=0.035$) to a normal model fiber with CF of 1.6 kHz and a Q_{10} of 3.65, produces the tuning properties with an impaired CF at 1.6 kHz, $Q_{10}\approx 3.65$, and a threshold shift of ~55 dB, similar to those from the example impaired fiber in Fig. 5(c). This example fiber has a slightly higher spontaneous discharge rate (44 spikes/s) than the model fiber. The model responses are broadband, meaning that significant synchrony is present at a large number of harmonics of the vowel. In particular, two prominent response components, one at F1 and the other at F2−F1, emerge at higher presentation levels, neither of which is present in the normal fiber's response, and significant responses are also present at other frequencies. However, the model responses show stronger synchrony to F2, whereas the physiological data for this impaired fiber shows moderate synchrony to F2 at all levels studied. It is also to be noted that at 93 dB SPL, the model prediction shows significant synchrony to a large number of harmonics other than formant related harmonics, which are not seen in the measured data. This smearing of the synchrony in the model prediction arises from the C1/C2 interaction, as the C1 and C2 responses are not completely out of

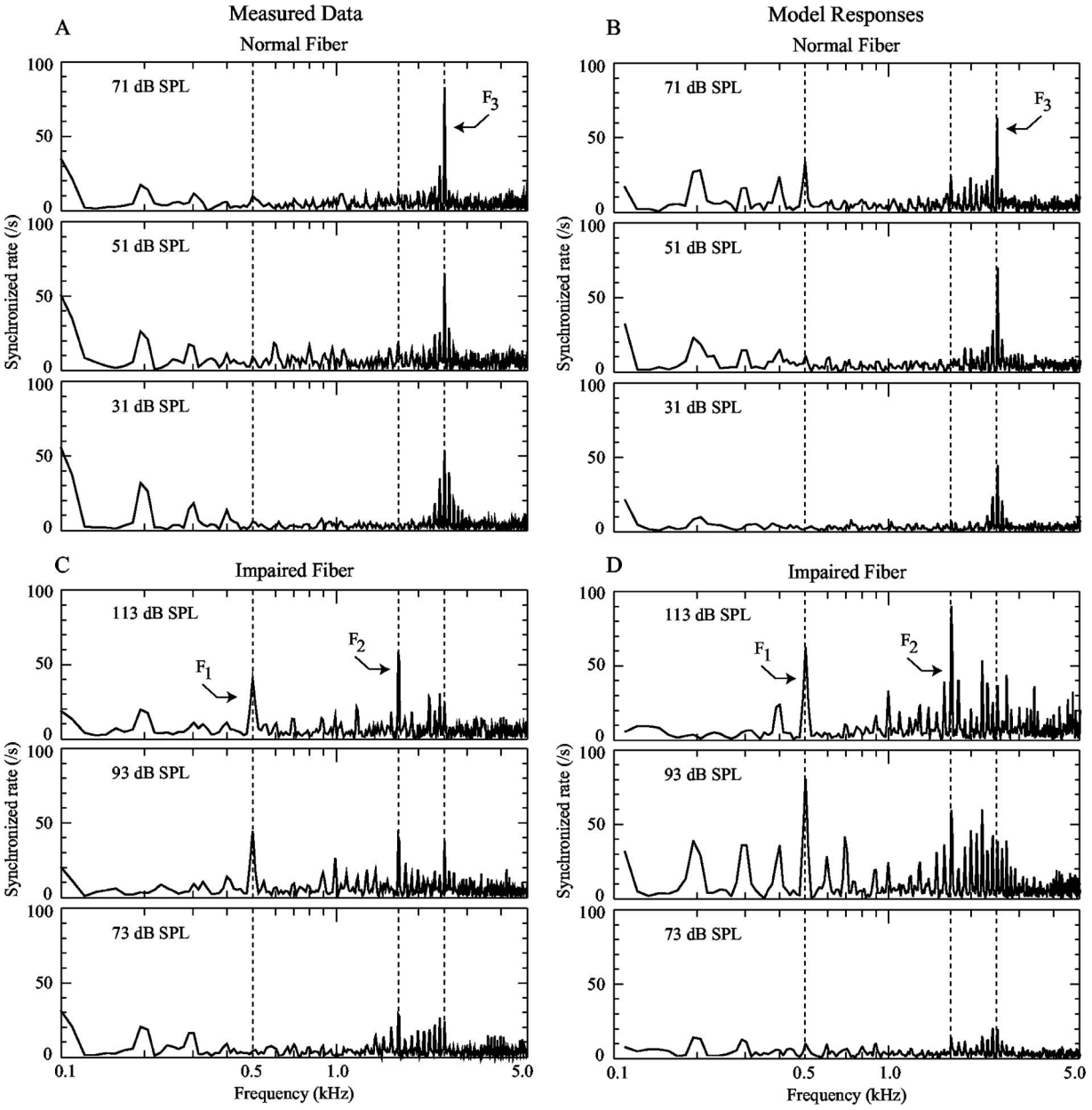


FIG. 6. Synchronized responses of normal [(A), (B)] and impaired [(C), (D)] fibers, all with CFs around F3, to the vowel /ɛ/ for three sound pressure levels as labeled. Left panels show the measured responses, reprinted from Fig. 10 of Miller *et al.* (1997) with permission from the Acoustical Society of America (© 1997), and model predictions are shown in the right panels. (A) Measured data from a normal fiber with CF of 2.5 kHz and Q_{10} of 4.3. (B) Model predictions from a normal model fiber with CF of 2.5 kHz and Q_{10} of 4.5. (C) Measured data from an impaired fiber with CF of 2.6 kHz, a Q_{10} of 1.4, and a threshold shift of ~60 dB. (D) Model predictions from an impaired model fiber; appropriate impairment in the OHC (C_{OHC}=0.18) and IHC (C_{IHC} =0.035) to a normal model fiber with CF of 2.95 kHz and Q_{10} of 4.87, produces the tuning properties with impaired CF at 2.6 kHz, $Q_{10}\approx 1.4$, and a threshold shift of ~60 dB, similar to those from the example fiber in (C).

phase at this presentation level (the C1/C2 transition occurs above 100 dB SPL for this fiber). Note that the outputs of the C1 and C2 filters become identical only when the control path output saturates at very high levels (Zilany and Bruce, 2006).

Figure 6 shows the synchronized responses for two fibers (normal and impaired) with CFs close to the third formant (F3); both measured (Miller *et al.*, 1997) and model responses are shown side-by-side for comparison. Here the normal fiber responses shown in Fig. 6(a) are synchronized to F3 and its adjacent harmonics. However, the synchrony capture is weaker than what was observed in the previous case [Fig. 5(a)]. The impaired fiber [Fig. 6(c)] again shows a broadband response with little synchrony to F3 and substantial responses at F1 and F2.

Shown in Fig. 6(b) are the synchronized responses for a model fiber with normal OHC and IHC function and CF, threshold and Q_{10} approximately matching the normal fiber from Fig. 6(a). Again, this example fiber has a slightly higher spontaneous discharge rate (48 spikes/s) than the model fiber. At all levels shown, the model fiber is phase locked to F3 and adjacent harmonics. As the presentation level increases, synchrony to F3 increases although some synchrony to adjacent harmonics still exists. At the highest presentation

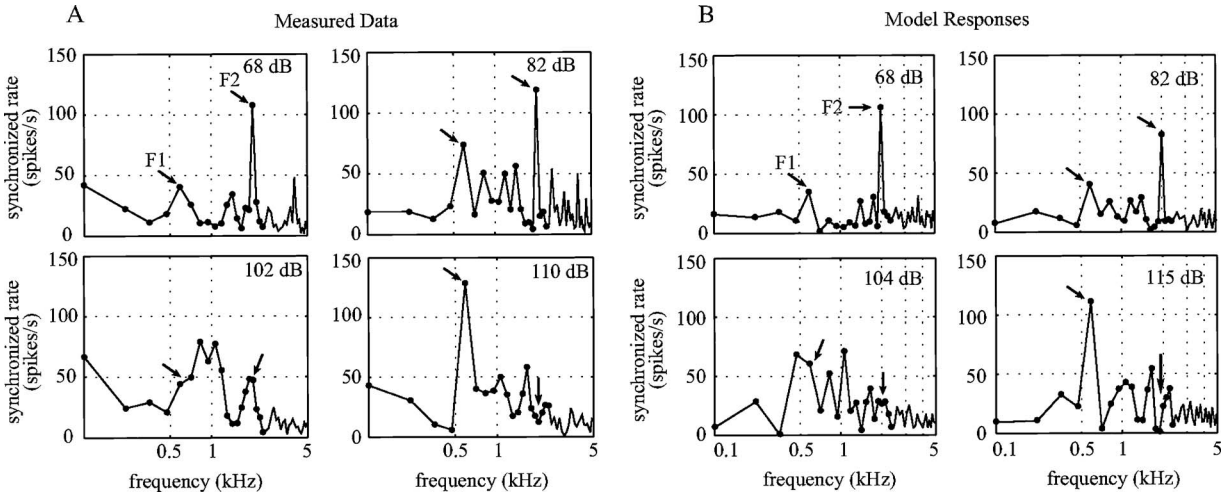


FIG. 7. Frequency spectra of responses to the vowel for a fiber with CF at 2.0 kHz for four sound pressure levels spanning from moderate to high levels as labeled. The four left panels show the measured synchronized rates, reprinted from Fig. 2 of Wong *et al.* (1998) with permission from Elsevier Science (© 1998), and the corresponding model predictions are shown in the four right panels. Synchronized rates to the first 20 harmonics are shown by closed circles in each case. Here, the original synthesized vowel (Fig. 1) is resampled such that the second formant is positioned exactly at the fiber's CF. Note that the levels at which the experimental and model fibers undergo the C1/C2 transition are 102 and 107 dB SPL, respectively.

level, a small amount of synchrony to F1 and F2 emerges that is not present in the measured data, which may result from the tuning of the model fiber at this stimulus level being broader than that of the particular cat fiber.

Appropriate impairment in the OHC (C_{OHC}=0.18) and IHC (C_{IHC}=0.035) to a normal model fiber with CF of 2.95 kHz and a Q_{10} of 4.87 produces the tuning properties with impaired CF of 2.6 kHz, $Q_{10}\approx 1.4$, and a threshold shift of ~60 dB, similar to those of the example impaired fiber from Fig. 6(c). This example fiber has a somewhat higher spontaneous discharge rate (66 spikes/s) than the model fiber. The responses of this model impaired fiber for three sound levels are shown in Fig. 6(d). Phase locking to F3 is mostly lost, while synchrony to other formants is seen at all presentation levels. At the two higher presentation levels, synchrony to F1 and F2 becomes more prominent. Like the example impaired fiber, some of the significant responses arise at frequencies that cannot be explained by the distortion products of the prominent formant frequencies.

The predictions of normal AN fiber responses to the vowel shown in this paper utilizing the model of Zilany and Bruce (2006) are quite similar to those of Bruce *et al.* (2003). However, the impaired fiber responses at high levels are better described by the newer model. Using Bruce *et al.* (2003), the synchrony to the formants in impaired fibers increases in almost all cases with presentation level [Figs. 9 and 12 of Bruce *et al.* (2003)], whereas the physiological data and also the model responses using Zilany and Bruce (2006) show somewhat different trends at higher levels. This improvement with the newer model comes from the interaction between the C1 and C2 responses at high levels.

B. Predictions of single normal fiber data at high levels

Wong *et al.* (1998) studied the temporal responses of normal fibers for the vowel /ɛ/ presented at high sound pressure levels. The vowel was resampled so that the second formant fell exactly at each fiber's CF. Figure 7 shows the spectra of responses to a vowel for a fiber with CF at 2.0 kHz. Both measured (Fig. 2 of Wong *et al.*, 1998) and model responses are shown for four different sound pressure levels. The synchronized rates to the first 20 harmonics of the vowel are shown by closed circles in each case.

At a vowel level of 68 dB SPL, both measured and model responses are synchronized to F2, consistent with the observation that the response of a fiber with a CF near a vowel's formant frequency is tightly phase-locked to that formant frequency. A weak response to F1 emerges at this level as the tuning of the fiber becomes broader. However, it is to be noted that the F1 response for the vowel is much weaker than that of the tonal (F1) response (Fig. 1 of Wong *et al.*, 1998), which is due to the suppression of F1 by F2 in the vowel response (Wong *et al.*, 1998).

As the level increases to 82 dB SPL, synchrony to F2 decreases slightly in the model responses, whereas responses to F1 and formant related harmonics (such as F2−2×F1, 2×F1, and F2−F1) increase. In fact, at this level the large components of the response are at F1, F2 and its distortion products, and hence are called multiformant response (Wong *et al.*, 1998). This indicates that the cochlea still retains some frequency selectivity. Another important observation is that the growth of response to F1 occurs at the expense of F2 before the C1/C2 transition, i.e., the growth of the response to F1 is not related to the C1/C2 interaction. So, it can be concluded that the multiformant response is not related in any way to the C1/C2 transition (Wong *et al.*, 1998). This is further supported by the model predictions to the vowel observed in Bruce *et al.* (2003). The model of Bruce and colleagues does not include a C2 filter and consequently is capable of generating only C1 responses. The vowel responses for the Bruce *et al.* (2003) model also show the fall of synchrony to F2 at high levels (although only for higher CF fibers) while the synchrony to F1 increases.

At 102 dB SPL in the measured data (or 104 dB SPL in

model responses), neither F1 nor F2 is dominant, and the responses become broadband, meaning that responses to the harmonics that are not related to the formants are substantial at this level. All components of the vowel for the model fiber undergo the C1/C2 transition at 107 dB SPL. Note that the C1 responses at this level are dominated by the formants and related harmonics, with the nonformant components being suppressed by the saturating nonlinearity of the C1 transduction function. In contrast, the C2 responses at this presentation level include relatively larger nonformant components than the C1 responses, because the C2 transduction function is not saturating at this level. Thus, the cancellation of the formant components in the antiphase C1 and C2 responses allows the nonformant components in the C2 response to dominate. At very high levels above the C2 threshold (115 dB SPL in model responses or 110 dB SPL in measured data), the synchrony to F2 is completely lost and goes to F1, although significant responses are seen in frequency components which are not related to the distortion products from F1 and F2, i.e., the fiber is responding in a broadband fashion.

It is to be noted that the two higher levels (104 and 115 dB SPL) used to determine model responses are slightly higher than the levels (102 and 110 dB SPL) of the corresponding physiological data. Since the model fiber in this case undergoes the C1/C2 transition at a slightly higher level (107 dB SPL) than the example fiber from the physiological data [102 dB SPL; cf. Figs. 12(c) and 12(f)], the two relatively higher levels have been used in determining model responses to illustrate the qualitative match between the model responses and the physiological data.

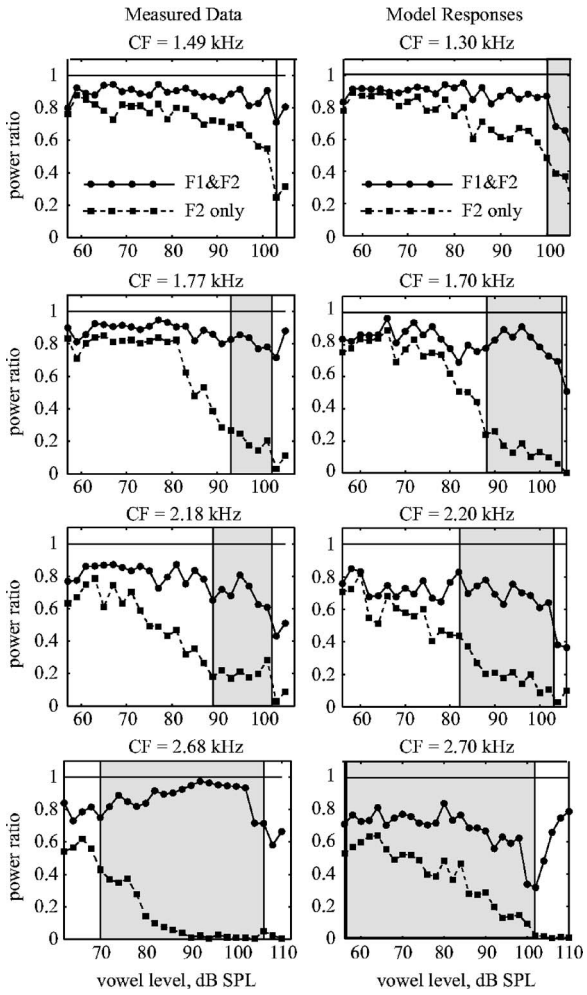


FIG. 8. Power ratio vs level functions for four normal fibers with CFs as labeled. The left column shows the measured PRs, reprinted from Fig. 4 of Wong *et al.* (1998) with permission from Elsevier Science (© 1998), and model predictions are displayed in the right column. Dashed lines with closed squares indicate the F2 PRs and solid lines with circles represent the F1&F2–related PRs, as shown in the legend. The lower and upper bounds of the shaded regions represent, respectively, the sound pressure levels at which synchrony capture by F2 is lost (VS$\leqslant$0.5) and the C2 threshold for F1.

C. Power ratios versus level

In order to quantify the degree of synchrony capture at different presentation levels, Wong *et al.* (1998) computed power ratios (PRs), which separate the responses into components related to the formants (including distortion products) and other components. Total power is defined as the sum of powers of the synchronized rates over the first 20 harmonics of the vowel response. The F2 PR is the fraction of total power of the synchronized rate that is phase locked to the second formant only. The F1&F2-related PR is the fraction of total power of the synchronized rate that is related to the F1 (5th harmonic), F2 (17th harmonic), and their distortion products (7th, 10th, 12th, 15th, and 20th harmonics).

Shown in Fig. 8 are the power ratios of four normal fibers with CFs in the range 1–3 kHz as a function of sound pressure level. The left and right columns show the measured and model responses, respectively. Normal model fiber responses are simulated with $C_{\mathrm{OHC}}=1$ and $C_{\mathrm{IHC}}=1$. The 400-ms synthetic vowel is sampled such that in each case the second formant falls exactly at the fiber's CF. The lower and upper bounds of the shaded region represent the levels at which synchrony capture by F2 is lost (i.e., vector strength,[2] VS$\leqslant$0.5) and the C2 threshold for F1, respectively. The C2 threshold is defined as the stimulus level at which the phase shift just exceeds 90°.

The measured data and model responses exhibit three distinct response areas. Before the shaded region (at low and moderate levels), responses are phase locked to F2 (VS $\geqslant$0.5), whereas in the shaded region, synchrony capture by F2 is lost and a multiformant response to F1 and F2 arises (see Fig. 7). After the shaded region (i.e., above the C2 threshold), the synchrony to F1 becomes dominant, and sometimes the unselective broadband response of the fiber is seen at these levels as judged by the drop in the F1&F2-related PR.

Both the measured data and model responses show that synchrony lost by F2 at high level is in no way related to the C1/C2 transition as the synchrony to F2 becomes progressively weaker before the C1/C2 transition (the upper bound of the shaded region). So, the C1 response itself shows gradual loss of synchrony by F2 at these levels while at the same time the response to F1 emerges for two reasons: (i) broadening of the C1 filter, and (ii) the shift in BF of the C1 filter. Thus, in the shaded region, the PR of F2 drops, whereas the F1&F2-related PR stays at the same level meaning that the F2 response is replaced by a response to F1.

However, the fall of F1&F2-related PR at or above C2 threshold seen in the model responses (also shown in Fig. 9 of Wong *et al.*, 1998) reflects the cancellation of the dominant formant components in the C1 and C2 responses, as the outputs of the C1 and C2 filters become comparable around this C1/C2 transition level. At very high levels, the dominant formant response is seen at F1 only (see Fig. 7), and thus F1&F2 related PR returns to a relatively high value.

With increasing CF, the loss of synchrony to F2 occurs at progressively lower stimulus levels and thus increases the width of the shaded region. It was argued in Wong *et al.* (1998) that the lower level switch of synchrony from F2 to F1 is due to the stronger two-tone suppression with increasing CF. This has been realized in the model as the CA gain is higher at higher CFs, and thus higher CF tuning becomes relatively broader (i.e., the relative change in Q_{10} is larger) than the tuning of a lower CF fiber at the same level. Consequently, the gain of the C1 filter is reduced, and then the low-pass filtering of the responses above 1 kHz in the IHC section further attenuates F2 more than F1 responses. Another potential reason for this early switch is that the frequency-scaled vowel has more power in F1 for higher CFs due to less attenuation by the ME filter. Bruce *et al.* (2003) found that this contributes to the switch of synchrony from F2 to F1 at an earlier level with increasing CF. Note that the levels at which the model predicts loss of synchrony capture by F2 are slightly lower than those of the experimental data for all CFs studied. This might be due to the fact that the model responses can only fit to the lower limits of the vector strength versus CF in the cat data (see Fig. 11 of Zhang *et al.*, 2001).

Figure 9 shows the PRs of four impaired fibers with substantial threshold shifts and broad tuning for sound pressure levels ranging from moderate to very high levels. The left column shows the data obtained experimentally (Wong *et al.*, 1998), and the right column is for model responses in which impaired Q_{10}, impaired CF, and threshold shifts for the fibers are matched with those of the physiological data. To do this, appropriate impairment in the OHC (C_{OHC}) and IHC (C_{IHC}) is applied to the model fibers with particular CF and Q_{10} values to obtain similar tuning curves to the experimental fibers. Like Fig. 8, the F2 PR and F1&F2-related PR as a function of level are shown by dotted and solid lines, respectively.

Both the measured data and the model predictions for impaired fibers show that the synchrony to F2 is mostly lost in all cases studied. The largest component of the response is at F1, which can be inferred from the observation that the F1&F2-related PR is typically more than twice the corresponding F2-related PR in each case. However, all of these responses are quite broadband as judged by the fall of F1&F2-related PR from the values near 1.0 observed in normal fibers (compare to Fig. 8).

It is to be noted that the model responses for the three fibers with higher CFs show slightly higher synchrony to F2 than those of measured data for sound levels up to 80 dB SPL, which is caused by the C1 responses. In all of these cases, the synchrony to F2 is near zero at ~60 dB SPL because of the loss of sensitivity (threshold shift) of these fibers, and then synchrony to F2 increases with level. Since these fibers are not completely impaired, the corresponding C1 responses still show some selectivity. However, above 80 dB SPL, the synchrony to F2 for all of these fibers again approaches zero, because at these levels the responses of the fibers are dominated by the broadband C2 response. In the model, the C2 filter has been implemented as the completely OHC-impaired version of the C1 filter, and thus has very

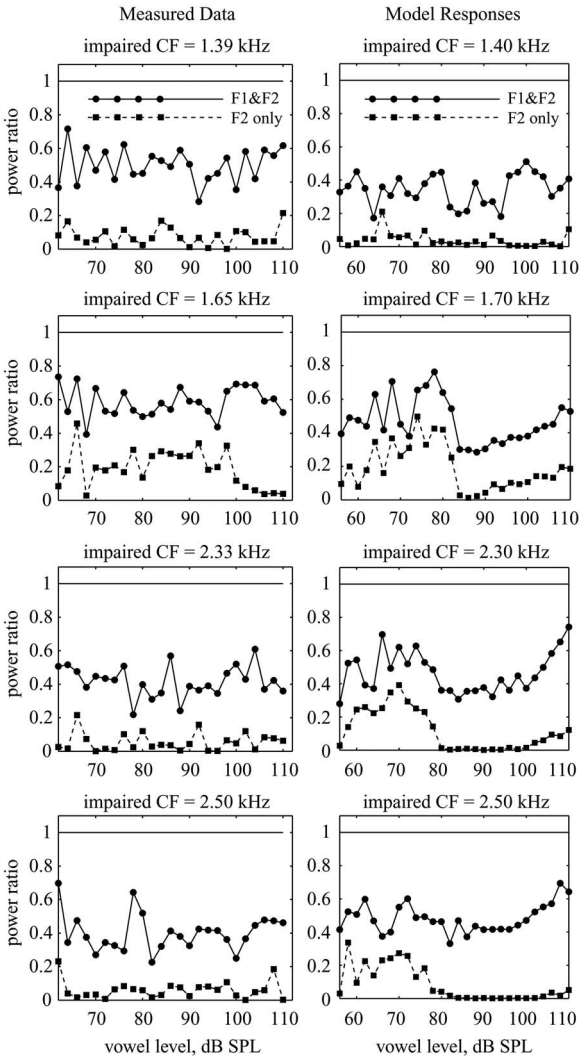


FIG. 9. Power ratio vs level functions for four impaired fibers with CFs as labeled. The left column shows the measured PRs, reprinted from Fig. 6 of Wong *et al.* (1998) with permission from Elsevier Science (© 1998), and model predictions are displayed in the right column. Dashed lines with closed squares indicate the F2 PRs and solid lines with circles represent the F1 and F2 PRs, as shown in the legend. Experimental impaired fibers are modeled by applying appropriate impairment (C_{OHC}, C_{IHC}) to the OHC and IHC of the normal model fibers: (i) A normal model fiber with CF of 1.8 kHz and a Q_{10} of 3.86, after being impaired with C_{OHC}=0.0 and C_{IHC} =0.0001, gives the tuning properties of the fiber with impaired CF at 1.4 kHz, impaired $Q_{10}\approx 0.8$, and a threshold shift of ~75 dB. (ii) A normal fiber with CF of 1.8 kHz and a Q_{10} of 6.0, after being impaired with C_{OHC}=0.4 and C_{IHC}=0.01, gives the tuning properties with impaired CF at 1.7 kHz, impaired $Q_{10}\approx 3.5$, and a threshold shift of ~55 dB. (iii) A normal fiber with CF of 2.5 kHz and a Q_{10} of 4.5, after being impaired with C_{OHC}=0.6 and C_{IHC}=0.008, gives the tuning with impaired CF at 2.3 kHz, impaired $Q_{10}\approx 2.8$, and a threshold shift of ~60 dB. (iv) A normal fiber with CF of 2.8 kHz and a Q_{10} of 4.75, after being impaired with C_{OHC} =0.4 and C_{IHC}=0.01, gives the tuning with impaired CF at 2.5 kHz, impaired $Q_{10}\approx 2.0$, and a threshold shift of ~60 dB.

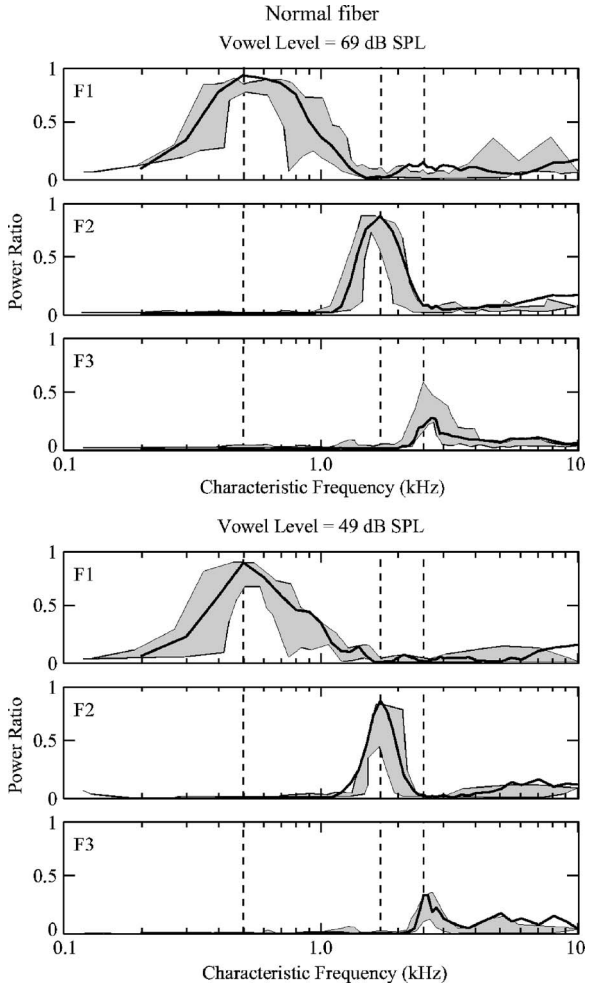


FIG. 10. Model predictions of F1, F2, and F3 power ratios for normal fibers as a function of CF at two sound pressure levels: 69 dB SPL (upper panels) and 49 dB SPL (lower panels). Thick solid lines indicate model predictions and gray shaded areas show the range of PR values observed in the physiological experiments (Miller *et al.*, 1997). Vertical dashed lines are aligned along the formant frequencies.

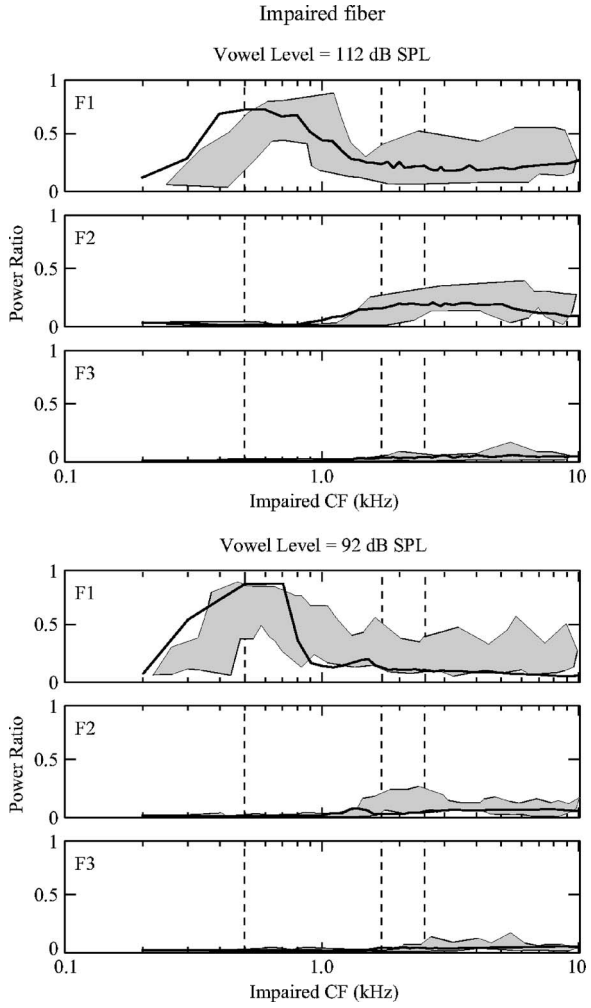


FIG. 11. Model predictions of F1, F2, and F3 power ratios for impaired fibers as a function of impaired CF at two sound pressure levels: 112 dB SPL (upper panels) and 92 dB SPL (lower panels). Impaired fibers are modeled by developing a functional relationship among C_{OHC}, C_{IHC}, and impaired CF—see Sec. II B 2. Thick solid lines show the model predictions and gray shaded areas indicate the range of PR values observed in the physiological experiments (Miller *et al.*, 1997). Vertical dashed lines indicate the formant frequencies.

broadband response properties. In addition, the C2 responses are not affected by the impairment of OHCs or IHCs. Thus, at very high levels, the synchrony behavior for the formants or harmonics is similar to that of normal fibers. So, the responses are dominated by the formant F1 with other substantial nonformant residual components (see the lowest panel of Fig. 7).

D. Synchrony capture as a function of CF

Miller *et al.* (1997) measured synchrony capture by the formants (F1, F2, and F3) of the vowel /ɛ/ in "besh" in terms of power ratios for a population of normal and impaired fibers across a wide range of CFs. Here the PR of a particular formant is defined as the fraction of total power (sum of powers of the first 50 harmonics) that is phase locked to the first, second, and third harmonics of that formant, as long as they are less than or equal to 5 kHz. Measured data and model predictions for populations of normal and impaired fibers are shown in Figs. 10 and 11, respectively. Thick solid lines indicate the model predictions and gray shaded areas show the range of PR values observed in the physiological experiments (Miller *et al.*, 1997).

Model predictions of PRs for normal fibers at 49 and 69 dB SPL shown in Fig. 10 fall predominantly within the range of values observed physiologically. In both the measured data and the model predictions, AN fibers are almost exclusively synchronized to the formant frequency nearest to the fiber's CF, and the peak of the synchrony to a particular formant becomes wider at higher levels which reflects a broader tuning at higher presentation levels.

To predict responses for a population of impaired fibers (Miller *et al.*, 1997), we utilized the functional relationships between C_{OHC} and CF and C_{IHC} and CF described in Sec. II B, such that the model population has impaired Q_{10} and threshold shift as a function of impaired CF similar to the experimental data. In order to compensate for the loss of sensitivity, higher presentation levels are used for the impaired fibers. Model predictions of PRs for impaired fibers

shown in Fig. 11 also fall reasonably within the range of values seen in the physiological data. Since the acoustic trauma caused impairment predominantly around the F2 and F3 regions, the synchrony to F2 and F3 falls drastically from those of the normal values. Higher presentation levels could not restore the normal response properties.

Both measured data and model predictions show that the synchrony to F3 stays near zero for both presentation levels, whereas fibers with CFs at or near F3 show some synchrony to F1 and F2, which is consistent with the results presented in Fig. 6. At 92 dB SPL, model F1 PRs exhibit lower values than the measured data for CFs higher than 800 Hz, because for these fibers the C1/C2 transition occurs at or around this presentation level. So, the response components contain noticeable amount of nonformant residual components, and thus the PR of F1 falls substantially. However, at 112 dB SPL, C2 responses dominate, and again F1 becomes the dominant component in the response. Consequently, the model fibers' synchrony to F1 stays within the range of physiological data in that CF region. At 112 dB SPL, the CF of peak synchrony to F1 shifts upward in the physiological data somewhat more than in the model predictions.

E. C1/C2 transitions in tonal and vowel responses

Wong *et al.* (1998) showed that the C1/C2 transition occurs even with complex stimuli. In fact, all components of the vowel undergo C1/C2 transition simultaneously, and the transition level is determined by the level at which the largest component reaches its C2 threshold. Note that the C2 threshold (or transition level) is defined as the level at which the phase is shifted by 90°. Figure 12 shows the comparison of the C2 thresholds for the first two formants in vowel responses and for tones at the frequencies of F1 and F2. For comparison, both measured responses from cats and model predictions are shown side by side; the left column shows the measured data and the right column shows the model predictions.

The model responses are shown for 21 fibers, and to be consistent with Wong *et al.* (1998), the playback sampling rate of the vowel /ɛ/ was changed in each case such that the second formant of the vowel falls exactly at the fiber's CF. The CFs of the model AN fibers range from 1 to 3 kHz (in steps of 100 Hz), which covers the usual range of second formant frequencies encountered in speech. It is worth noting that in determining C2 thresholds for the formants, the phase is only computed when the vector strength is greater than 0.15. It has been observed that when the synchrony is weak, the phase estimates are erratic and hence cannot be computed reliably. To make a comparison with the published data, the model results are analyzed based on the assumption that the levels of F1 and F2 are always 2 and 18 dB, respectively, below the overall vowel level.[3]

Figures 12(a) and 12(d) show the comparison of the C2 thresholds (in dB SPL) for the tone at F1 (along the abscissa) and for the formant F1 in vowel responses (along the ordinate). The model predictions, consistent with the measured data, fall within a band of ±5 dB SPL of the equality line. A similar comparison is shown for F2 in Figs. 12(b) and 12(e).

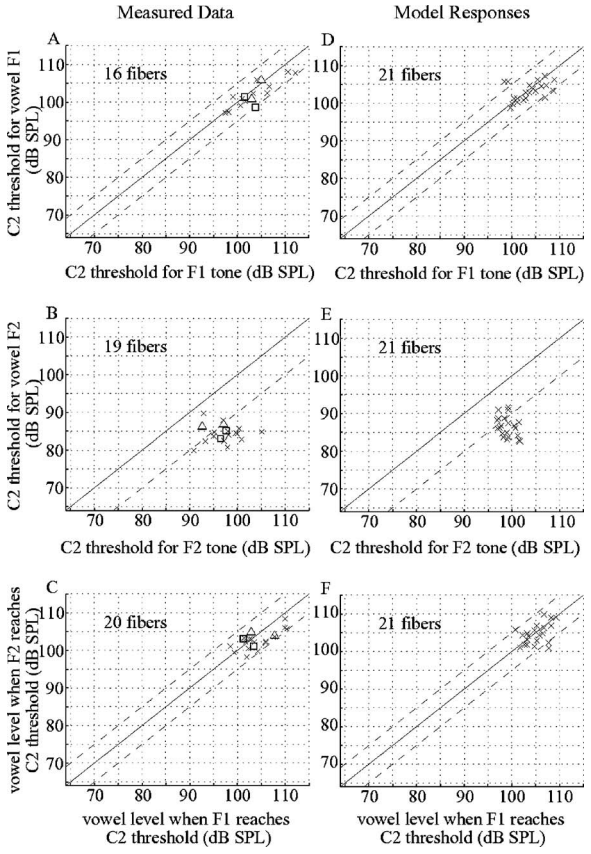


FIG. 12. Comparison of C2 thresholds for the two formants (in vowel responses) and for tones at the frequencies of F1 and F2. Left column [(A)–(C)] shows the measured responses, reprinted from Fig. 3 of Wong *et al.* (1998) with permission from Elsevier Science (© 1998), and model predictions are shown in the right column [(D)–(F)]. Model predictions are based on the responses from 21 fibers with CFs ranging from 1 to 3 kHz. In each part, the solid line indicates where the abscissa and ordinate values are equal. (A), (D) Comparison of the C2 thresholds (in dB SPL) for the formant F1 in vowel responses and for the frequency F1 in tonal responses. Dashed lines represent ±5 dB SPL deviations from the equality line. (B), (E) Comparison of the C2 thresholds for the formant F2 in vowel responses and the frequency F2 in tonal responses. Dashed line shows 10 dB SPL difference. (C), (F). Comparison of the overall vowel levels at which the two formants F1 and F2 undergo their corresponding C1/C2 transitions. Again, dashed lines indicate ±5 dB SPL deviations from the equality line.

Here the F2 level in the vowel is 5–20 dB lower than the level of the F2 tone to reach the C2 threshold. This early transition is attributed to the presence of other harmonics in the vowel. However, the vowel level when F2 reaches its C2 threshold is within ±5 dB of the vowel level when F1 reaches its C2 threshold, as shown in Figs. 12(c) and 12(f). Since the level of F1 in the vowel is almost the same as the level of the F1 tone to reach C2 threshold, and for F2 the formant reaches its C2 threshold at a level substantially lower than the level required for the corresponding tone, it can be concluded that it is the largest component in the vowel (in this case F1) that determines the level for C2 threshold for all components of the vowel. Both physiological data and model predictions rule out the possibility of separate processing of each component of the vowel, otherwise individual component might undergo C1/C2 transition separately. So, the overall vowel level determines the transition behavior for each component of the vowel.

This behavior can be understood qualitatively from the model architecture. The design of the model is such that the outputs of the C1 and C2 filter become equal when the control path output saturates, and the two separate transduction functions following the C1 and C2 filters make them completely out of phase. In response to a single tone, a very high sound pressure level is required to generate a large enough control signal to reduce the gain of the C1 filter to that of the C2 filter. In contrast, for a wideband stimulus such as a vowel, the control signal depends on the overall level of the stimulus, because the bandwidth of the control-path filter is wider than that of the signal-path C1 filter. Consequently, the level at which the outputs of the C1 and C2 filter become equal is determined by the overall level of a multicomponent stimulus. The overall level of a harmonic complex such as a vowel is determined primarily by the largest harmonic component. Following the IHC transduction functions, the C1 and C2 responses are out of phase for all components of the stimuli, and the steep growth of C2 transduction allows each component of the stimuli to undergo the C1/C2 transition at the same high presentation level, which is determined by the largest harmonic component, as observed earlier.

IV. DISCUSSION

The normal and impaired AN fiber responses in cats to the vowel /ɛ/ are described in this paper using the computational model proposed by Zilany and Bruce (2006). The vowel presented to the tympanic membrane during physiological experiments is replicated here, and then applied as the input to the model. To do this, the stimulus is first filtered by the HRTF of the human head and then modified by the frequency response of the acoustic delivery system. This is necessary, because the change in the relative levels of the stimulus components (especially the formants) affects the AN responses substantially, and part of the improvements in predictions over the Bruce *et al.* (2003) results from this. The model predictions are compared to the responses from the physiological experiments (Miller *et al.*, 1997; Wong *et al.*, 1998), which clearly demonstrates the ability of the model to closely represent the measured data in both normal and impaired ears. The measured data and model predictions include responses of individual fibers, as well as a population of fibers over a wide range of CFs.

A. C1/C2 transitions for the vowel

The observation that all components of the vowel undergo the C1/C2 transition simultaneously (Wong *et al.*, 1998) has been accurately modeled in this paper, and the transition level is determined by the level of the largest component of the vowel. To provide further evidence, we have determined the transition levels in the model AN responses to a contrast enhanced frequency-shaped (CEFS) version of the vowel /ɛ/. The CEFS vowel has a spectrum (i.e., the relative magnitude of the formants and harmonics) that is substantially different from that of the standard /ɛ/, such that it is dominated by the second formant (Miller *et al.*, 1999a). The line spectrum of the CEFS vowel is shown in the upper panel of Fig. 13. The playback sampling rate of this vowel is changed so that F2 always falls at the fiber's CF. Again, model AN fibers with CFs ranging from 1 to 3 kHz with steps of 100 Hz have been considered. The lower panel of Fig. 13 shows the comparison of the C2 thresholds (in dB SPL) in model AN fiber responses for a tone at the F2 frequency (along the abscissa) and for the formant F2 in the CEFS vowel (along the ordinate). The result falls within a band of ±5 dB SPL from the equality line, indicating that it is the largest component F2 in the CEFS vowel that determines the C1/C2 transition for all components of the vowel [cf. Fig. 12(e)]. It is to be noted that the synchrony to F1 for this vowel is not enough to produce a steady phase response before the C1/C2 transition, and that is why the C1/C2 transition for F1 is not shown here. However, in the C2 responses (i.e., at very high levels), the synchrony to F1 is substantial and thus gives a steady phase response.

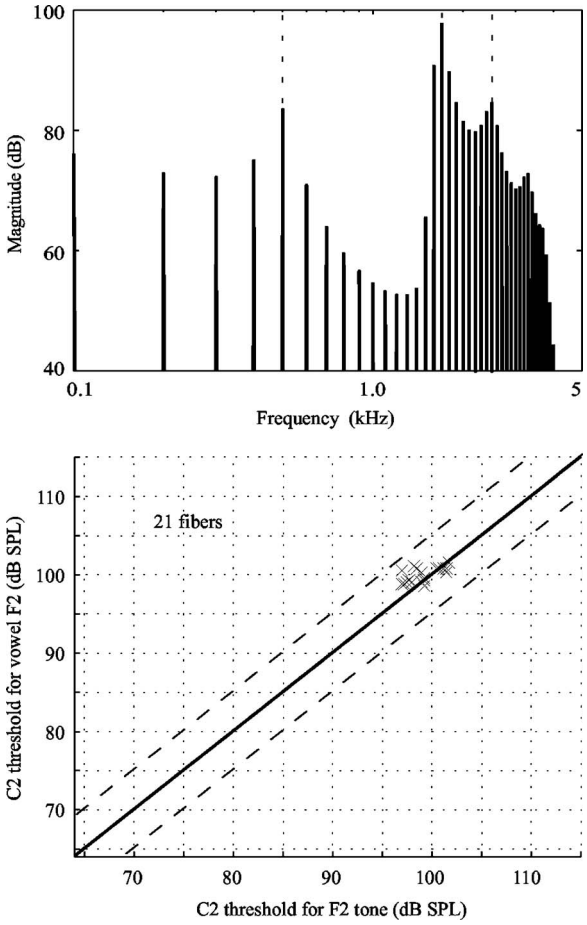


FIG. 13. Upper panel: Line spectrum of the CEFS vowel (Miller *et al.*, 1999a). Lower panel: Comparison of the C2 thresholds for the formant F2 in vowel responses and the frequency F2 in tonal responses. Dashed lines indicate ±5 dB SPL deviations from the equality line.

B. Comparison with previous models

A number of earlier computational models attempted to predict the AN fiber responses to speech-like sounds (e.g., Deng and Geisler, 1987a; Geisler, 1989; Jenison *et al.*, 1991; Sachs *et al.*, 2002). However, most of them lacked several important aspects of nonlinearity observed in the cochlea that are essential in predicting responses to speech. For ex-

ample, Geisler (1989) studied the effects of broadened tuning on responses to speech using a computational model that did not include two-tone suppression and BM compression. Recently, Bruce *et al.* (2003) and Holmes *et al.* (2004) attempted a more systematic investigation of AN responses to vowels in cats and guinea pigs, respectively. Both studies employed more complete AN models, in that they are able to produce a wide range of physiological phenomena in response to simple and complex stimuli.

The predictions of AN fiber responses using the model by Zilany and Bruce (2006) are almost the same as those of Bruce *et al.* (2003) when operating at low to moderate sound pressure levels. However, the input to the model is modified here by the frequency response of the sound delivery system to closely match the vowel's spectral shape at the tympanic membrane, as was presented during experiments. The change in relative levels of the formants explains the onset of loss of synchrony capture by F2 at moderate levels for CFs above 2 kHz (see Fig. 8), which was not observed by Bruce *et al.* (2003). At high levels, the predictions presented in this paper are superior to those of Bruce *et al.* (2003) for several reasons. First, the C1 filter, originally a gammatone filter in Bruce *et al.* (2003), has been replaced by a chirp filter in Zilany and Bruce (2006), and thus the impulse response of the model AN fiber shows the instantaneous frequency glide in it, which is responsible for producing the shift in BF at high levels in normal and impaired fibers. The loss of synchrony capture in impaired fibers, as well as in normal fibers at high levels, is partly attributed to this. Second, the inclusion of the C2 filter in Zilany and Bruce (2006) enables the model to address the AN responses seen at the C1/C2 transition and above. For instance, the significant loss of synchrony capture by the dominant formant components observed around the C1/C2 transition level cannot be described without the C1/C2 interaction. Third, the CA gains for fibers with lower CFs in Bruce *et al.* (2003) are not sufficient to account for the significant loss of synchrony capture by F2 seen at very high levels in normal and impaired fibers.

Holmes *et al.* (2004) predicted guinea pig data for single and double vowels (Palmer *et al.*, 1986; Palmer, 1990) using a cochlear model applying the dual resonance nonlinear filter bank approach (Meddis *et al.*, 2001; Sumner *et al.*, 2003). They showed that the BM nonlinearity has a very small effect in the representation of speech at low frequencies, and it is the filter shape not the nonlinearity that is responsible for the vowel formant representation in their model predictions. This observation strongly contrasts with the model responses in cats (Deng and Geisler, 1987a; Sachs *et al.*, 2002; Bruce *et al.*, 2003), where BM nonlinearity plays an important role in the representation of speech over a wide range of CFs, and thus synchrony capture is relatively stronger. Holmes *et al.* (2004) suggest that the differences in the guinea pig and cat model predictions may arise from a lower degree of BM nonlinearity in guinea pigs at lower CFs. Indeed, compressive nonlinearity appears to be absent from BM vibration recordings made in the apex of the guinea pig cochlea (Cooper and Rhode, 1995; Khanna and Hao, 1999). Unfortunately, no similar BM data are available in cat, so we must infer the degree of BM nonlinearity from the cat AN data. In this work, we have further increased nonlinearity (CA gain) at lower CFs than was employed in Bruce *et al.* (2003)—see Fig. 3. Human psychoacoustic data show substantial nonlinearity in low frequency hearing (Oxenham and Plack, 1997; Plack and Oxenham, 2000; Lopez-Poveda *et al.*, 2003). If our estimates of cat BM nonlinearity are at least qualitatively correct, then this may suggest that the cat is a more suitable animal model than the guinea pig for predicting neural responses to speech in humans.

The main focus of this paper is to predict AN responses to vowels at high levels (Miller *et al.*, 1997; Wong *et al.*, 1998), especially the spread of synchrony and the C1/C2 transition. Lin and Goldstein (1995) incorporated the two-factor hypothesis in their model to explain high-level nonlinearities such as the C1/C2 transition and peak splitting, and conducted a detail analysis of responses to tones. As their model incorporates only static nonlinearities, Lin and Goldstein (1995) found it necessary to introduce an *ad hoc* stimulus-dependent phase adaptation mechanism to produce the correct stimulus-dependent phase changes as a function of time. In addition, the C1/C2 transition occurs in the model BM responses, inconsistent with most of the physiological observations (Ruggero *et al.*, 1997; Cheatham and Dallos, 1998; Robles and Ruggero, 2001). It is unlikely that this model can predict vowel data at high levels, especially the C1/C2 transition. Holmes *et al.* (2004) predicted guinea pig data to vowels at low to moderate levels, and in fact, only cat data are available at high levels. However, the model used in Holmes *et al.* (2004) is of the type with parallel filter paths with static nonlinearities like that of Lin and Goldstein (1995), and hence it is not clear if this form of model could predict the vowel C1/C2 transition behavior observed in cats.

C. Implications for speech intelligibility and hearing aids

Although the predictions of normal AN fiber responses at low to moderate levels shown in this paper utilizing the model of Zilany and Bruce (2006) are quite similar to those of Bruce *et al.* (2003), the normal and impaired fiber responses at high levels are better described by the newer model. It is of interest that the method of implementing impairment in this paper was the same as that of Bruce *et al.* (2003). Consequently, the improved predictions of the impaired AN fiber data arise solely as a result of the improved description of the normal AN fiber behavior that was obtained with the new C1 filter properties and the addition of the C2 filter.

In predicting impaired fiber responses to the vowel using Bruce *et al.* (2003), the synchrony capture by F2 remains substantial even at very high levels [Figs. 14 and 16 of Bruce *et al.* (2003)]. In contrast, measured data and predictions using the new model (Figs. 9 and 11 of this paper) show significant loss of synchrony to F2 at high levels. The new model also now better predicts the loss of synchrony to F2 in normal fibers at high levels—compare Fig. 13 of Bruce *et al.* (2003) with Fig. 8 of this paper. However, in both the measured data and the new model predictions, the degradation of the representation of F2 at high presentation levels is more severe in the impaired fibers. Wong *et al.* (1998) sug-

gested that these results could at least partly explain the roll-off observed in speech intelligibility at high levels that is stronger in hearing impaired listeners than normal hearing listeners, even when both are operating at the same presentation level (Stelmachowicz *et al.*, 1985; Glasberg and Moore, 1986). We have tested this hypothesis with a speech-intelligibility predictor utilizing the Zilany and Bruce (2006) model and have found that the model can indeed explain the effects of presentation level and hearing impairment on intelligibility (Zilany and Bruce, 2007). Consequently, this model appears to be a useful tool for the design and evaluation of hearing aid amplification schemes (e.g., Miller *et al.*, 1999a; Bruce, 2004; Bondy *et al.*, 2004).

D. Future directions

The results of this study show good qualitative and quantitative match of the model predictions with the vowel data in cats. The model responses to other simple and complex stimuli are described in Zilany and Bruce (2006). Now the model appears accurate enough to be useful as a front end in many research areas, such as speech recognition in noisy conditions (e.g., Ghitza, 1988; Tchorz and Kollmeier, 1999), modeling of auditory scene analysis (e.g., Brown and Cooke, 1994), modeling of neural circuits in the auditory brain-stem (e.g., Hewitt and Meddis, 1992), design of speech processors in cochlear implants (e.g., Wilson *et al.*, 2005), and design of hearing aid amplification schemes (e.g., Miller *et al.*, 1999a; Bruce, 2004; Bondy *et al.*, 2004).

Although the model is primarily designed to replicate AN fiber data, it does give some insight about the possible physical mechanisms producing the C1/C2 interaction in the cochlea. The model architecture has been designed such that two modes of BM vibration, active and passive, are transduced by two separate IHC transduction functions. This separate transduction mechanism seems consistent with two important observations: (i) C1 responses are highly affected by cochlea impairment, while C2 is resistant to trauma, and (ii) all components of a multicomponent stimulus undergo C1/C2 transition simultaneously irrespective of their individual levels. However, further experimental and modeling studies of cochlear micromechanics and hair cell transduction are clearly required in order to better determine the details of the physical basis for the C2 response.

In addition, our model provides an explanation of why all the formants of a vowel undergo the C1/C2 transition together. Consistent with the hypothesis of Wong *et al.* (1998), we found that the model's C1/C2 transition for the vowel formants is dependent on the overall level of the vowel, which is dominated by the formant with the greatest amplitude. We tested this hypothesis further using the CEFS vowel, a variant on the standard /ɛ/ vowel. However, in physiological experiments the CEFS vowel has only been presented to cats with impaired ears (Miller *et al.*, 1999a). Further experiments in normal-hearing cats with a range of different vowels or multitone complexes would be beneficial in further understanding the C1/C2 transition behavior for such stimuli.

ACKNOWLEDGMENTS

The authors would like to thank Eric Young and Roger Miller for supplying acoustic calibration and auditory nerve data from the physiological studies and Laurel Carney, Michael Heinz, Eric Young, and three anonymous reviewers for giving invaluable feedback on earlier versions of the manuscript. This research was supported by NSERC Discovery Grant No. 261736 and the Barber-Gennum Chair Endowment.

[1]In this paper, best frequency (BF) is the frequency at which the fiber response is maximum, which can vary with sound level, whereas the characteristic frequency (CF) is the stimulus frequency at which the fiber has the lowest threshold. Note that the model parameter CF is invariant to the stimulus sound pressure level and hair cell impairment. However, the value for CF that is measured from a model tuning curve may change as a function of hair cell impairment (see Fig. 5 of Zilany and Bruce, 2006), although the model parameter CF has not changed. In such cases, we refer to the CF measured from the tuning curve as the *impaired CF*. Note that in this paper the CFs (normal or impaired) are estimated from the corresponding tuning curves following the method described in Liberman (1984).

[2]Vector strength is a dimensionless measure of phase locking, and is defined, for a particular frequency, as the ratio of the synchronized rate at that frequency and the average discharge rate of the fiber; $\mathrm{VS}(kf_0) = |R(kf_0)|/R(0)$, where $|R(kf_0)|$ is the magnitude of the Fourier transform of the period histogram at frequency kf_0, the kth harmonic of the stimulus fundamental f_0 (Wong *et al.*, 1998).

[3]Wong *et al.* (1998) erroneously stated that the levels of F1 and F2 in the vowel driving the headphone were always 2 and 18 dB, respectively, below the overall vowel level. However, this actually corresponds to the average levels of the first two formants of the vowel *after* being modified by the acoustic delivery system. In fact, the acoustic delivery system changes the levels of the various components of the vowel to an extent of ±6 dB for frequencies below 3 kHz (Wong *et al.*, 1998). So, the relative levels of the various vowel components, including F1 and F2, change as the vowel is resampled. Thus, the analysis should be based on the exact levels of F1 and F2 in the vowel at the tympanic membrane. However, when the model responses were analyzed with the exact levels of F1 and F2 in the vowel at the tympanic membrane (input to the model), the results were qualitatively the same as when using the average levels of F1 and F2 in the vowel used in Wong *et al.* (1998). Consequently, to avoid reanalyzing the data, we have used the same analysis (average rather than exact levels) as Wong *et al.* (1998).

Bondy, J., Becker, S., Bruce, I., Trainor, L., and Haykin, S. (**2004**). "A novel signal-processing strategy for hearing-aid design: Neurocompensation," Signal Process. **84**, 1239–1253.

Brown, G. J., and Cooke, M. (**1994**). "Computational auditory scene analysis," Comput. Speech Lang. **8**, 297–336.

Bruce, I. C. (**2004**). "Physiological assessment of contrast-enhancing frequency shaping and multiband compression in hearing aids," Physiol. Meas. **25**, 945–956.

Bruce, I. C., Sachs, M. B., and Young, E. D. (**2003**). "An auditory-periphery model of the effects of acoustic trauma on auditory nerve responses," J. Acoust. Soc. Am. **113**, 369–388.

Carney, L. H., McDuffy, M. J., and Shekhter, I. (**1999**). "Frequency glides in the impulse responses of auditory-nerve fibers," J. Acoust. Soc. Am. **105**, 2384–2391.

Cheatham, M. A., and Dallos, P. (**1998**). "The level dependence of response phase: Observations from cochlear hair cells," J. Acoust. Soc. Am. **104**, 356–369.

Cooper, N. P., and Rhode, W. S. (**1995**). "Nonlinear mechanics at the apex of the guinea-pig cochlea," Hear. Res. **82**, 225–243.

Cooper, N. P., and Rhode, W. S. (**1997**). "Mechanical responses to two-tone distortion products in the apical and basal turns of the mammalian cochlea," J. Neurophysiol. **78**, 261–270.

Delgutte, B. (**1980**). "Representation of speech-like sounds in the discharge patterns of auditory-nerve fibers," J. Acoust. Soc. Am. **68**, 843–857.

Delgutte, B., and Kiang, N. Y. S. (**1984**). "Speech coding in the auditory nerve. V. Vowels in background noise," J. Acoust. Soc. Am. **75**, 908–918.

Deng, L., and Geisler, C. D. (**1987a**). "A composite auditory model for

processing speech sounds," J. Acoust. Soc. Am. **82**, 2001–2012.
Deng, L., and Geisler, C. D. (**1987b**). "Responses of auditory-nerve fibers to nasal consonant-vowel syllables," J. Acoust. Soc. Am. **82**, 1977–1988.
Geisler, C. D. (**1989**). "The responses of models of 'high-spontaneous' auditory-nerve fibers in a damaged cochlea to speech syllables in noise," J. Acoust. Soc. Am. **86**, 2192–2205.
Ghitza, O. (**1988**). "Temporal non-place information in the auditory-nerve firing patterns as a front-end for speech recognition in a noisy environment," J. Phonetics **16**, 109–123.
Glasberg, B. R., and Moore, B. C. J. (**1986**). "Auditory filter shapes in subjects with unilateral and bilateral cochlear impairments," J. Acoust. Soc. Am. **79**, 1020–1033.
Goldstein, J. L. (**1990**). "Modeling rapid waveform compression on the basilar membrane as multiple-bandpass-nonlinearity filtering," Hear. Res. **49**, 39–60.
Goldstein, J. L. (**1995**). "Relations among compression, suppression, and combination tones in mechanical responses of the basilar membrane: Data and MBPNL model," Hear. Res. **89**, 52–68.
Hewitt, M. J., and Meddis, R. (**1992**). "Regularity of cochlear neucleus stellate cells: A computational modeling study," J. Acoust. Soc. Am. **93**, 3390–3399.
Holmes, S. D., Summer, C. J., O'Mard, L. P., and Meddis, R. (**2004**). "The temporal representation of speech in a nonlinear model of the guinea pig cochlea," J. Acoust. Soc. Am. **116**, 3534–3545.
Jenison, R. L., Greenberg, S., Kluender, K. R., and Rhode, W. S. (**1991**). "A composite model of the auditory periphery for the processing of speech based on the filter response functions of single auditory-nerve fibers," J. Acoust. Soc. Am. **90**, 773–786.
Khanna, S. M., and Hao, L. F. (**1999**). "Reticular lamina vibrations in the apical turn of a living guinea pig cochlea," Hear. Res. **132**, 15–33.
Kiang, N. Y.-S. (**1990**). "Curious oddments of auditory-nerve studies," Hear. Res. **49**, 1–16.
Liberman, M. C. (**1984**). "Single-neuron labeling and chronic cochlear pathology. I. Threshold shift and characteristic-frequency shift," Hear. Res. **16**, 33–41.
Liberman, M. C., and Dodds, L. W. (**1984**). "Single-neuron labeling and chronic cochlear pathology. III. Stereocilia damage and alterations of threshold tuning curves," Hear. Res. **16**, 55–74.
Liberman, M. C., and Kiang, N. Y.-S. (**1984**). "Single-neuron labeling and chronic cochlear pathology. IV. Stereocilia damage and alterations in rate- and phase-level functions," Hear. Res. **16**, 75–90.
Lin, T., and Goldstein, J. L. (**1995**). "Quantifying 2-factor phase relations in non-linear responses from low characteristic-frequency auditory-nerve fibers," Hear. Res. **90**, 126–138.
Lopez-Poveda, E. A., Plack, C. J., and Meddis, R. (**2003**). "Cochlear nonlinearity between 500 and 8000 Hz in listeners with normal hearing," J. Acoust. Soc. Am. **113**, 951–960.
Meddis, R., O'Mard, L. P., and Lopez-Poveda, E. A. (**2001**). "A computational algorithm for computing nonlinear auditory frequency selectivity," J. Acoust. Soc. Am. **109**, 2852–2861.
Miller, R. L., Calhoun, B. M., and Young, E. D. (**1999a**). "Contrast enhancement improves the representation of /ɛ/-like vowels in the hearing-impaired auditory nerve," J. Acoust. Soc. Am. **106**, 2693–2708.
Miller, R. L., Calhoun, B. M., and Young, E. D. (**1999b**). "Discriminability of vowel representations in cat auditory-nerve fibers after acoustic trauma," J. Acoust. Soc. Am. **105**, 311–325.
Miller, R. L., Schilling, J. R., Franck, K. R., and Young, E. D. (**1997**). "Effects of acoustic trauma on the representation of the vowel /ɛ/ in cat auditory nerve fibers," J. Acoust. Soc. Am. **101**, 3602–3616.
Nuttall, A. L., and Dolan, D. F. (**1996**). "Steady-state sinusoidal velocity responses of the basilar membrane in guinea pig," J. Acoust. Soc. Am. **99**, 1556–1564.
Oxenham, A. J., and Plack, C. J. (**1997**). "A behavioral measure of basilar-membrane nonlinearity in listeners with normal and impaired hearing," J. Acoust. Soc. Am. **101**, 3666–3675.
Palmer, A. R. (**1990**). "The representation of the spectra and fundamental frequencies of steady-state single- and double-vowel sounds in the temporal discharge patterns of guinea pig cochlear-nerve fibers," J. Acoust. Soc. Am. **88**, 1412–1426.
Palmer, A. R., Winter, I. M., and Darwin, C. J. (**1986**). "The representation of the steady-state vowel sounds in the temporal discharge patterns of the guinea pig cochlear nerve and primarylike cochlear nucleus neurons," J. Acoust. Soc. Am. **79**, 100–113.
Plack, C. J., and Oxenham, A. J. (**2000**). "Basilar-membrane nonlinearity estimated by pulsation threshold," J. Acoust. Soc. Am. **107**, 501–507.
Robles, L., and Ruggero, M. A. (**2001**). "Mechanics of the mammalian cochlea," Physiol. Rev. **81**, 1305–1352.
Ruggero, M. A., Rich, N. C., Recio, A., Narayan, S. S., and Robles, L. (**1997**). "Basilar-membrane responses to tones at the base of the chinchilla cochlea," J. Acoust. Soc. Am. **101**, 2151–2163.
Sachs, M. B., Bruce, I. C., Miller, R. L., and Young, E. D. (**2002**). "Biological basis of hearing-aid design," Ann. Biomed. Eng. **30**, 157–168.
Schilling, J. R., Miller, R. L., Sachs, M. B., and Young, E. D. (**1998**). "Frequency-shaped amplification changes the neural representation of speech with noise-induced hearing loss," Hear. Res. **117**, 57–70.
Sinex, D., and Geisler, C. (**1983**). "Responses of auditory-nerve fibers to consonant-vowel syllables," J. Acoust. Soc. Am. **73**, 602–615.
Stelmachowicz, P. G., Jesteadt, W., Gorga, M. P., and Mott, J. (**1985**). "Speech perception ability and psychophysical tuning curves in hearing-impaired listeners," J. Acoust. Soc. Am. **77**, 620–627.
Sumner, C. J., O'Mard, L. P., Lopez-Poveda, E. A., and Meddis, R. (**2003**). "A non-linear filter-bank model of the guinea-pig cochlear nerve: Rate responses," J. Acoust. Soc. Am. **113**, 3264–3274.
Tan, Q., and Carney, L. H. (**2003**). "A phenomenological model for the responses of the auditory-nerve fibers. II. Nonlinear tuning with a frequency glide," J. Acoust. Soc. Am. **114**, 2007–2020.
Tchorz, J., and Kollmeier, B. (**1999**). "A model of auditory perception as a front end for automatic speech recognition," J. Acoust. Soc. Am. **106**, 2040–2050.
Wiener, F. M., and Ross, D. A. (**1946**). "The pressure distribution in the auditory canal in a progressive sound field," J. Acoust. Soc. Am. **18**, 401–408.
Wilson, B. S., Schatzer, R., Lopez-Poveda, E. A., Sun, X., Lawson, D. T., and Wolford, R. D. (**2005**). "Two new directions in speech processor design for cochlear implants," Ear Hear. **26**, 73S–81S.
Wong, J. C., Miller, R. L., Calhoun, B. M., Sachs, M. B., and Young, E. D. (**1998**). "Effects of high sound levels on responses to the vowel /ɛ/ in cat auditory nerve," Hear. Res. **123**, 61–77.
Young, E. D., and Sachs, M. B. (**1979**). "Representation of steady-state vowels in the temporal aspects of the discharge patterns of populations of auditory nerve fibers," J. Acoust. Soc. Am. **66**, 1381–1403.
Zhang, X., Heinz, M. G., Bruce, I. C., and Carney, L. H. (**2001**). "A phenomenological model for the responses of auditory-nerve fibers. I. Nonlinear tuning with compression and suppression," J. Acoust. Soc. Am. **109**, 648–670.
Zilany, M. S. A., and Bruce, I. C. (**2006**). "Modeling auditory-nerve responses for high sound pressure levels in the normal and impaired auditory periphery," J. Acoust. Soc. Am. **120**, 1446–1466.
Zilany, M. S. A., and Bruce, I. C. (**2007**). "Predictions of speech intelligibility with a model of the normal and impaired auditory-periphery," in *Proceedings of the Third International IEEE EMBS Conference on Neural Engineering* (IEEE, Piscataway, NJ), pp. 481–485.

Individual differences in auditory abilities

Gary R. Kidd, Charles S. Watson, and Brian Gygi
Department of Speech and Hearing Sciences, Indiana University, Bloomington, Indiana 47405

(Received 30 June 2006; revised 13 January 2007; accepted 7 April 2007)

Performance on 19 auditory discrimination and identification tasks was measured for 340 listeners with normal hearing. Test stimuli included single tones, sequences of tones, amplitude-modulated and rippled noise, temporal gaps, speech, and environmental sounds. Principal components analysis and structural equation modeling of the data support the existence of a general auditory ability and four specific auditory abilities. The specific abilities are (1) loudness and duration (overall energy) discrimination; (2) sensitivity to temporal envelope variation; (3) identification of highly familiar sounds (speech and nonspeech); and (4) discrimination of unfamiliar simple and complex spectral and temporal patterns. Examination of Scholastic Aptitude Test (SAT) scores for a large subset of the population revealed little or no association between general or specific auditory abilities and general intellectual ability. The findings provide a basis for research to further specify the nature of the auditory abilities. Of particular interest are results suggestive of a familiar sound recognition (FSR) ability, apparently specialized for sound recognition on the basis of limited or distorted information. This FSR ability is independent of normal variation in both spectral-temporal acuity and of general intellectual ability.
[DOI: 10.1121/1.2743154]

PACS number(s): 43.66.Fe, 43.66.Mk, 43.71.Es [MSS] Pages: 418–435

I. INTRODUCTION

The study of individual differences has a long history in psychology, most notably in the area of mental or cognitive abilities (e.g., Cattell, 1885; Galton, 1869, 1883; Binet, 1903; Thurstone, 1938; Carroll, 1993). Rather than simply documenting the range of abilities on various tasks, a primary goal of the subfield of differential psychology has been to identify the independent dimensions on which individuals vary. The identification of such dimensions, often termed *abilities*, has been a crucial component in the development of theories of intelligence and cognition. The nature of these abilities, and the extent to which they are utilized in different cognitive tasks, has been investigated with a wide range of tests (e.g., Sternberg, 1977; Hunt, 1978; Carroll, 1993). In some cases, experimental manipulations have both validated and further elucidated the nature and significance of specific abilities (e.g., Snow and Lohman, 1988).

The present study applies the individual differences approach to the study of auditory abilities. The goal was to identify the number and nature of distinct auditory abilities that underlie performance on a broad range of auditory discrimination and recognition tasks. The potential usefulness of such a determination is illustrated by the hypothesis that problems with speech recognition are a consequence of limited temporal (or spectral) processing ability. Rather than test this hypothesis by comparing speech recognition performance to performance on some single test of temporal resolution, a more valid approach would ask whether speech tests in general correlate strongly with a variety of different temporal and spectral processing measures. An extension of this approach would determine the relative contributions of each of several independent auditory abilities (assuming that such abilities do exist) to performance on speech recognition tasks. That strategy is employed in the research reported in this article.

"Auditory ability" is an abstract concept, comparable to general intelligence or physical strength. It is often a practical convenience to characterize individuals as more or less intelligent or strong, despite the uncertain referents of those concepts or *intervening variables* (MacCorquodale and Meehl, 1948). The validity of intervening variables depends on two properties, the first of which is the existence of a group of different measures (behavioral or psychophysical tests in the case of auditory abilities), all of which are significantly correlated with each other. The second property, which can elevate an intervening variable's status from a statistical inference to something more substantial, is a demonstrated neurophysiological correlate of those psychophysical measures, such as the rate of neural transmission, length of the cochlea, or volume of neural tissue in one of the auditory areas of the CNS. If multiple psychophysical measures of an ability are reliably demonstrated, there is a strong likelihood that a neural correlate will eventually be discovered. The work reported here addresses the first of these properties by using principal components analysis and structural equation modeling to identify distinct auditory abilities that underlie performance on a wide range of auditory tests.

Despite a number of earlier efforts, there is only modest agreement about the number or nature of auditory abilities. An exception to this generalization is found in one basic ability, auditory sensitivity as reflected in the audiogram. If persons with clinically significant hearing loss are included, this is clearly a primary dimension on which individuals differ and which predicts performance on many tasks, including speech recognition. This observation was made early in auditory research and has been reflected in the development of

the Articulation Index (French and Steinberg, 1947) and in more recent derivatives of that index (Pavlovic, 1984; Humes and Riker, 1992). Other basic auditory abilities have also been discussed, particularly ones that appear essential for tasks demanding resolution of complex stimuli, notably temporal acuity (e.g., Hirsh, 1959; Green, 1971; Watson, 2004) and spectral resolving power (e.g., Feth and O'Malley, 1977; Moore and Glasberg, 1986).

Auditory sensitivity is thus regarded as the major dimension on which listeners differ, reflecting the severe consequences of variation in this ability, and the frequency with which it is diminished, particularly in old age. Loss or developmental impairment of other auditory abilities, such as temporal or spectral resolving power, has also been postulated as the cause of disorders such as delayed language development or dyslexia (e.g., Tallal *et al.*, 1993). No specific measures of these abilities have, however, achieved the canonical status of the pure-tone audiogram as the measure of sensitivity.

The identification of basic auditory abilities using a purely psychophysical methodology requires the testing of a large number of people on a large number of auditory tests. The size requirements are dictated by well-established principles relating test reliability, intertest correlations, and the expected range of performance (see, for example, Gorsuch, 1983). The set of auditory tests must be sufficiently large that it can be reasonably assumed to exceed the number of discrete auditory abilities and include multiple tests of each ability.

Application of factor analysis, or principal components analysis, is a systematic method by which to identify subsets of similar tasks that provide measures of a common underlying ability. These techniques, originally developed by Spearman (1904), have been widely employed in the search for primary mental abilities by Thurstone (1947) and many more recent investigators. A limitation of this approach is that the discrete abilities that can be identified depend on the selection of measures included in the test battery. This requirement has made it difficult to compare the results of earlier factor analytic studies of auditory abilities, because of only moderate overlap among the tests that were employed. This is illustrated by the range of tests included in the Seashore battery (Seashore *et al.*, 1939, 1960), which includes many tests of rhythm and of the perception of musical segments that are not found in any of the other major studies, but it has only a few measures of basic auditory temporal or spectral discrimination abilities. Nevertheless, there has been sufficient commonality among the tests used in some of the earlier studies to suggest the existence of a few discrete primary auditory abilities, as described in the following section.

A. Previous factor analytic studies of auditory abilities

Factor analytic studies of auditory abilities conducted prior to the mid-1980s have been reviewed in an earlier article (Johnson, Watson, and Jensen, 1987). Ten such studies used widely differing collections of auditory tests, methodologies, and numbers and types of subjects, ranging from high school sophomores (Elliott *et al.*, 1966) to prison inmates (Stankov and Horn, 1980). While the variety of tests employed makes it difficult to compare across the studies, most of the studies supported the existence of multiple auditory abilities. These included the abilities to discriminate differences in pitch, intensity, and duration, plus the higher-level ability of auditory memory. One of the earliest and most complete of these studies was conducted by Karlin (1942), a student of Thurstone. Karlin found that the ability to recognize speech was surprisingly independent of the listeners' spectral and temporal discrimination abilities.

More recently, in a remarkable survey of factor analytic studies of human cognitive abilities, Carroll (1993) devoted one of 11 ability-specific chapters to "Abilities in the Domain of Auditory Reception." Carroll observed that, "The domain of individual differences in auditory receptive abilities has received relatively little attention in the factor analytic literature." Nevertheless, he was able to find 38 published datasets that provide some basis for establishing the structure of the domain of auditory abilities. Carroll considered only one of these sets to provide a reasonable statistical account of individual differences in speech processing (Hanley, 1956). Carroll's list of candidate factors included "hearing acuity" (auditory sensitivity measured with either tones or speech), "speech sound discrimination," "spectral, temporal and intensive discrimination," "speech perception under distortion," "cognitive relations" (among tonal patterns), "musicality," and several others.

Watson *et al.* (1976) observed large ranges of performance in tonal pattern discrimination studies, spurring an interest in individual differences. In an effort to learn more about individual listeners who showed exceptionally good or poor ability to detect changes in these spectral-temporal patterns, the Test of Basic Auditory Capabilities (TBAC) was developed (Watson *et al.*, 1982a, b). This battery has since been used in a number of investigations (e.g., Christopherson and Humes, 1992; B. Watson and Miller, 1993; Drennan and Watson, 2001; Surprenant and Watson, 2001; Jakobson *et al.*, 2003). The original TBAC included three single-tone discrimination tests, three tests of temporal pattern discrimination, and two speech tests. (These are subtests 1 through 8 of the current test battery and are described below.) In the first use of the TBAC, 127 adults with normal hearing were tested in a free field. There was virtually no association between performance on the nonspeech subtests and the two speech tests, supporting Karlin's (1942) finding of a dissociation between the recognition of "social sounds" and measures of auditory acuity. Later, Surprenant and Watson (2001) conducted a replication of this study in which three additional speech tests (identification of sentences, words, and CVs) were added to the eight TBAC subtests. Analysis of the data from this extended battery showed three factors, one for speech recognition, a second for nonspeech discrimination tests with both simple and complex stimuli, and a third for temporal-order discrimination, supporting the earlier findings of independence between performance with speech and nonspeech stimuli.

Although spectral and temporal resolving power vary significantly among normal-hearing adults, the average spectral and temporal abilities of normal-hearing adults are

clearly more than sufficient for the demands of speech recognition under quiet listening conditions. The lack of association between speech and nonspeech tests in these studies was not, however, due to a limited range of variation in speech recognition or in spectral-temporal acuity. Surprenant and Watson (2001) found a range of speech-to-noise ratios of about 7.0 dB for 50% correct identification of words in CID sentences (Davis and Silverman, 1970), and threshold measures for spectral and temporal discrimination differed by factors of from 3 to 10 times. Failure of measures of discrimination acuity to predict performance in the identification of nonsense syllables, words, or sentences pointed to some other ability as the source of the considerable range of speech recognition scores in noise. Another observation suggested that this hypothesized ability might be cognitive rather than auditory (more central than peripheral). Watson *et al.* (1996) found a modest correlation ($r=0.52$, $p<0.005$, $N=90$) between speech recognition by ear alone and that by eye alone (speechreading). While not a particularly strong association, this correlation is considerably larger than that between speech recognition and any single measure of spectral or temporal acuity obtained with nonspeech stimuli.

While this correlational evidence might appear to support arguments that speech recognition is "special" (i.e., explained by a speech-specific mechanism not utilized in the perception of nonspeech stimuli), the data are subject to other interpretations. In Surprenant and Watson (2001), as well as in virtually all prior factor analytic studies of auditory abilities, the speech tests differed from all of the nonspeech tests in at least two ways, in addition to the apparent differences between speech and most laboratory-generated stimuli. The first is that speech perception has been evaluated with recognition tests rather than with discrimination tests, placing quite different demands on the listeners. Discrimination tests generally reward the strategy of learning to focus on a dimension or property that is subject to change, while recognition tests encourage attention to larger patterns of spectral and temporal information. A second, more fundamental, difference is that the speech of one's native language is an extremely familiar stimulus, while the nonspeech stimuli used in these studies are novel laboratory-generated sounds. Most of a listener's knowledge about the nonspeech stimuli is gained during a relatively brief period of testing. However, listeners have extensive explicit and tacit knowledge about the structure of speech sounds, including the many constraints (linguistic and physical) that govern this class of sounds. An individual's ability to effectively use this knowledge is likely a major factor accounting for differences in speech recognition performance, especially in tests that require listeners to identify speech stimuli that have been degraded or masked by noise. Under these conditions, listeners who can use their knowledge to develop better attentional strategies (or better "guessing" strategies) might have an advantage over others with equally acute spectral and temporal resolution.

B. Expanded test battery for the present project

Despite the support provided by Karlin (1942) and Surprenant and Watson (2001) for the independence of the abilities to process speech and nonspeech sounds, there is ample evidence in the literature that temporal and spectral acuity are necessary to distinguish among the sounds of speech [e.g., the long line of research with vocoders, begun by Dudley (1939)]. Thus, persons with significantly worse than average spectral or temporal acuity might reasonably be expected to be correspondingly worse at recognizing speech. It is a common assertion among audiologists that, although amplification can return sensitivity to normal values, speech perception by the hearing impaired remains below that of nonimpaired persons because their auditory resolving power is inadequate. While this assertion may be true for persons with impaired hearing, it does not necessarily follow that the variation in resolving power among those with normal sensitivity will predict their speech recognition abilities.

An obvious problem with the earlier studies is the limited number of tests of spectral and temporal acuity, perhaps failing to include other discrimination tests that would predict speech recognition. Several spectral and temporal measures were therefore added to yield an expanded TBAC battery, including: (1) ripple-noise discrimination, a measure of spectral resolution as a function of the depth of troughs in the frequency spectrum (Yost *et al.*, 1978); (2) detection of amplitude modulation in Gaussian noise, measured at different modulation rates (a measure of temporal resolution first studied by Viemeister, 1979); and (3) detection and discrimination of temporal gaps, a test reported to be correlated with speech recognition (e.g., Glasberg and Moore, 1989; Snell *et al.*, 2002; Tyler *et al.*, 1982). The association between gap detection and speech recognition is perhaps the most widely cited finding in support of a relation between nonspeech and speech processing abilities. However, that association has been observed only with older or hearing-impaired listeners, and some studies have failed to find it even with these populations (e.g., Nelson, Nittrouer, and Norton, 1995; Strouse *et al.*, 1998). To some extent, the discrepancies may be due to differences in the details of the tests used to assess nonspeech and speech abilities (e.g., the duration and complexity of the stimulus in which a gap appears, the location of the gap, the type of speech stimuli and background noise). The present study assesses the relation between temporal processing and speech recognition in a large population of normal-hearing listeners, using several measures of temporal processing in addition to gap detection and gap discrimination, and several measures of speech recognition. The use of a large population ($N=340$) was expected to provide a wide range of performance on all measures and allows for a more definitive assessment of the relation among speech and nonspeech abilities.

An additional test was added to the battery to evaluate the hypothesis suggested earlier, that the difference between performance on speech and nonspeech tests is a consequence of differences in the familiarity of the stimuli used in these tests. For this purpose, a recognition test for familiar *nonspeech* sounds was added to the battery. The stimuli were environmental sounds produced by both animate and inanimate sources (e.g., dogs barking, doors slamming, cars starting). These sounds were presented in Gaussian noise, follow-

TABLE I. A summary of the 19 subtests of the extended version of the Test of Basic Auditory Capabilities (TBAC-E).

Subtest	Stimulus	Detect/identify	Manipulation
(1) Pitch discrimination	250-ms 75 dB SPL 1000 -Hz tone	ΔF	Frequency
(2) Single-tone intensity discrimination	250-ms 75 dB SPL 1000 -Hz tone	ΔI	Intensity
(3) Single-tone duration discrimination	100-ms 75 dB SPL 1000 -Hz tone	ΔT	Duration
(4) Pulse-train discrimination	Rhythmic sequence of six 20-ms 1000-Hz tones	ΔT	Relative duration of pauses between tones
(5) Embedded test-tone loudness	Sequence of nine (or eight) contiguous 40-ms tones (300–3000 Hz)	Presence/absence of middle tone	Duration of middle tone
(6) Temporal order for tones	Sequence of four contiguous tones (550–710 Hz)	Change in temporal order of middle two tones	Duration of middle two tones
(7) Temporal order for syllables	Sequence of four contiguous CV syllables	Change in temporal order of middle two syllables	Syllables duration
(8) Syllable identification	VC syllables in cafeteria noise	Syllable (3AFC)	Natural variation in cafeteria noise
(9–12) Sinusoidal amplitude modulation detection	500-ms noise sample with sinusoidal amplitude modulation (8, 20, 60, and 200 Hz)	Presence of AM	Modulation depth
(13) Ripple noise discrimination	500-ms noise sample with spectral "ripple"	Presence of ripple	Ripple depth
(14) Gap detection	750-ms noise sample	Silent gap in temporal center	Gap duration
(15) Gap duration discrimination	750-ms noise sample	Change in gap duration	Gap duration
(16) Nonsense syllable identification	Nonsense CVC syllable in noise	Syllable (4AFC)	S/N
(17) Word identification	One, two, and three-syllable words, in noise	Word (4AFC)	S/N
(18) Sentence identification	Meaningful sentences (4 to 10 words) in noise	Sentence (free recall)	S/N
(19) Environmental sound recognition	Familiar environmental sounds in noise	Familiar sounds (3AFC)	S/N

ing procedures similar to those used with the speech stimuli.

The new expanded test battery consists of 19 tests, as listed in Table I. Although there are certainly other tests or stimulus conditions that might have provided additional useful information, no earlier study has assessed performance on as wide a range of speech and nonspeech tests with a population of this size. The availability of SAT scores for a large portion of the subject population made it possible to assess the extent to which performance on these tests is determined by general intellectual abilities.[1]

II. METHOD

A. Subjects

The listeners were 340 adults (100 men, 240 women, ages 18–31, mean=22.4) recruited primarily from the Indiana University student population, but also from the nonstudent population in and around Bloomington. All tested within normal limits [<20 dB HL from 250 to 4000 Hz (ANSI, 1989)] on a hearing screening. All student listeners gave permission to access their academic grade point average and their scores (verbal and quantitative) on the Scholastic Aptitude Test (SAT). The listeners were paid for their participation.

B. Stimuli

Stimuli were delivered using a digital audio tape deck (Panasonic, SV3500). The stimuli were presented diotically over EAR 3A insert earphones. All stimuli were presented at 75 dB SPL, with the exception of the increased-level comparison tones in the intensity discrimination test.

This expanded version of the Test of Basic Auditory Capabilities (TBAC, Watson *et al.*, 1982a, b) includes the original eight subtests, plus 11 new subtests that were added to examine auditory abilities that may not have been addressed by the original battery. Trials in each of the subtests, except for speech identification, are structured in a modified 2AFC format in which a standard stimulus is followed by two test stimuli, one of which is different from the standard. The listeners use a computer keyboard to indicate their response selections. Trials are arranged in groups of six, and the level of difficulty is systematically increased from trial to trial, within each group, in logarithmic steps. For most tests, eight levels of difficulty are tested over 72 trials, presenting the six easiest levels in the first 36 trials, followed by an increase in difficulty of two log steps for trials 37–72. The actual numbers of easy and difficult groups of trials varies among the subtests, as described below. Standard stimuli in all subtests were presented at 75 dB SPL.

Details of the 19 subtests of the Test of Basic Auditory Capabilities, Expanded (TBAC-E)

Subtest 1: *Pitch discrimination*. The standard is a 1-kHz, 250-ms tone. The values of the frequency increments range from 2 to 256 Hz in equal logarithmic steps.

Subtest 2: *Single-tone intensity discrimination*. The standard is a 1-kHz, 250-ms tone. The increments range from 0.5 to 8.0 dB.

Subtest 3: *Single-tone duration discrimination*. The standard is a 1-kHz, 100-ms tone. Increments in duration range from 8 to 256 ms in equal logarithmic steps.

Subtest 4: *Pulse-train discrimination*. The standard stimulus consists of six 20-ms pulses of a 1-kHz tone. These pulses are arranged in three pairs, with 40 ms of silence between members of a pair and 120 ms between pairs. The temporal structure of the "different" sequence is varied by increasing the separation between members of each pair, with a corresponding decrease in the between-pair time (and, thus, a constant interval between the first tones in each of the successive pairs). Increases in within-pair separation ranged from 5- to 50-ms increments, in equal logarithmic steps. Thus, the first, third, and fifth tones are fixed in time, while the onsets of the second, fourth, and sixth tones are delayed by varying amounts. This is a minimal test of the ability to detect changes in the relative timing of events in a repeated temporal sequence (i.e., rhythm). However, it should be noted that, in principle, this test can be performed solely on the basis of changes in absolute durations.

Subtest 5: *Embedded test-tone loudness*. Subjects listen for one member of a sequence of nine tones with frequencies ranging from 300 to 3000 Hz. A different, randomly selected series of nine tones is presented on each trial. The task is to detect the presence of the fifth tone in the sequence. The tone is absent in the standard. The duration of all tones except the fifth, or target tone, is 40 ms. The test is made more difficult by reducing the duration of the target tone from 200 to 10 ms, in equal logarithmic steps. The name of this test derives from the listener's experience. Only the duration of the target tone is directly manipulated.

Subtest 6: *Temporal order for tones*. The task is to discriminate the order in which two equal-duration tones (550 and 710 Hz) are presented. The duration of the two tones is varied from 20 to 200 ms in equi-log steps. The tones are presented without a gap between them and are preceded and followed, without gaps, by 100-ms "leader" and "trailer" tones at 625 Hz.

Subtest 7: *Temporal order for syllables*. This is a speech analog to subtest 6, in which the listener is to discriminate the syllable sequence /ta/ka/ from /ka/ta/, when the two CV syllables are preceded and followed by the syllables /fa/ and /pa/. The listener's task is thus to discriminate /fa/-/ta/-/ka/-/pa/ from /fa/-/ka/-/ta/-/pa/, and subjects are told that only the middle syllables vary. The duration of the syllables was varied from 75 to 250 ms in five steps.

Subtest 8: *Syllable identification*. This is a subset of the Dubno and Levitt (1981) nonsense syllable test. Nonsense VC syllables are presented in cafeteria noise and a three-alternative forced-choice identification procedure is used. Syllables are presented in a carrier phrase (e.g., "You will mark *oothe* please"). The original test was modified by reducing the number of alternatives on each trial to the correct response plus the two most likely errors, as shown in confusion matrices provided to us by the authors.

Subtests 9–12: *Sinusoidal Amplitude Modulation (SAM) noise discrimination*. Fifty-four independent 500-ms samples of Gaussian noise are sinusoidally amplitude modulated at four different rates: 8, 20, 60, and 200 Hz. For each AM rate, eight different modulation depths are used, ranging from −18 to −32 dB for the slowest modulation rate, to −10 to −24 dB for the fastest rate. [Modulation depth is here expressed as $20\log(m)$, where m is a modulation index that ranges from 0.0 to 1.0.] The modulated and unmodulated stimuli are equated for total rms energy. The standard is always a 500-ms broadband noise with no amplitude modulation, and one of the two test stimuli is modulated. Each modulation rate is tested in separate blocks of 54 trials, in the order of increasing modulation rate. Each block consists of four 6-trial groups with the largest modulation depths, followed by five 6-trial groups with the smallest depths.

Subtest 13: *Ripple noise discrimination*. Seventy-two independent 500-ms digital samples of Gaussian noise are low-pass filtered at 3000 Hz. Sinusoidal ripples are created in the power spectrum by delaying the samples by 5 ms and adding them to the input. Before adding, the delayed signal is attenuated by an amount ranging from 0 dB (maximum ripple) to 14 dB (slight ripple), in seven 2-dB steps. The resulting eight "rippled" (i.e., spectrally modulated) noises are then high-pass filtered at 300 Hz and are equated for rms. The standard is always a 500-ms broadband noise with the same bandpass filtering as the "rippled" samples, but with a uniform power spectrum. One of the two comparison patterns is rippled. The test consists of 72 trials: six easier 6-trial groups (deep ripple: 0–10-dB attenuation) followed by six more difficult 6-trial groups (shallower ripple: 4–14-dB attenuation).

Subtest 14: *Gap detection*. Seventy-two independent 750-ms digital samples of Gaussian noise have gaps of silence of eight different durations at their temporal centers. The gap durations range from 0.5 to 64 ms, in log steps, while the total durations remained constant. The noises have 0.5-ms cosine ramps at the beginning and end of the silent intervals. The standard is always a 750-ms broadband noise that has no gap in it. One of the two comparison stimuli contains a gap. The test consists of 72 trials: six easier 6-trial groups (gaps ranging from 2 to 64 ms) followed by six more difficult 6-trial groups (gaps from 0.5 to 16 ms).

Subtest 15: *Gap-duration discrimination*. Seventy-two independent 750-ms digital samples of Gaussian noise have silent gaps of 40 ms placed at their temporal centers. Increments in gap duration range from 4 to 200 ms, in log steps. The total duration of the noise is kept constant while the gap duration increases, resulting in an increase in total duration with increases in gap duration. The noise bursts have a 0.5-ms cosine ramp at onset and offset. The standard is always a 750-ms broadband noise with a 40-ms silent gap at its temporal center. One of the comparison patterns contains a larger gap (with total noise duration remaining constant), while the other is identical to the standard. The test consists of 72 trials: six easier 6-trial groups (gap increments ranging from 12 to 200 ms) followed by six more difficult 6-trial groups (gap increments from 4 to 65 ms).

Subtests 16–18: *Speech tests*. Equal speech and noise levels for 0-dB speech-to-noise ratios were set by digitally recording at a peak VU-meter reading equal to that of the Gaussian noise samples. Various speech-to-noise ratios were achieved by attenuating the speech. Presentation levels were

determined by setting a 1-kHz calibration tone, recorded with the same rms value as the Gaussian noise, to 75 dB SPL. Talkers were two male and two female students in the Department of Theatre and Drama at Indiana University, all of whom were native speakers of Standard American English (Midwestern dialect). Responses for all speech tests are recorded either on paper response sheets or by using a computer keyboard and monitor. In either case, the response alternatives are presented on each trial for the nonsense-syllable and word tests, while the sentence test is given in open-set format.

For all speech tests, the speech-to-noise (S/N) ratios were adjusted after the first 146 subjects had been tested and estimates of psychometric functions were examined. (Original and adjusted values are provided for each test below.) The new S/N values were selected to minimize testing at values that produce near-chance or near-perfect performance and to provide equal increments in percent correct between each step. New S/N values were interpolated from the psychometric functions, for values of percent correct from 40% (15 percentage points above chance) to 90% in steps of 12.5 percentage points. For the sentence identification test (which used an open-ended response protocol), S/N values were interpolated for percent-correct values from 20% to 90% in steps of 17.5 percentage points. The resulting S/N values for each subtest are given below.

Subtest 16: *Nonsense syllable identification*. A set of 100 nonsense CVC syllables uses all of the standard English consonants and vowels in random combinations. The stimuli are mixed with broadband Gaussian noise at five speech-to-noise ratios (initial values: −3, −7, −11, −15, −19 dB; adjusted values: −0.4, −5.1, −8.4, −11.5, −15.4 dB). A four-alternative forced-choice procedure is used. Three foils (incorrect CVCs) for each of the stimuli were generated by changing either the initial consonant, the vowel, or the final consonant (one for each foil). The response options are presented, and listeners are asked to indicate the sound that was presented on each trial. Listeners are given 3 s to indicate their response. The set of 100 stimuli is presented twice in separate blocks of trials using two different random orders. Stimuli are presented with decreasing S/N values within each group of five trials.

Subtest 17: *Word identification*. The set of 100 stimuli consisted of 60 one-syllable, 25 two-syllable, and 15 three-syllable words. The stimuli were mixed with broadband Gaussian noise at five speech-to-noise ratios (initial values: −3, −7, −11, −15, −19 dB; adjusted values: −2.9, −7.8, −11.1, −14.4, −18.3 dB). The method was the same as that for nonsense syllables, except for a slightly different strategy for generating foils. Three foils were generated by modifying the beginning, middle, or ending portion of each word (one for each foil). To accommodate the constraints of English words with different numbers of syllables, it was necessary for the modified portion to range from single consonants or vowels to full syllables.

Subtest 18: *Sentence identification*. A set of 40 sentences was composed for this test. The number of words per sentence ranged from four to ten. All sentences consisted of familiar words and concepts. Sentences were highly predictable so that listeners' abilities to use context would be evaluated by this test. The stimuli were mixed with broadband Gaussian noise at five speech-to-noise ratios (initial values: −4, −6, −8, −10, −12 dB; adjusted values: −6.3, −7.9, −9.0, −10.1, −11.7 dB). Listeners were allowed 6 s to write the words they could identify, in order, on a response sheet. All 40 sentences were presented twice, using two random orders in two blocks of trials. As with the other speech tests, stimuli were presented with decreasing S/N values within each group of five trials. The total number of correctly identified words (regardless of order) was recorded for each sentence.

Subtest 19: *Environmental sound identification*. Twenty-five digitally recorded sounds were selected from two high-quality sound effects collections (Hollywood Edge and Sound FX The General). All were 16-bit stereo sounds sampled at 44.1 kHz. The sounds were converted to monaural format and presented diotically. The particular sounds were selected from among several hundred options, based on criteria of familiarity, detectability, and pairwise discriminability. These 25 sounds and several others from the collections cited above have been extensively studied by our research group (Gygi, 2001; Gygi, Kidd, and Watson, 2004, in press). The sounds were all perfectly recognizable at high event-to-noise ratios (e.g., in excess of +20 dB). To simulate difficult listening conditions, samples were prepared by mixing each sound with Gaussian noise at eight different event-to-noise ratios (Ev/N). Equal sound and noise levels for 0 dB Ev/N were achieved by equating the rms values. The level of the environmental sound was adjusted to create other Ev/N values. The levels were selected so that the probability of identification would range from near chance to near perfect, which was achieved in seven 3-dB steps. The range for each event is centered on the Ev/N level corresponding to an average probability of detection of 0.66, as determined in a preliminary experiment with a group of four normal-hearing college students. Minimum Ev/Ns ranged from −31.6 to −16.9 across the set of sounds. A 3AFC format is employed, in which one sample is played to the listeners who then select the correct response from among three alternatives. The two incorrect alternatives on each trial are those sounds from the same 25-stimulus catalog that were determined in pilot studies to be most likely to be confused with the sound actually presented. Trials are presented in groups of six with increasing Ev/N levels within each group. The test consists of two 150-trial blocks, each with a different a random sequence of the 25 sounds. In the first block, the sounds are presented at the six highest Ev/N levels. In the second block, the sounds are presented at the six lowest Ev/N ratios.

C. Procedures

Participants were tested in a sound-treated room in groups of up to 12 listeners per session. Testing of each participant was conducted over four 90-min sessions on consecutive weekdays (one session per day). The sequences of tests and of stimuli within each test were the same for all participants. The rationale is that, in an individual differences study, the goal is to eliminate variance due to any factors other than differences among the subjects. Subtests 1–8 were

TABLE II. Split-half reliability coefficients for the 19 TBAC subtests.

TBAC-I Subtests	Reliability	Added subtests	Reliability
Embedded tone	0.723	SAM 8 Hz	0.786
Pitch	0.819	SAM 20 Hz	0.815
Pulse train	0.816	SAM 60 Hz	0.702
Duration	0.755	SAM 200 Hz	0.717
Loudness	0.878	Ripple	0.759
Temporal order	0.807	Gap discrim.	0.560
Syllable ID	0.501	Gap detect.	0.612
Syllable sequence	0.766	Nonwords	0.787
		Words	0.635
		Sentences	0.795
		Env. sounds	0.827

administered in the first session, subtests 9–15 in the second, subtests 16–18 in the third, and subtest 19 in the final session. All tests began with detailed recorded instructions, which included two practice trials. Subjects were encouraged to make their "best guess" on each trial, regardless of their confidence level.

III. RESULTS

A. Reliability of TBAC measures

Christopherson and Humes (1992) examined the reliability of the original eight TBAC subtests. The tests were administered multiple times to the same listeners and were found to be reliable; Cronbach's alpha values were above 0.7 for all but the Syllable Identification subtest, which had a value of 0.58. Performance on all subtests changed little over six repeated administrations of the TBAC. The new TBAC tests employed here were constructed following the same principles as the earlier tests and utilized stimuli and tasks that have been used extensively in previous research. The reliability of all subtests was estimated as part of this research project, using a split-half procedure, rather than the repeated-test procedure used by Christopherson and Humes. Following a resampling strategy (see Good, 2006), correlations for 1000 randomly selected pairs of split halves were computed for each subtest, and reliability was computed using the Spearman-Brown prediction formula applied to the mean correlation. The results are shown in Table II. Consistent with the findings of Christopherson and Humes, reliability coefficients for all of the first eight subtests were above 0.7, except for the Syllable Identification subtest. Reliability for the newer subtests were in the same range, with only three of the subtests falling below 0.7 (with the lowest coefficient at 0.56). Thus, reliability of most subtests is fairly high, and the less reliable subtests have more than adequate reliability for the present purposes. As with any test, performance on these tests may be influenced by factors other than the abilities they were designed to measure (such as general intelligence or motivation). However, these tests appear sufficiently stable and reliable for use in estimating the associations among them. The question of the nature and number of the abilities that account for performance on these tests is addressed in Sec. III C.

B. Range of abilities

The range of performance on each of the 19 subtests for 338 subjects is summarized in Table III. Two subjects were excluded from the analysis because of extremely low performance on one or more subtests, suggesting either misunderstanding of the instructions or lack of motivation. The entries are estimated threshold values for the group of subjects in each decile, with decile assignment based on subjects' overall percent correct scores for each subtest. Thresholds are based on estimates from psychometric functions fitted to data for all subjects within a decile, as shown in Fig. 1. This method was chosen in preference to fitting the psychometric functions to data from individual subjects because of the

TABLE III. Threshold (70% correct) estimates for each decile for the 19 subtests of the TBAC-E.

Subtest	1	2	3	4	5	6	7	8	9	10
(1) Pitch (ΔF, Hz)	36.27	15.45	13.99	12.55	10.96	8.78	7.88	6.59	4.95	3.08
(2) Intensity (ΔI, dB)	2.37	1.16	0.84	0.70	0.65	0.55	0.48	0.45	0.34	0.29
(3) Duration (ΔT, ms)	67.20	41.34	30.58	27.47	24.71	22.54	19.67	18.04	14.11	9.68
(4) Pulse train (ΔT, ms)	22.83	14.92	11.53	10.38	9.69	7.94	7.26	6.21	4.93	3.22
(5) Embeded tone (T, ms)	55.95	42.64	35.89	33.54	28.16	25.15	21.73	19.49	15.75	10.63
(6) Temporal order tones (T, ms)	119.97	77.73	64.74	56.97	51.22	45.99	41.65	35.73	28.13	21.96
(7) Temporal order syllables (T, ms)	256.11	159.45	135.64	114.15	105.42	100.10	86.94	87.59	80.89	73.70
(8) Syllable ID: VC [$P(c)$]	0.61	0.69	0.71	0.73	0.74	0.75	0.77	0.78	0.80	0.84
(9) SAM 8 Hz (mod. depth, dB)	−16.34	−20.52	−22.60	−23.48	−25.02	−25.90	−26.48	−27.31	−28.25	−29.73
(10) SAM 20 Hz (dB)	−14.05	−16.52	−20.84	−22.60	−23.94	−24.53	−26.08	−26.81	−27.61	−30.87
(11) SAM 60 Hz (dB)	−12.95	−17.70	−19.82	−20.60	−21.38	−22.38	−22.94	−23.55	−24.37	−25.88
(12) SAM 200 Hz (dB)	−7.08	−14.30	−16.32	−16.97	−17.34	−17.94	−18.52	−19.13	−20.12	−22.04
(13) Ripple noise (dB)	0.64	−1.75	−3.58	−4.70	−5.06	−5.90	−7.43	−8.08	−8.65	−11.00
(14) Gap detection (T, ms)	5.10	3.06	2.59	2.34	2.00	1.83	1.54	1.40	1.09	0.71
(15) Gap discrimination (ΔT, ms)	67.68	48.91	46.56	41.73	37.04	31.39	30.93	28.24	24.36	14.61
(16) Syllable ID (CVC) (S/N)	−4.49	−5.77	−6.53	−7.08	−7.62	−8.03	−8.45	−8.82	−9.41	−9.35
(17) Word ID (S/N)	−7.37	−9.23	−9.56	−9.96	−10.46	−10.73	−11.04	−11.54	−11.99	−13.39
(18) Sentence ID (S/N)	−7.07	−7.70	−7.88	−8.05	−8.25	−8.34	−8.39	−8.57	−8.84	−9.22
(19) Environmental sounds ID (S/N)	−10.73	−11.98	−12.42	−12.91	−12.94	−13.15	−13.53	−13.72	−14.17	−14.92

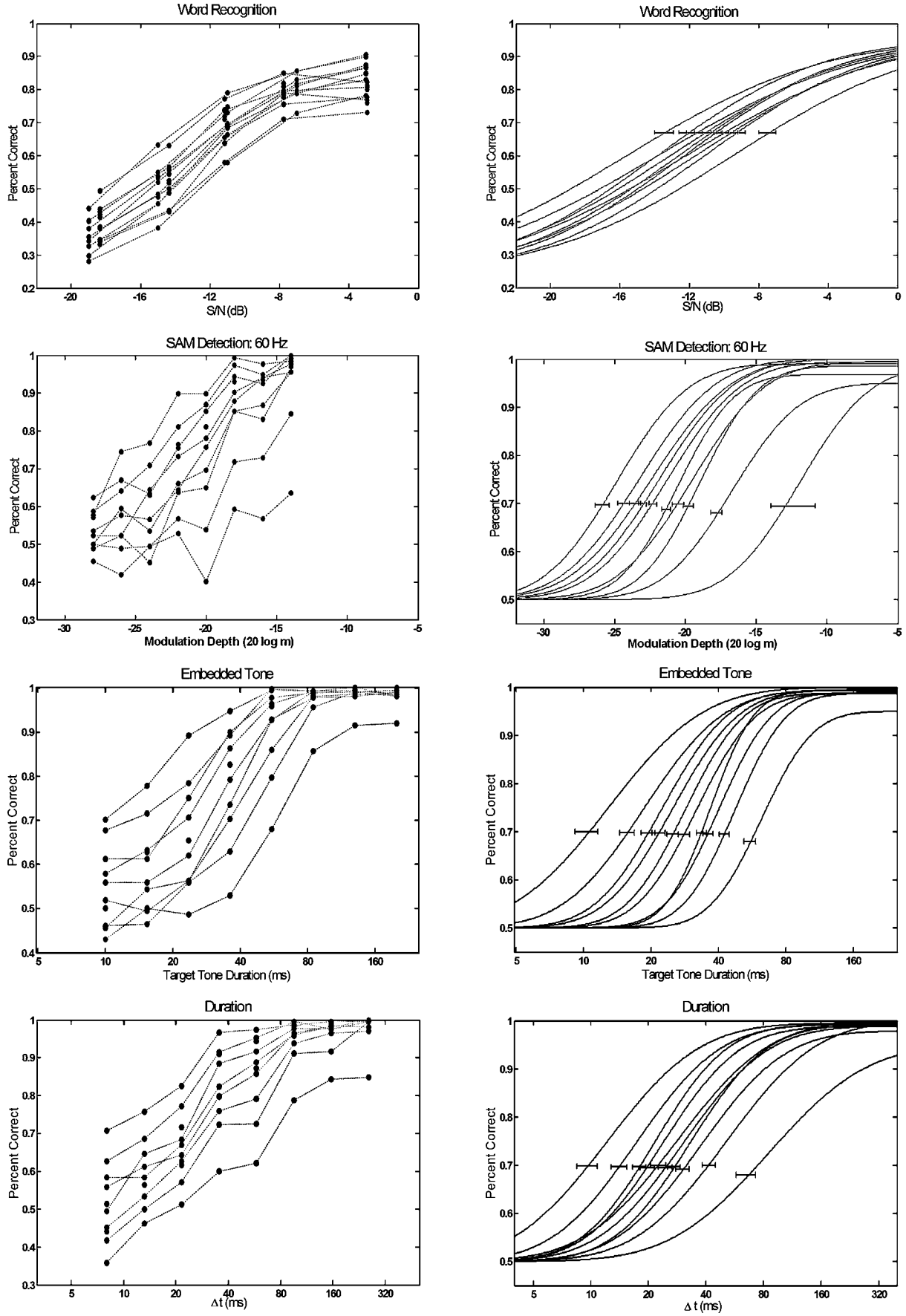


FIG. 1. Performance of each decile group on 18 of the 19 TBAC-E subtests. (The required data were not available for the Syllable Identification subtest.) Both percent-correct data and fitted functions are shown for the first four subtests. Panels on the left show percent correct at each stimulus level. Panels on the right show fitted functions and 90% confidence intervals for the threshold estimates derived from the fitted functions. Fitted functions and confidence intervals are shown for the remaining subtests.

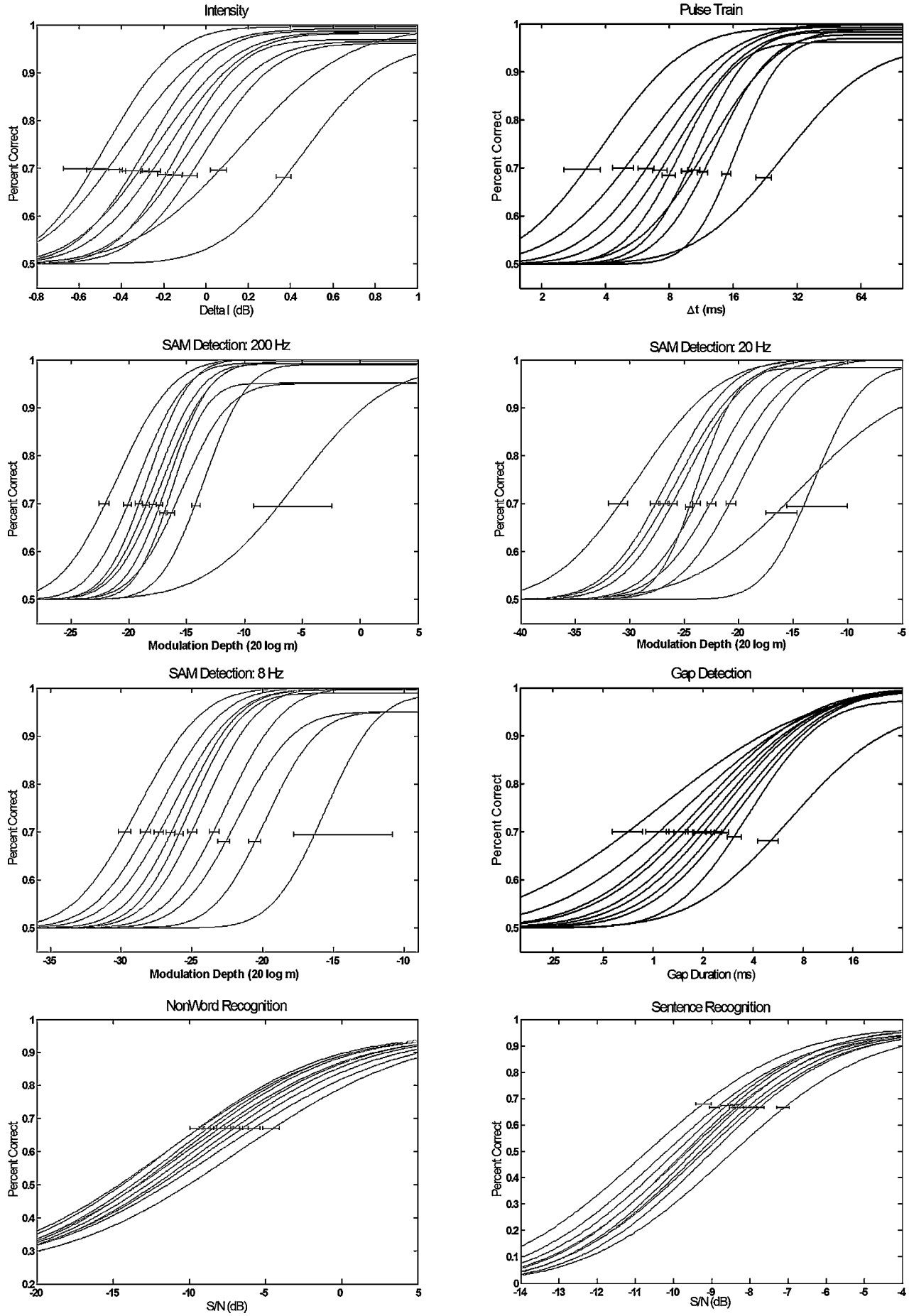


FIG. 1. (*Continued*).

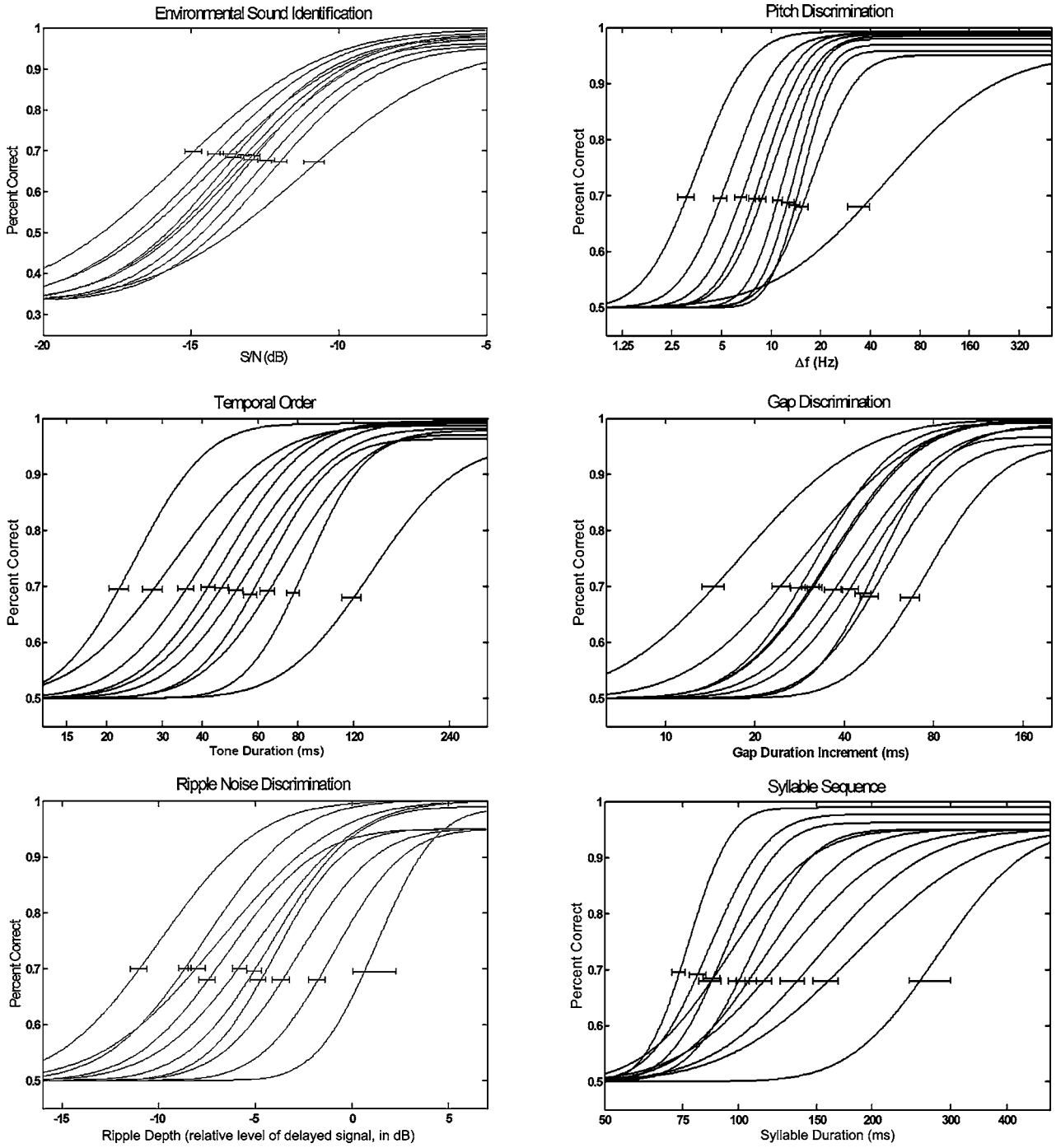


FIG. 1. (*Continued*).

limited number of trials per subject. Thus, the range of threshold values in Table III does not represent the full range of thresholds for individual listeners. The underlying distributions were approximately normal, with significant deviations from normality occurring primarily at the low end of the distribution. Because equal numbers of subjects were included in each decile (by definition), deciles near the tails of the distribution include a greater range of performance than those near the middle. Although exceptional performance was observed at both ends of the distributions, all distributions were unimodal, with no evidence of separate populations of listeners with exceptionally low or high acuity on any measure.

Thresholds were estimated by fitting a sigmoid function (cumulative Gaussian) to the data for each decile using a maximum-likelihood method developed by Wichmann and Hill (2001a, b). This method uses a parametric bootstrap technique (a Monte Carlo resampling method with maximum-likelihood parameter estimates) that allows for the estimation of confidence intervals for threshold estimates. In this application, the lower limit of the function was constrained to be at chance (which varied with the task), and lambda (a lapse parameter that sets the upper limit of performance) was constrained to be within the range of 0.0 to 0.5. The threshold was defined as the stimulus level corresponding to 70% correct (for cases in which lambda=0.0) or

slightly less for nonzero values of lambda. This represents a type of correction for inattention in that the actual percent-correct value at which the threshold is estimated is a constant proportional distance between chance and the upper limit of the fitted function (defined by lambda).

Perhaps the most remarkable aspect of the data in Table III is the broad range of performance on many of the tests. The ranges of threshold values are close to those reported by Surprenant and Watson (2001) for the first seven TBAC subtests, with substantial differences occurring for only two of the subtests. The best performance on the Embedded Tone subtest (i.e., in the highest deciles) was considerably better than in the earlier study, as was the worst performance (first decile) on the Duration subtest. In general, performance of the top 10% on each subtest approaches that found with highly trained listeners using more traditional psychophysical procedures, while the worst 10% performed quite poorly. For example, the top group was able to detect a 0.3% change in frequency (or 5.3 cents), while the bottom group required a 3.6% change in frequency (or 61.7 cents). Thresholds in terms of duration, or of changes in duration, in temporal tasks increased from 3.5 to 7.0 times across the deciles, and thresholds for modulation depth and ripple depth ranged from 12 to 16 dB. The range of threshold S/N values for recognition of familiar sounds (speech and environmental sounds) may seem small in comparison (2 to 6 dB) to those of the unfamiliar laboratory test sounds. But, it should be considered that because of the steep psychometric functions for these stimuli, relatively small changes in S/N translate into large changes in percent-correct identification. For example, at a S/N of −10 dB, the worst subjects could understand only about 18% of the words in sentences, while the best subjects could understand about 52%. Others (Surprenant and Watson, 2001; Bronkhorst and Plomp, 1992) have found a larger range for sentence recognition thresholds of roughly 6 to 7 dB. This represents a substantial range in the ability of normal-hearing listeners to understand speech in noise.

Fitted psychometric functions for each decile are shown in Fig. 1 for 18 TBAC subtests.[2] [No slope or threshold estimates could be derived for the Syllable Identification subtest (subtest 8) because of the stimulus values used.] Raw data and fitted functions are shown in separate plots for four of the subtests.[3] These four subtests were chosen because they are representative of the different patterns of results in the full set of tests and because they represent each of the four factors identified in the factor analysis described below. The 90% confidence intervals indicated for each of the threshold estimates show consistent accuracy, with the exception of the first decile for the 8-Hz SAM subtest. The reduced accuracy results from the restricted range of above-chance performance for this decile, suggesting that even the easiest stimulus levels were too challenging for the worst listeners. This was an issue for all SAM subtests, the Ripple subtest, and the Syllable Identification subtest (subtest 8). Despite the reduced accuracy, these are valid threshold estimates. They indicate a low level of performance at the same stimulus levels that result in near-perfect performance for most listeners. Such a large difference in performance at these levels may indicate that, for these tasks, the worst subjects are not simply less sensitive to the to-be-detected stimulus dimension; they may be using a different listening strategy that is inefficient or inappropriate for the task.

The families of psychometric functions are quite orderly, with thresholds decreasing consistently across the deciles. The largest change in threshold typically occurs between the first and second deciles. Slopes of functions for adjacent deciles are generally similar across the deciles, and any changes in slope tend to be gradual. For 13 of the subtests, slopes tended to increase slightly with increasing thresholds, while the opposite was true for the remaining cases. Thus, there was no consistent tendency for slopes to change systematically with the ability of the listeners.

C. Identification of basic auditory abilities

The considerable variance on most of the subtests suggests a broad range of abilities. It also provides an opportunity to examine the covariance as a means of identifying the number of distinct abilities that underlie performance on this collection of tests. As a first step in an attempt to identify these underlying abilities, a principal components analysis was conducted using arcsine-transformed overall percent-correct scores for each of the 338 subjects and all 19 subtests. This analysis produced four orthogonal factors with eigenvalues greater than 1.0, which together accounted for approximately 50% of the total variance. Attempts to extract additional factors produced uninterpretable factors with eigenvalues less than 1.0. The factor loadings (varimax rotated) for each of the subtests are shown in Fig. 2.[4] In this figure, the tests are grouped according to the factor on which they have the highest loading. Within each factor, the tests are ordered according to the strength of the loading on that factor. The cross loadings of each test with the other three factors are also shown. Examination of the cross loadings reveals that, within each factor, there are some tests with high primary loadings and low cross loadings, while other tests have one or more relatively high cross loadings. Thus, some tests appear to be close to pure measures of each of the underlying factors, while others are influenced by more than one underlying factor or ability.

The factor names were chosen to reflect the tests with the highest loadings on each factor and also the variety of tests that have their primary loading on that factor. The four factors can be interpreted as four independent abilities that underlie performance on this collection of auditory tasks. Although these will be referred to as auditory abilities, this is not meant to imply that these abilities are exclusively in the auditory domain. The same abilities may also be important for the performance of tests with stimuli in other modalities, especially when the test requires the discrimination of temporal structure or the identification of stimuli presented in noise. The nature of the four factors or abilities is discussed in more detail below.

1. Loudness-duration factor

The strongest loadings for intensity and duration discrimination on this factor suggest a basis in the sensitivity to

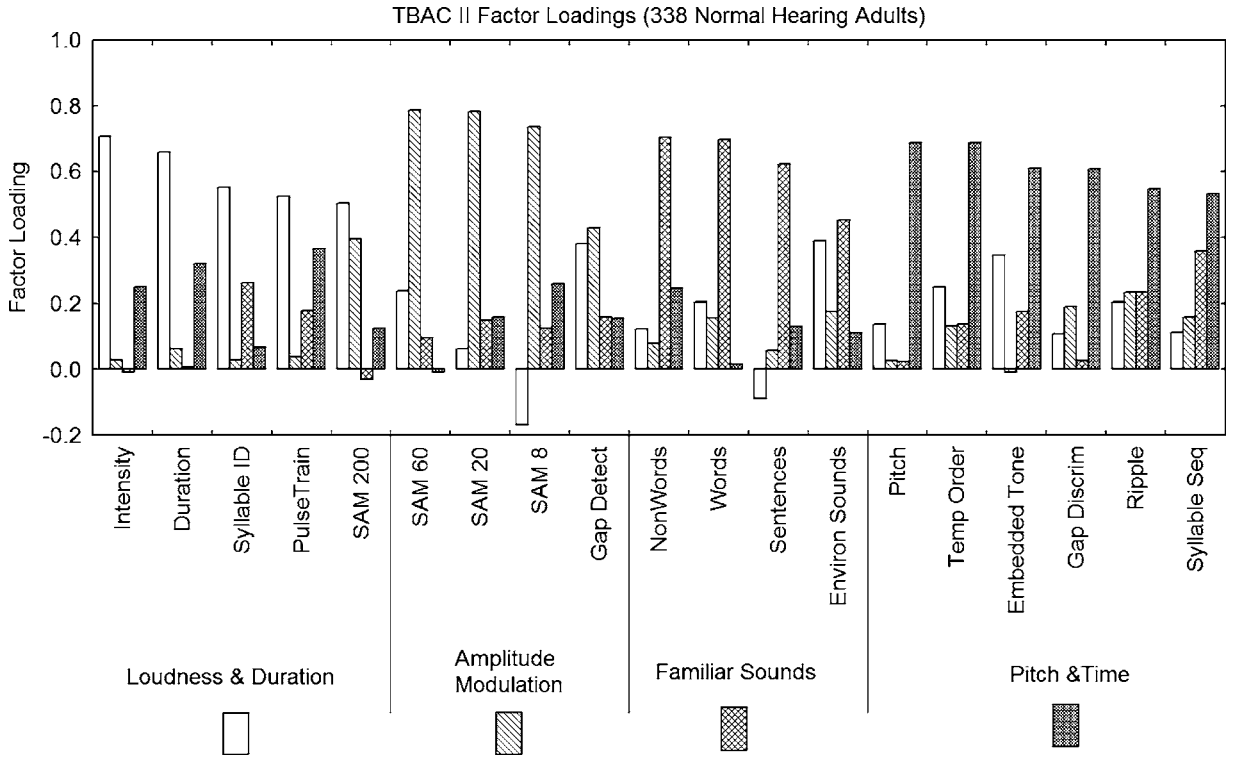


FIG. 2. Factor loadings for each of the 19 TBAC-E subtests. Subtests are grouped according to the factor on which they have the highest loading and ordered within each factor according to the magnitude of the primary loading on that factor.

changes in overall energy as an explanation of the commonality of these subtests. The remaining three subtests that have their highest loadings on this factor have their highest cross loading on the factor to which they seem logically to belong: Syllable ID on the Familiar Sound Factor, Pulse Train on the Pitch-Time Factor, and SAM 200 on the Amplitude Modulation Factor. However, the Pulse Train subtest does require subjects to detect changes in duration, and the 200-Hz SAM test involves rapid (i.e., short-duration) changes in intensity. Note that the duration of the intensity dips for the 200-Hz SAM test are comparable to threshold gap durations in the Gap Detection subtest, which has similar loadings on the Loudness-Duration and Amplitude Modulation factors. (A reviewer suggested that the 200-Hz SAM stimuli might have stronger pitch strength when deeply modulated, which might in turn be perceived as greater loudness, which seems a plausible argument.) The high cross loadings indicate that these tests require a second ability, almost to the same degree as the primary Loudness-Duration ability. The role of this ability in the Syllable Identification test ("you will mark *ooze*, please") is less obvious than in the other tests with high loadings on this factor, but the loadings also suggest that the use of duration and intensity cues may be especially important in discriminating among the alternatives when nonsense VCs are presented in this context (cafeteria noise).

2. Amplitude modulation factor

This factor includes the SAM noise discrimination tests with modulation rates of 8, 20, and 60 Hz, as well as the gap detection subtest. The modulation rates of subtests in this factor are considerably slower than the 200-Hz rate for the SAM test that loaded most strongly on the Loudness-Duration Factor. This is consistent with the change in perceptual quality as the modulation rate increases. At the slower rates, the change in amplitude has a time-varying or "fluttering" quality, while at the fastest rate, the modulation is heard as an overall "roughness" or "buzzing" quality. Inclusion of the gap detection test in this factor seems understandable as it is yet another measure of the ability to follow brief changes in level. As with the 200-Hz SAM test, the amplitude fluctuation is brief (roughly 1 to 5 ms), and both tests have relatively high (and similar) loadings on the Loudness-Duration and Amplitude Modulation Factors. It appears that both tests utilize the two underlying abilities (envelope following and duration/intensity discrimination) in roughly equal amounts.

3. Familiar sounds factor

In previous studies with the TBAC, this factor has been designated a speech factor (e.g., Surprenant and Watson, 2001). However, earlier studies did not include any familiar sounds other than speech. While the strongest loading for the environmental sounds test is on this factor, unlike the nonsense-word, word, and sentence tests, it has a relatively strong cross loading on another factor (Loudness-Duration). This suggests the existence of a familiar sound recognition ability that is not specific to speech. Implications of a single factor with high loadings for tests of both speech and other familiar sounds are discussed in a later section.

4. Pitch and time factor

The six subtests that load most strongly on this factor require two types of pattern processing: spectral (the pitch and ripple-noise discrimination tests) and spectral-temporal (temporal order for tones, embedded test tone, gap discrimination, and syllable sequence). The syllable sequence subtest might have been expected to have its highest loading on the familiar sounds factor. However, it is perceptually quite different from the other speech tests in that it consists of sequences of natural speech tokens (/fa/ /ta/ /ka/ /pa/) that are made difficult to discriminate by deleting multiple cycles of the fundamental pitch from the vowels. As the components become brief, they lose their similarity to natural speech, and the test becomes similar to the nonspeech temporal-order test in which temporal-order judgments are based on changes in holistic pattern properties rather than an identification of the order of the components. Because all of the spectral-temporal subtests in this factor consist of brief patterns that are made more difficult by shortening durations that change the overall sound of the patterns, it may be that these tests are performed by attending to holistic properties that affect the overall sound quality. This interpretation is consistent with earlier reports indicating that subjects can *detect* changes in the temporal order of brief tones based on holistic properties, even when they are unable to identify the order of the tones (Hirsh, 1959; Watson, 2004).

Although the four factors described above do not capture the wide range of auditory abilities included in Carroll's (1993) analysis (primarily because of the narrower range of auditory tests included), the variety of tests within this narrower domain provides considerable information about the nature of the underlying abilities. The types of tests and stimuli were similar to those included in two of Carroll's factors: (1) Spectral, Temporal, and Intensive Discrimination of Nonspeech Sounds and (2) Speech Recognition under Distortion. Three of the present factors can be seen as a finer differentiation of the first factor, suggesting different abilities underlying different types of temporal and spectral tests. The Familiar Sounds factor supports Karlin's (1942) finding of a separate factor for understanding speech under difficult listening conditions (Carroll's "Speech Recognition under Distortion"). The inclusion of an environmental sound test in the present study and its major loading on the factor that otherwise contains only speech tests suggests that this factor may be related to a more general ability.

5. General auditory and intellectual abilities

Before considering the nature of these four factors and the underlying abilities they may represent in more detail, the possible role of general auditory or intellectual abilities in the performance of these tests should be considered. The influence of a general ability, and the degree to which this ability may be auditory or intellectual, will be evaluated through comparison of different models of the latent variable structure that underlies the measured variables, using structural equation modeling.[5] The role of intellectual abilities will be evaluated by examining the correlations between SAT and latent variable scores.

The collection of models evaluated for this analysis is illustrated in Fig. 3. The simplest, model 1, has a single latent variable and represents the hypothesis that a single general ability underlies all measured variables. Model 2 is based on the principal components analysis shown in Fig. 2. This model includes four latent variables (corresponding to the four factors), each associated with a different set of measured variables. Each measured variable is associated with the latent variable that corresponds to the factor on which it had the highest loading in the principal components analysis. For simplicity, no cross loadings from the principal components analysis are represented in this model. The third model is a combination of the first two, in which there are four independent specific auditory factors and a general factor associated with all of the measured variables (and independent of the four specific factors). Model 4 is a version of the second model in which the four auditory factors are allowed to correlate with each other. (Model 1 can also be seen as a special case of model 4, in which all correlations among the factors are set to unity.) Finally, model 5 is the most general model in that it includes all of the components of the other four models.

A summary of the goodness of fit for all five models is given in Table IV. Models 1 and 2 are not particularly good fits to the data. Root-mean-square-error of approximation (RMSEA) values less than 0.05 and comparative-fit index (CFI) values of 0.95 or higher are generally taken as evidence of a good fit (see Bollen, 1989; Raykov and Marcoulides, 2000). Although a four-factor solution was considerably better than a single-factor solution in the principal components analysis, there is no advantage for the simplified four-factor model over a single-factor model in this context. This is largely due to the absence of cross loadings in this model and to a different standard in evaluating the goodness of fit. Also, recall that the four-factor principal components solution was far from a perfect fit, accounting for roughly half of the variance.

Models 3 and 4 both provide fairly close fits to the data. However, the model with a general factor provides a significantly better fit to the data than the correlated-factors model ($\chi^2=64$, $df=13$, $p<0.001$). This provides further support for the existence of four independent factors, but it also indicates the existence of a general factor accounting for additional variance. Model 5, which allows correlations among the four auditory factors, results in a small, but statistically significant ($\chi^2=19$, $df=6$, $p<0.01$), improvement in the fit measures. However, given the similarity of the fit measures (and the power of the significance test with such a large N), there is no compelling argument for the more complex and difficult-to-interpret model 5.

The standardized factor loadings for model 3 are shown in Fig. 4. All loadings for associations with the general factor are significant. Nearly all loadings for associations between measured variables and the four specific factors are also significant, the exceptions being those for SAM at 200 Hz, Gap Discrimination, Ripple Noise Discrimination, and Syllable Sequence. The loadings for the four specific auditory factors are generally consistent with those from the principal components analysis, supporting the interpretation given earlier.

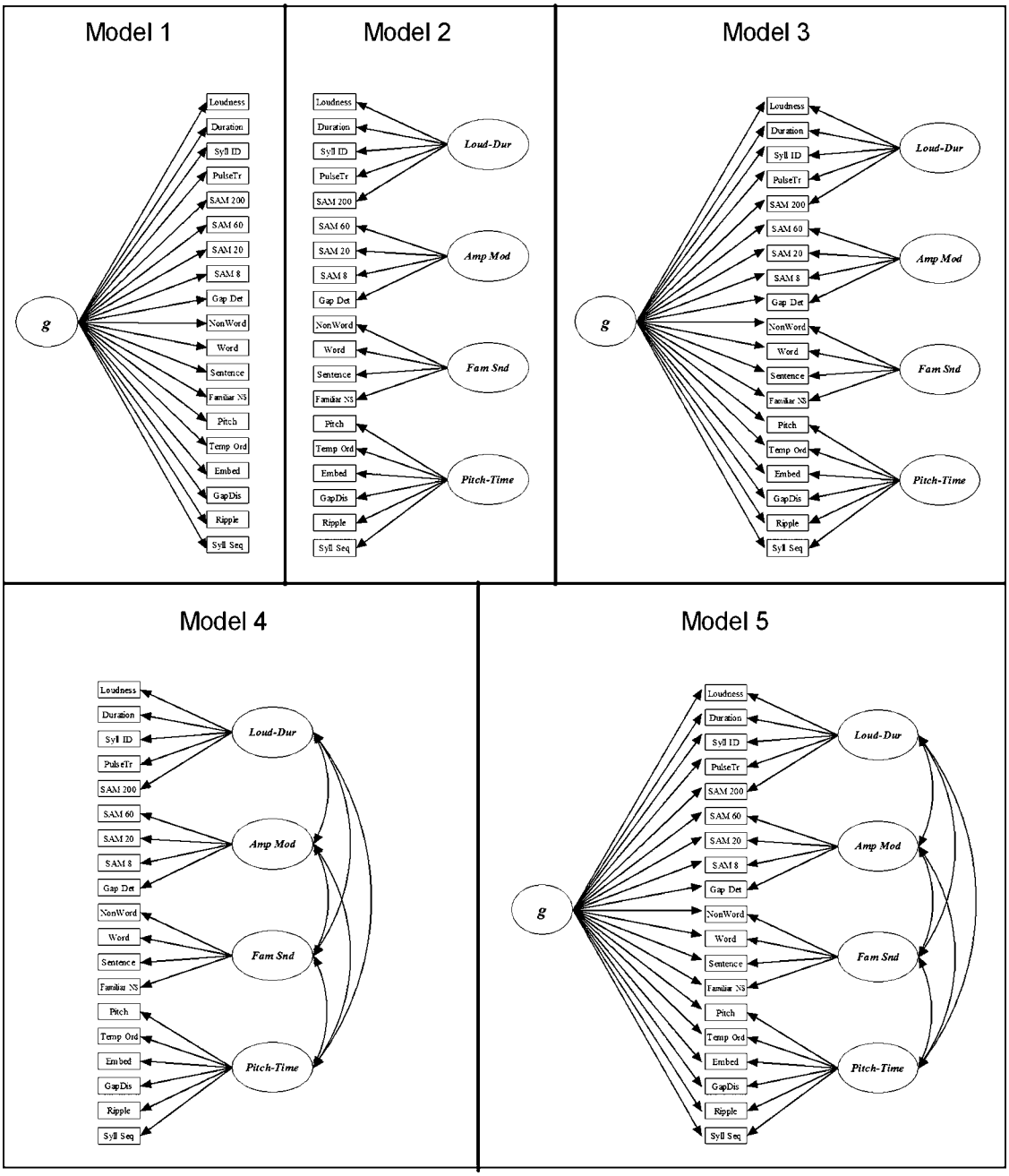


FIG. 3. Five models of auditory abilities. Model 1 represents a single general auditory ability accounting for performance on all tasks. Model 2 represents four independent auditory abilities associated with specific subsets of tasks (based on the principal components analysis) and no general ability. Model 3 includes a general ability and four independent abilities. Model 4 is a version of model 2 that allows correlations among the four auditory abilities. Model 5 is a version of model 3 that allows correlations among the four auditory abilities.

The loadings for the general factor are greater than those for the specific factors for 11 of the 19 subtests. This is most pronounced for the subtests of the Pitch and Time factor (with the exception of the Pitch subtest) and, to a lesser extent, for subtests of the Loudness-Duration factor (with the exception of the Loudness subtest). Loadings for the specific factors are generally greater than or equal to those for the general factor for subtests in the Amplitude Modulation factor and the Familiar Sounds factor, although the general factor plays a greater role in Gap Detection and Environmental Sound Identification.

Although all measured variables are based on tests designed to assess auditory abilities, it is important to determine the extent to which nonauditory intellectual abilities contribute to performance. The extent to which any of the latent variables in model 3 reflect intellectual abilities was

TABLE IV. Goodness-of-fit statistics for five models of auditory abilities.

Model	*df*	$\chi 2$	RMSEA	CFI
(1) One general factor	152	556	0.089	0.907
(2) Four auditory factors	152	614	0.095	0.887
(3) One general and four independent factors	133	189	0.035	0.983
(4) Four correlated factors	146	253	0.047	0.969
(5) One general and four correlated factors	127	170	0.032	0.986

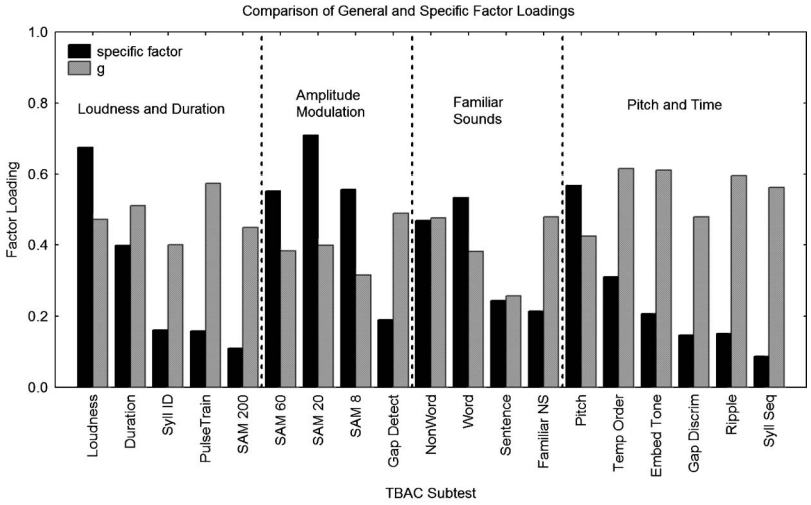


FIG. 4. Factor loadings for the latent variables in model 3. This shows the degree to which performance on each subtest is influenced by the general auditory ability (gray bars) and the specific auditory ability (black bars) with which it is associated in the model. Each section of the figure shows the loadings for a different specific factor.

evaluated by examining the correlations between the latent variable scores and SAT scores for each of the latent variables.

Math and verbal SAT scores were available for approximately 70% of the subjects (235 of 338 subjects).[6] The correlations between these scores and the latent variable scores from model 3 are shown in Table V. All correlations with latent variables are relatively low, indicating that variation in general intellectual abilities within this population accounts for an extremely small portion of the variance in performance. Some correlations are significant, however, with the highest occurring between SAT scores and the latent variable scores associated with the "general" ability. The data also indicate that verbal abilities play a minor role in familiar sound recognition and that both math and verbal abilities have a small influence on performance with the Pitch/Time tests. Overall, the magnitude of the correlations indicates that the abilities associated with these latent variables are closely associated with the ability to discriminate or identify different types of auditory stimuli, rather than with more general intellectual abilities associated with nonauditory or amodal aspects of the tests. However, the possible involvement of some other type of general nonauditory ability (e.g., temporal processing, attention, memory), independent of intellectual abilities but not specific to auditory stimuli, cannot be ruled out.

TABLE V. Correlations between latent variable scores and SAT scores (verbal and quantitative).

	Loud dur	Amp mod	Fam snd	Pitch time	*g*
SAT-V	−0.02	0.03	0.18[a]	0.14[b]	0.19[a]
SAT-M	−0.03	0.03	0.08	0.15[b]	0.22[c]

[a] $p < 0.05$.
[b] $p < 0.01$.
[c] $p < 0.001$.

IV. DISCUSSION

A. Familiar sound recognition ability

The data suggest the existence of an FSR ability, but additional evidence will be required to more precisely specify the nature of that ability. This is closely related to an ability identified by Carroll (1993), based largely on the work of Stankov and Horn (1980), which was labeled "Speech Perception Under Distortion." The factor loadings suggest that the perceptions of familiar speech and nonspeech sounds rely on the FSR ability to different degrees. However, there is clearly a common ability that partially accounts for the recognition of both types of familiar sounds under difficult listening conditions. This may be an ability to make use of limited information that does not fully specify the stimulus event. Such an ability might be based on one or more of the following:

1. Efficiency in the location or activation of stored information (memory) about familiar sounds. Rapid access to items in the lexicon and in an equivalent storage system for environmental sounds could facilitate the consideration of alternatives in a recognition task.
2. A problem-solving or guessing strategy that effectively fills in the missing pieces of an incomplete signal (i.e., an ability to identify the "whole" based on fragments). This might reflect a more effective use of stimulus knowledge to generate hypotheses and rule out alternatives with regard to the missing information. Evidence of such an ability in both the auditory and visual modalities was reported by Watson *et al.* (1996).
3. Effective dynamic focusing of attention on the most informative spectral-temporal locations within a familiar sound. The locations may be those where distinguishing information is present or where the background noise is least disruptive. Such attentional strategies would require

the use of information about the properties of the familiar sounds to be identified and of the background or interfering sounds.

These three skills are not mutually exclusive, and the FSR ability may reflect a combination of them. That is, efficient access to stored information about familiar sounds can facilitate guessing or stimulus-completion strategies or help allocate attention. It should also be noted that these abilities are not necessarily purely auditory. The memory retrieval, stimulus-completion, or attentional strategies that make up the FSR ability may facilitate the recognition of any familiar stimulus (or perhaps any familiar temporal pattern) regardless of its sensory modality. That FSR is independent of the other abilities identified in this study does not mean that spectral and temporal resolving power are unimportant for the recognition of speech and other familiar sounds. The findings indicate only that individual differences in spectral and temporal resolving power among young normal-hearing listeners do not account for individual differences in the ability to recognize speech or environmental sounds under difficult listening conditions. A more severe degradation of spectral acuity can certainly impair the ability to understand speech, as is evident in the effects of processing speech through a vocoder with only a few channels (Dudley, 1939; Shannon *et al.*, 1995; Friesen *et al.*, 2001). The identification of an FSR ability, independent of spectral-temporal resolving power does, however, strongly suggest that cognitive approaches to improving speech understanding in hearing-impaired individuals should be explored. Although no hearing-impaired individuals were tested in this study, a corresponding range of the FSR ability seems certain to exist in the hearing-impaired population and thus to account for differences in speech recognition abilities among listeners with similar hearing loss.

B. Spectral and temporal processing abilities

The remaining three specific factors identified in this study reflect a variety of spectral or temporal processing abilities. These three factors represent a further differentiation of a factor identified by Carroll (1993), which he described as encompassing spectral and temporal processing for nonspeech complex stimulus patterns. This division of Carroll's single factor into three is a consequence of a wider range of relevant tests than in any of the earlier investigations considered by Carroll. There was some evidence in the datasets analyzed by Carroll for a distinction between spectral abilities (e.g., frequency and timbre discrimination) and temporal abilities (e.g., intensity, duration, and rhythm discrimination), but the data did not provide clear support for the existence of these two independent factors.

Rather than revealing a distinction between spectral and temporal abilities, the present study suggests the presence of three distinct types of temporal processing abilities. The *Loudness-Duration* factor reflects the ability to detect changes in overall energy (and duration). This ability has an obvious element of temporal processing, suggestive of the energy detection model (the "leaky integrator" model) developed within signal detection theory (Green and Swets, 1966; Jeffress, 1967). The *Pitch-and-Time* factor suggests a combined spectral-temporal pattern processing ability that is closely associated with spectrum and timbre discrimination. A range of tests with both spectral (i.e., frequency or timbre) and temporal (i.e., duration or timing) aspects is included. This factor may largely reflect the ability to discriminate holistic changes in the sound quality of a spectral-temporal pattern, whether due to spectral changes, temporal changes, or both. Finally, the *Amplitude Modulation factor* (consisting of SAM detection at the slower rates and gap detection) suggests an envelope-following ability that is distinct from an energy detection ability.

Despite the relatively large number of auditory tests (19) employed in the present study, compared to earlier investigations, a still wider range appears necessary to achieve a satisfactory definition of each of the four auditory abilities. The current findings provide a basis for the selection of additional tests that utilize the hypothesized abilities to different degrees. For example, the inclusion of other modulation-based tasks (e.g., modulation masking, modulation detection interference, modulation rate discrimination) might help clarify the nature of the ability that underlies the Amplitude Modulation factor, perhaps clarifying the role of modulation filters (Dau *et al.*, 1997a, b; Langner *et al.*, 2002) in these abilities. The inclusion of a rhythm discrimination test that required the discrimination of relative timing (e.g., by using time-transposed rhythmic patterns) plus other temporal discrimination tasks with different types of temporal structure over a range of time frames might help clarify the relations among the different temporal abilities observed in the present study. The inclusion of tasks that require the *discrimination* of familiar sounds with different types of stimulus changes would help to clarify the nature of the FSR ability and its relation to other auditory abilities.

C. The possibility of an "auditory *g*"

The latent variable analysis revealed the existence of a *general ability* in addition to the four independent auditory abilities. This represents common variance among all auditory tests that was not accounted for by the four factors. The influence of the general factor on performance was significant for all tests and was greater than that of the specific factors for a majority of the tests (11 of 19). However, both specific and general factors had a significant influence on all but four of the tests. The influence of specific factors was somewhat greater than that of the general factor for tests associated with the Amplitude Modulation and Familiar Sounds factors, while the opposite was true for the Pitch and Time factor and the Loudness-Duration factor.

Although a small portion of this general ability is related to math and verbal abilities, as measured by SAT scores, this ability is largely independent of intellectual abilities and appears to reflect a general facility with auditory stimuli. While many cognitive abilities, such as working memory or controlled attention, have been shown to be correlated with SAT scores (Daneman and Hannon, 2001; Engle *et al.*, 1999), as has general intelligence (Frey and Detterman, 2004; Beaujean *et al.*, 2006), there may be other amodal abilities

that influence performance on auditory tests. For example, performance may be influenced by a general ability to process patterns in time or by the ability to detect subtle changes in complex multidimensional stimuli. It would be necessary to include a wider variety of nonauditory tests (e.g., visual, tactile, cognitive) to determine the extent to which the general ability identified in this study is specific to the auditory modality.

D. Summary and conclusions

The following conclusions are based on modeling of the data from 338 normal-hearing participants on 19 auditory discrimination and identification tests and associations with verbal and quantitative SAT scores.

(1) The best-fitting model of these data is one in which there is one general auditory ability and four independent specific abilities.
(2) The four specific auditory abilities are
    (a) loudness-duration discrimination
    (b) amplitude modulation detection
    (c) familiar sound recognition (speech and environmental sounds)
    (d) spectral and temporal pattern discrimination
(3) Intellectual abilities, as measured by SAT scores, had very weak associations with the general and specific abilities, accounting for less than 5% of the variance in each case.
(4) The independence of the four specific abilities suggests that individual differences in speech recognition under difficult listening conditions are not due to differences in spectral or temporal acuity, but to differences in a *familiar sound recognition (FSR) ability*.
(5) A more precise specification of the nature of the abilities identified in this study will require further individual-differences research.

Although no hearing-impaired listeners participated in this study, it is assumed that clinical populations vary along the same auditory-ability dimensions as the young normal-hearing listeners tested in the present study. Thus, differences in the ability to understand speech under difficult listening conditions among hearing-impaired listeners with similar hearing loss are likely due, at least in part, to differences in the FSR ability.

ACKNOWLEDGMENTS

This research was supported by Grant No. R01 DC00250 from the National Institute on Deafness and other Communicative Disorders. The authors would like to thank Patricia Knapp for her help in preparing the manuscript for publication. The associate editor and two reviewers also made significant contributions to this article. We are especially grateful for the constructive criticisms made by a third reviewer.

[1]Although SAT scores may not be ideal measures of general intellectual abilities, two recent studies (Frey and Detterman, 2004; Beaujean *et al.*, 2006) have shown that the relation between SAT scores and IQ measures is quite robust and clearly sufficient for the current purpose. These scores were available, while the cost of individually administered IQ tests would have been prohibitive.

[2]Because two sets of S/N values were used for the speech subtests (as described in the Method section), deciles were computed separately for each of the two groups of listeners and then combined to form one set of deciles. The data plotted in Fig. 1 for the Word subtest thus include 10 points (representing two sets of five S/N values) for each decile.

[3]At the most difficult stimulus values tested, performance becomes nearly independent of the stimuli. At these levels, some listeners will perform above chance and others below (for finite numbers of trials). When listeners are grouped by overall performance, as in the present case, those in the lowest deciles will tend to be those with below-chance responses, yielding the appearance of below-chance asymptotes for those deciles. This appearance is, of course, merely an artifact of this method of grouping the listeners.

[4]A different version of this figure, based on an earlier analysis, appears as part of an overview of individual differences research in Watson and Kidd (2002).

[5]Although structural equation modeling (SEM) has become popular in many other fields, there are only a handful of uses of this approach in articles published in this journal. In brief, SEM is a family of statistical concepts that include (and interrelate) multiple regression analysis, factor analysis, (causal) path analysis, and linear regression analysis (see Bentler, 1980, 1986; Bollen, 1989). Its use in the present project is to compare several alternative models that might account for the performance of the 338 subjects on 19 subtests in this project, in terms of a variety of hypothesized auditory and cognitive abilities. LISREL version 8.7 from Scientific Software International was used for all SEM analyses.

[6]Some subjects were not students, some did not take the SAT, and some records were not available at the Indiana University Office of Records for unknown reasons.

ANSI (**1989**). ANSI S3.6-1989: Specifications for Audiometers. (American National Standards Institute, New York).

Beaujean, A. A., Firmin, M. W., Knoop, A. J., Michonski, J. D., Berry, T. P., and Lowrie, R. E. (**2006**). "Validation of the Frey and Detterman (2004) IQ prediction equations using the Reynolds Intellectual Assessment Scales," Personality & Individual Differences **41**, 353–357.

Bentler, P. M. (**1980**). "Multivariate analysis with latent variables: Causal modeling," Annu. Rev. Psychol. **31**, 419–456.

Bentler, P. M. (**1986**). "Structural equation modeling and Psychometrika: An historical perspective on growth and achievements," Psychometrika **31**, 35–51.

Binet, A. (**1903**). *The Experimental Study of Intelligence* (Schleicher, Paris).

Bollen, K. A. (**1989**). *Structural Equations with Latent Variables* (Wiley, New York).

Bronkhorst, A. W., and Plomp, R. (**1992**). "Effect of multiple speechlike maskers on binaural speech recognition in normal and impaired hearing," J. Acoust. Soc. Am. **92**, 3132–3139.

Carroll, J. B. (**1993**). *Human Cognitive Abilities: A Survey of Factor-analytic Studies* (Cambridge University Press, New York).

Cattell, J. M. (**1885**). "On the time required for recognizing and naming letters and words, pictures and colors," in *James McKeen Cattell: Man of Science, Vol. 1: Psychological Research*, edited by A. T. Poffenberger (Science, Lancaster, PA, **1947**), pp. 13–25.

Christopherson, L. A., and Humes, L. E. (**1992**). "Some psychometric properties of the Test of Basic Auditory Capabilities (TBAC)," J. Speech Hear. Res. **35**, 929–935.

Daneman, M., and Hannon, B. (**2001**). "Using working memory theory to investigate the construct validity of multiple-choice reading comprehension tests such as the SAT," J. Exp. Psychol. Gen. **130**, 208–223.

Dau, T., Kollmeier, B., and Kohlrausch, A. (**1997a**). "Modeling auditory processing of amplitude modulation. I. Detection and masking with narrow-band carriers," J. Acoust. Soc. Am. **102**, 2892–2905.

Dau, T., Kollmeier, B., and Kohlrausch, A. (**1997b**). "Modeling auditory processing of amplitude modulation. II. Spectral and temporal integration," J. Acoust. Soc. Am. **102**, 2906–2919.

Davis, H., and Silverman, S. R. (**1970**). *Hearing and Deafness* (Holt, Rinehart & Winston, New York).

Drennan, W., and Watson, C. S. (**2001**). "Sources of variation in profile analysis. I. Individual differences and extended training," J. Acoust. Soc. Am. **110**, 2491–2503.

Dubno, J. R., and Levitt, H. (**1981**). "Predicting consonant confusions from acoustic analysis," J. Acoust. Soc. Am. **69**, 249–261.
Dudley, H. (**1939**). "Remaking speech," J. Acoust. Soc. Am. **11**, 169–177.
Elliott, D. N., Riach, W. D., Sheposh, J. P., and Trahiotis, C. (**1966**). "Discrimination performance of high school sophomores on a battery of auditory tests," Acta Oto-Laryngol., Suppl. **216**, 1–59.
Engle, R. W., Tuholski, S. W., Laughlin, J. E., and Conway, A. R. (**1999**). "Working memory, short-term memory, and general fluid intelligence: A latent-variable approach," J. Exp. Psychol. Gen. **128**, 309–331.
Feth, L. L., and O'Malley, H. (**1977**). "Two-tone auditory spectral resolution," J. Acoust. Soc. Am. **62**, 940–947.
French, N. R., and Steinberg, J. C. (**1947**). "Factors governing the intelligibility of speech sounds," J. Acoust. Soc. Am. **19**, 90–119.
Frey, M. C., and Detterman, D. K. (**2004**). "Scholastic assessment or g? The relationship between the scholastic assessment test and general cognitive ability," Psychol. Sci. **15**, 373–378.
Friesen, L. M., Shannon, R. V., Baskent, D., and Wang, X. (**2001**). "Speech recognition in noise as a function of the number of spectral channels: Comparison of acoustic hearing and cochlear implants," J. Acoust. Soc. Am. **110**, 1150–1167.
Galton, F. (**1869**). *Hereditary Genius: An Inquiry into its Laws and Consequences* (Macmillan, London).
Galton, F. (**1883**). *Inquiries into Human Faculty and its Development* (Macmillan, London).
Glasberg, B. R., and Moore, B. C. J. (**1989**). "Psychoacoustic abilities of subjects with unilateral and bilateral cochlear hearing impairments and their relationship to the ability to understand speech," Scand. Audiol. Suppl. **32**, 1–25.
Good, P. I. (**2006**). *Resampling Methods: A Practical Guide to Data Analysis*, 3rd ed. (Birkhauser, Boston).
Gorsuch, R. L. (**1983**). *Factor Analysis* (Erlbaum, Hillsdale, NJ).
Green, D. M. (**1971**). "Temporal auditory acuity," Psychol. Rev. **78**, 540–551.
Green, D. M., and Swets, J. A. (**1966**). *Signal Detection Theory* (Wiley, New York).
Gygi, B. (**2001**). "Factors in the identification of environmental sounds," Unpublished doctoral dissertation, Indiana University, Bloomington.
Gygi, B., Kidd, G. R., and Watson, C. S. (**2004**). "Spectral-temporal factors in the identification of environmental sounds," J. Acoust. Soc. Am. **115**, 1252–1265.
Gygi, B., Kidd, G. R., and Watson, C. S. (**2007**). "Similarity and categorization of environmental sounds," Percept. Psychophys. (in press).
Hanley, C. N. (**1956**). "Factorial analysis of speech perception," J. Speech Hear Disord. **21**, 76–87.
Hirsh, I. J. (**1959**). "Auditory perception of temporal order," J. Acoust. Soc. Am. **31**, 759–767.
Humes, L. E., and Riker, S. (**1992**). "Evaluation of two clinical versions of the articulation index," Ear Hear. **13**, 406–409.
Hunt, E. (**1978**). "Mechanics of verbal ability," Psychol. Rev. **85**, 109–130.
Jakobson, L. S., Cuddy, L. L., and Kilgour, A. R. (**2003**). "Time-tagging: A key to musicians' superior memory," Music Percept. **20**, 307–313.
Jeffress, L. A. (**1967**). "Stimulus-oriented approach to detection reexamined," J. Acoust. Soc. Am. **41**, 480–488.
Johnson, D. M., Watson, C. S., and Jensen, J. K. (**1987**). "Individual differences in auditory capabilities. I," J. Acoust. Soc. Am. **81**, 427–438.
Karlin, J. E. (**1942**). "A factorial study of auditory function," Psychometrika **7**, 251–279.
Langner, G., Albert, M., and Briede, T. (**2002**). "Temporal and spatial coding of periodicity information in the inferior colliculus of awake chinchilla (Chinchilla laniger)," Hear. Res. **168**, 110–130.
MacCorquodale, K., and Meehl, P. E. (**1948**). "On a distinction between hypothetical constructs and intervening variables," Psychol. Rev. **55**, 95–107.
Moore, B. C. J., and Glasberg, B. R. (**1986**). "The role of frequency selectivity in the perception of loudness, pitch and time," in *Frequency Selectivity in Hearing*, edited by B. C. J. Moore (Academic, London), pp. 251–308.
Nelson, P. B., Nittrouer, S., and Norton, S. J. (**1995**). "Say-stay identification and psychoacoustic performance of hearing-impaired listeners," J. Acoust. Soc. Am. **97**, 1830–1838.
Pavlovic, C. V. (**1984**). "Use of the articulation index for assessing residual auditory function in listeners with sensorineural hearing impairment," J. Acoust. Soc. Am. **75**, 1253–1258.
Raykov, T., and Marcoulides, G. A. (**2000**). *A First Course in Structural Equation Modeling* (Erlbaum, Mahwah, NJ).
Seashore, C. E., Lewis, D., and Saetveit, J. (**1939**, **1960**). *Seashore Measures of Musical Talents* (The Psychological Corporation, New York).
Shannon, R. V., Zeng, F-G., Kamath, V., Wygonski, J., and Ekelid, M. (**1995**). "Speech recognition with primarily temporal cues," Science **270**, 303–304.
Snell, K. B., Mapes, F. M., Hickman, E. D., and Frisina, D. R. (**2002**). "Word recognition in competing babble and the effects of age, temporal processing, and absolute sensitivity," J. Acoust. Soc. Am. **112**, 720–727.
Snow, R. E. and Lohman, D. F. (**1988**). "Implications of cognitive psychology for educational measurement," in *Educational Measurement*, 3rd ed., edited by R. L. Linn (American Council on Education & Macmillan, New York), pp. 263–331.
Spearman, C. (**1904**). "General intelligence objectively determined and measured," Am. J. Psychol. **15**, 201–293.
Stankov, L., and Horn, J. L. (**1980**). "Human abilities revealed through auditory tests," J. Educ. Psychol. **72**, 21–44.
Sternberg, R. J. (**1977**). *Intelligence, Information Processing, and Analogical Reasoning: The Componential Analysis of Human Abilities* (Erlbaum, Hillsdale, NJ).
Strouse, A., Ashmead, D. H., Ohde, R. N., and Grantham, D. W. (**1998**). "Temporal processing in the aging auditory system," J. Acoust. Soc. Am. **104**, 2385–2399.
Surprenant, A. M., and Watson, C. S. (**2001**). "Individual differences in the processing of speech and nonspeech sounds by normal-hearing listeners," J. Acoust. Soc. Am. **110**, 2085–2095.
Tallal, P., Miller, S., and Fitch, H. A. (**1993**). "Neurobiological basis of speech: A case for the preeminence of temporal processing," in *Temporal Information Processing in the Nervous System: Special Reference to Dyslexia and Dysphasia, Annals of the New York Academy of Sciences, Vol. 682* edited by P. Tallal, A. M. Galaburda, R. R. Llinas, and C. v. Euler (NYAS, New York), pp. 27–47.
Thurstone, L. L. (**1938**). "Primary mental abilities," Psychometric Monographs, No. 1.
Thurstone, L. L. (**1947**), *Multiple Factor Analysis* (University of Chicago Press, Chicago).
Tyler, R. S., Summerfield, Q., Wood, E. J., and Fernandes, M. A. (**1982**). "Psychoacoustic and phonetic temporal processing in normal and hearing impaired listeners," J. Acoust. Soc. Am. **72**, 740–752.
Viemeister, N. F. (**1979**). "Temporal modulation transfer functions based on modulation thresholds," J. Acoust. Soc. Am. **66**, 1364–1380.
Watson, B. U. and Miller, T. K. (**1993**). "Auditory perception, phonological processing and reading ability/disability," J. Speech Hear. Res. **36**, 850–863.
Watson, C. S. (**2004**). "Temporal acuity and the identification of temporal order: Related, but distinct, auditory abilities," Seminars in Hearing **25**, 219–227.
Watson, C. S., and Kidd, G. R. (**2002**). "On the lack of association between basic auditory abilities, speech processing, and other cognitive skills," Seminars in Hearing **23**, 83–93.
Watson, C. S., Kelly, W. J., and Wroton, H. W. (**1976**). "Factors in the discrimination of tonal patterns. II. Selective attention and learning under various levels of stimulus uncertainty," J. Acoust. Soc. Am. **60**, 1176–1186.
Watson, C. S., Qiu, W. W., Chamberlain, M., and Li, X. (**1996**). "Auditory and visual speech perception: Confirmation of a modality-independent source of individual differences in speech recognition," J. Acoust. Soc. Am. **100**, 1153–1162.
Watson, C. S., Jensen, J. K., Foyle, D. C., Leek, M. R., and Goldgar, D. E. (**1982a**). "Performance of 146 normal adult listeners on a battery of auditory discrimination tasks," J. Acoust. Soc. Am. **71**, S73.
Watson, C. S., Johnson, D. M., Lehman, J. R., Kelly, W. J., and Jensen, J. K. (**1982b**). "An auditory discrimination test battery," J. Acoust. Soc. Am. **71**, S73.
Wichmann, F. A., and Hill, N. J. (**2001a**). "The psychometric function: I. Fitting, sampling, and goodness of fit," Percept. Psychophys. **63**, 1293–1313.
Wichmann, F. A., and Hill, N. J. (**2001b**). "The psychometric function: II. Bootstrap-based confidence intervals and sampling," Percept. Psychophys. **63**, 1314–1329.
Yost, W. A., Hill, R., and Perez-Falcon, T. (**1978**). "Pitch and pitch discrimination of broadband signals with rippled power spectra," J. Acoust. Soc. Am. **64**, 1166–1173.

Theoretical feasibility of suppressing offensive sports chants by means of delayed feedback of sound

Sander J. van Wijngaarden[a] and Johan A. van Balken
TNO Defence, Security & Safety, P.O. Box 23, 3769 ZG Soesterberg, The Netherlands

(Received 3 July 2006; revised 24 April 2007; accepted 24 April 2007)

A novel approach for disrupting offensive chants at sporting events is proposed, based on attacking synchronization between individuals. Since timing is crucial for coordination between chanters, disruption of timing is expected to be effective against undesired chants. Delayed auditory feedback is known to disrupt timing in individual sound production. It may be expected to have similar effects on groups of chanters. To test this hypothesis, a controlled laboratory study was carried out. This showed that the timing of individuals joining in with sports chants can indeed be severely disrupted by also presenting an artificially delayed version of this chant (distracter). This effect is reduced as an individual is given more cues (direction, fidelity) to differentiate between original chant and distracter. However, informal field trials showed that it may be hard to exploit the perceptual effects discussed here for countering offending sports chants in a real-life setting, particularly due to feedback distortion at the required high levels. © *2007 Acoustical Society of America.*
[DOI: 10.1121/1.2740054]

PACS number(s): 43.66.Lj, 43.75.Rs [DOS] Pages: 436–445

I. INTRODUCTION

Chanting is a common phenomenon at sporting events. Many sports chants reinforce a positive atmosphere (Armstrong and Young, 1999), but others can be highly offensive. Attempts are often made to disrupt offensive chants, usually by means of masking sounds such as loud music (McRobb Calder, 2002). Earlier, we have proposed an alternative approach, which has the potential of being effective at lower sound levels (Van Wijngaarden and Van Leeuwen, 2006; Van Wijngaarden and Van Balken, 2006). This approach is based on attacking synchronization between individuals. Since timing is crucial for coordination between musicians as well as in speech (e.g., Jungers *et al.*, 2002), disruption of timing is expected to be effective against chants. A mechanism known to affect an individual's timing of speech and music is delayed auditory feedback.

When a person's speech is fed back to this person at a certain delay, for instance by means of headphones, then effects on speech production are noticed (Lee, 1950; Yates, 1963). Depending mostly on the delay time, delayed auditory feedback may induce a reduction of the speaking rate, skipping or repeating of words and syllables, or even total confusion. Speech production is influenced by one's perception of one's own speech; introducing time delays disrupts this natural feedback process. Similarly, timing of music performances is affected by delayed feedback (Pfordresher and Palmer, 2002).

Delayed auditory feedback is primarily known to lay audiences (if at all) as a party trick sometimes used in television game shows, but is also used in a more serious professional context, such as in stuttering therapy (Kuniszyk-Jözkowiak, 1996). Under certain delayed feedback conditions, many stutterers are able to speak more fluently. Applications are also found in clinical audiology (Tye-Marry, 1992). Delayed auditory feedback can be used to determine whether children who have been given cochlear implants are able to decode information that is becoming available through these implants; they will then show the "normal" response to delayed auditory feedback, such as reduction of the speaking rate through prolonging of speech sounds.

In groups, delayed feedback may have slightly more complex effects. Presumably, each individual in a group producing synchronous speech or music does not only use his own speech as a feedback signal, but the joint speech signal from the entire group; otherwise accurate synchronization would be impossible to achieve. Also, the combined contributions of other group members will normally have a much higher sound level than just one's own speech, with the possible exception of small groups of just a few people. In other words, timing performance of each member of the group affects all other members of the group. For a group to maintain an overall acceptable synchrony and rhythm, a certain process of social self-organization is needed. Introducing delays between different contributors to the overall sound may have surprising consequences. Pairs of musicians, clapping rhythms together while listening to a delayed version of each other's sound, were found to show a deceleration of the clapping rate that seems consistent with classical effects of auditory feedback (Chafe *et al.*, 2004). However, at relatively short time delays, a surprising increase in clapping rate was observed.

Rhythmic applause by audiences of music performances has been studied using models taken from statistical physics (Néda *et al.*, 2000). After a well-received musical performance, a period of a few seconds of incoherent applause is often followed by a phase of accurately synchronized clapping. The transition between these two phases can be understood by modeling the audience as a set of globally coupled

[a]Electronic mail: sander.vanwijngaarden@tno.nl

 0001-4966/2007/122(1)/436/10/$23.00

oscillators. The condition for synchronization is that the dispersion (statistical spread in timing) between group members must be smaller than a certain critical value. Hence, one would expect disruption of coordination between chanters (due to delayed feedback channels) to be potentially effective in undermining synchronization, by increasing dispersion between group members. This can only be true if timing of individual chanters is, in fact, affected by delayed feedback from other chanters. The purpose of the current study is to test this hypothesis.

Whether or not a suppression technique based on delayed sound feedback can be effective in practice will also depend on other (largely nonacoustic) factors. Practice has proven sports chants to be remarkably resistant against suppression by external influences. Sociologists view sports chants as a collective expression of social-cultural identity (Armstrong and Young, 1999). Comparisons to primitive rituals to prove one's masculinity, or even ritual warfare, have even been drawn (Bromberger, 1993). In all, it seems safe to say that very potent measures are needed to successfully suppress a chant once sports fans have set their hearts on singing it. Our approach, which is essentially based on subconscious, involuntary perceptual mechanisms, is appealing since chanters may be expected to find it hard to devise counterstrategies.

Two laboratory experiments were carried out in order to determine the conditions under which delayed feedback affects individuals in a group (Secs. II and III). These experiments place participants, one by one, in a group of simulated fellow chanters (based on digitally manipulated recordings). The participant tries to chant along with this group. Since the group is simulated, any arbitrary delay can be added to the feedback signal presented to the participant.

Although these laboratory experiments do not recreate the dynamic, interactive group process observed in real life, it does give insight into the behavior of individuals in reaction to delayed contributions from other chanters. It allows us to test the formulated main hypothesis: Can chanting individuals be disrupted by delayed feedback of other chanters' voices? Additionally, two less formal field trials took place, which are briefly presented in Sec. IV.

II. EXPERIMENT 1: EFFECT OF DELAY AND DISTRACTER SIGNAL LEVEL

A. Goal of the experiment

The primary goal of experiment 1 was to test the hypothesis that timing of individual chanters can be adversely affected by presenting them with delayed versions of the contributions by other chanters. Also, optimum time delays and signal-to-noise ratios were investigated. Subjects were presented with "target" and "distracter" chanters, and instructed to synchronize with the target chanters, while ignoring the distracter chanters. Sound level and delay of the distracters was varied.

Additionally, the question whether or not individual chanters experience distracter chanters to be disturbing was investigated.

B. Method

1. Participants

Eight paid volunteers, all male students with an interest in sports, took part in this experiment. They varied in age between 19 and 33 years (mean 23 years). They completed a questionnaire specifically inquiring about hearing impairments, speech impairments, and linguistic disorders. None of the subjects reported abnormal speech or hearing.

2. Stimuli

Participants were instructed to sing along with four different chants (see the Appendix). These chants were derived from relatively well-known Dutch football chants, and would generally not be considered offensive. The target signal consisted of high-quality recordings of a (synchronized) group of seven chanters (six males, one female). The distracter signal was another recording of the same seven chanters; because of between-session differences, subtle differences between target and distracter signals existed. The target signal was presented with zero delay, and at a fixed signal level of 80 dB(A) at the position of the participant's head.

Depending on the test conditions, the distracter signal was presented with time delays ranging from 0 to 900 ms, and signal levels ranging from 6 dB below the target signal level to 15 dB above the target signal level. In other words, the target-to-distracter ratio ranged from +6 to −15 dB.

3. Apparatus and procedure

Participants were placed in an anechoic room, in a standing position. The target signal as well as the distracter signal was presented from a loudspeaker at a distance of 1.40 m straight in front of the participant, at the approximate height of the participant's ears. Subjects wore a lightweight boom-microphone which was used to record their own chanting during the experiment.

Before starting the actual experiment, subjects were given several practice runs to get acquainted with equipment and procedures. The experiment was divided into 3 min runs of singing the same chant, structured in the following way:

(1) 12 s of singing along without disruption (no delay of the distracter signal).
(2) 45 s during which either the signal level or the delay of the chanter was increased (depending of the current test condition).
(3) This whole pattern (12 plus 45 s) was repeated three times, leading to a total of nearly 3 min.

Two different types of test conditions were presented: runs during which the delay was constant and the distracter level increased, versus runs during which the delay was increased at a fixed level of the distracter. The conditions included in experiment 1 are given in Table I.

It appears likely that the effect of delayed feedback will not only depend on the delay and distracter sound level at the time of measuring the effect, but also on the way these parameters varied in the moments before measurement. This should become apparent by investigating differences be-

TABLE I. Conditions included in experiment 1. Each condition was repeated three times per subject. The presentation order was counterbalanced across participants and repetitions, in order to cancel out any effects of presentation order on the mean results.

	Chant	Delay (ms)	Target-to-distracter ratio (dB)	Number of conditions
Baseline	1–4	0	-	4
Fixed delay	1–4	200	−15 to 3	4
Fixed delay	1–4	300	−15 to 3	4
Fixed delay	1–4	400	−15 to 3	4
Fixed delay	1–4	600	−15 to 3	4
Fixed delay	1–4	800	−15 to 3	4
Fixed target-to-distracter ratio	1–4	0–900	6	4
Fixed target-to-distracter ratio	1–4	0–900	0	4
Fixed target-to-distracter ratio	1–4	0–900	−6	4
Fixed target-to-distracter ratio	1–4	0–900	−12	4

tween the two types of conditions (fixed delay and increasing distracter level, versus increasing delay and fixed distracter level).

Subjects were instructed, and frequently reminded, that their objective was to maintain synchrony with the original target chant. A small financial bonus was promised (and given) to subjects who were able to synchronize well, in order to provide additional motivation to maintain synchrony.

4. Data analysis

Two different measures were extracted from each condition: a subjective rating of the degree of disruption, and an objective synchronization measurement. The subjective rating, on a 1–5 scale, was given by each participant upon completion of each 3 min run. The subjective rating reflects the overall judgment of the entire run, some parts of which may have been more disturbing than others.

The synchronization measure was derived from the boom microphone which recorded the participant's chanting. This measure was derived using a correlation technique based on the amplitude envelope of the signal, which was calculated from the recorded audio and resampled to 200 Hz. This was done for each tested condition, but also for a normalization condition in which the participant was allowed to synchronize as accurately as possible, without any disruption. The two envelopes (test condition and optimally synchronized condition) were compared by calculating a correlation coefficient for every segment of 2 s. A correlation coefficient of 1 indicates perfect synchronization of the segment, a value close to 0 indicates no (time) relation between the segments. Hence, each minute of recorded signal resulted in 30 measures reflecting the degree of synchronization at that particular moment. These measures were used for further analysis in three ways.

(1) *Calculating the proportion of accurately timed segments.* In order to decide whether a 2 s segment was accurately timed (synchronous) or not, a threshold on the measured correlation coefficients of 0.70 was used. This threshold was chosen somewhat arbitrarily; it corresponds to the 75th percentile when correlating different instances of the same, undisturbed but somewhat sloppily timed chant. Other choices for the threshold between 0.50 and 0.90 were verified to produce similar results.

(2) *Calculating the required time shift to restore synchronization.* Another measure of synchronization is how much adjustment (time shift) the chant contribution by a participant would need in order to be optimally synchronous with the target. If a participant's timing is ahead or behind relative to the target chant, then the correlation with the target will be reduced; it can be calculated how much time shift is needed to reach an optimal correlation.

(3) *Direct statistical analysis on mean correlation coefficients, after normalization.* Methods (1) and (2) are useful to gain qualitative insight into the data, but do not easily allow statistical analysis. For this reason, the correlation coefficients measured from the participant's chant contributions were also considered more directly, i.e., by calculating average correlation coefficient values and associated standard errors for each delay and target-to-distracter ratio. Before averaging, an additional normalization was applied to the correlation coefficients. Even if a chant is repeated several times under perfect conditions, the correlation between successive instances of the chant will always be less than optimal, i.e., smaller than 1. The reason for this is that perfect timing is nearly impossible to achieve. However, since difficulty of timing may vary between chants, it is desirable to account for this natural variability. This is done by dividing all correlation coefficients by the average correlation between two undisrupted instances of the same chant.

C. Results

The results from experiment 1 are presented in three sections. First, we show the general pattern of responses within typical test runs. This section serves to create a qualitative impression of the way that chants become disrupted by the distracter signal. The second section contains an analysis of objective synchronization data for each condition, averaged across subjects. This gives an impression of the magnitude of the disruptive effects, as a function of delay and distracter sound level. The third section presents subjective disruption ratings, showing the degree to which participants are consciously aware of (and suffering from) disruption attempts.

1. Synchronization within a test run

Figure 1 shows the average proportion of appropriately timed (synchronous) 2 s segments, as a function of time, averaged over a series of typical test runs. In this example, the target-to-distracter ratio was fixed at −12 dB while the delay increased gradually from 0 to approximately 900 ms.

During the first 12 s the target and distracter were still synchronous. This leads to a relatively high proportion of accurately timed segments, although some initialization ef-

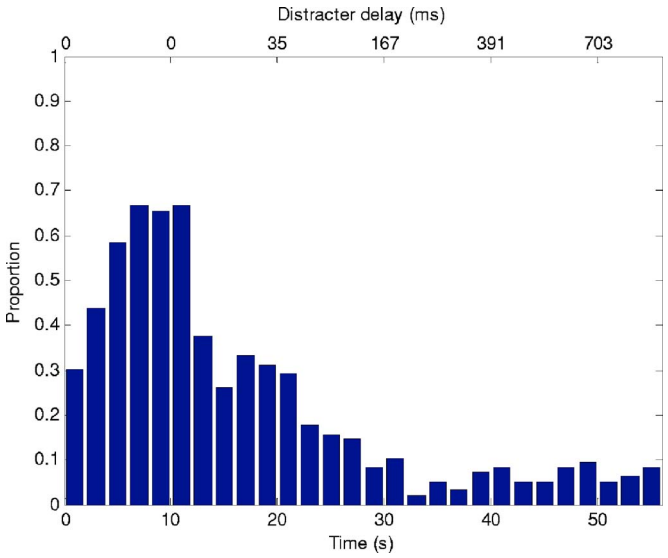


FIG. 1. (Color online) Proportion of accurately timed chant segments, as a function of position within a test run (N=96; 8 participants, 4 chants, 3 repetitions). The target-to-distracter ratio for this condition was −12 dB. The delay of the distracter was 0 ms during the first 12 s, then increased up to approximately 900 ms by slowing the distracter down at a rate of 0.5%/s.

fects appear to be present in the first 6 s. Once an (initially small) delay of the distracter is introduced, the proportion of accurately timed segments decreases rapidly.

Figure 2 is based on the same condition as Fig. 1, but this time the required time shift to restore optimal synchronization is shown.

Figure 2 shows some erratic data points at any moment during the test run, including during the undisrupted beginning of the run. However, as a delay is introduced to the distracter chant, the data points appear to fall apart into two distinct groups: one following the target chant (around a delay of 0 ms), and one following the distracter chant, at an

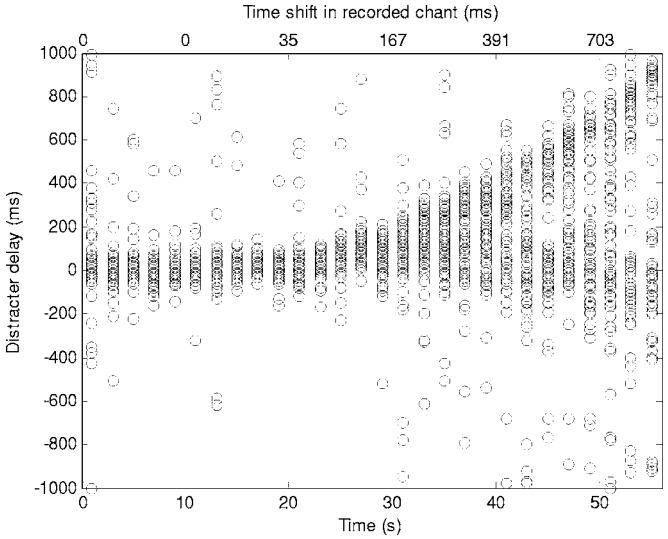


FIG. 2. Time shift (delay) of the participant's recorded chant for which timing correlates optimally to the target chant. Data from each participant, chant, and repetition are shown as separate data points (N=96; same condition as Fig. 1).

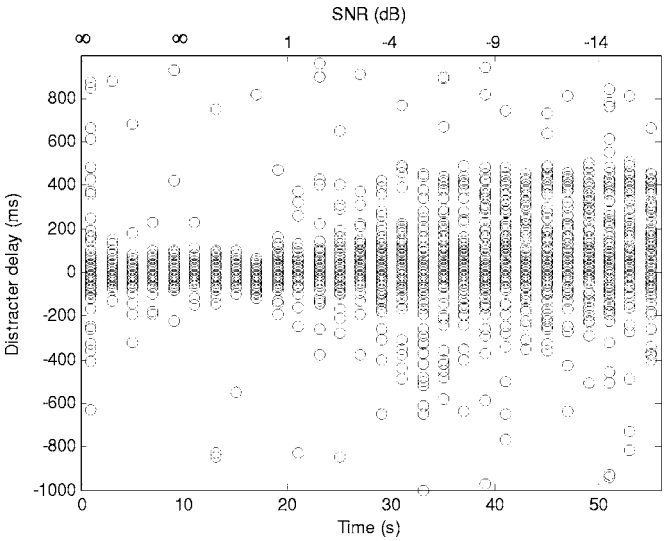


FIG. 3. Time shift (delay) of the participant's recorded chant for which timing correlates optimally to the target chant, at a fixed distracter delay of 400 ms and an increasing distracter loudness (N=96; target-to-distracter ratio decreasing at 0.5 dB/s).

increasing delay. These are not specifically between-participant differences; all participants showed both types of behavior.

A different pattern is observed for conditions where the distracter delay was fixed and the distracter level was increased gradually; see Fig. 3 for a typical example. In this case, the distracter was introduced after 12 s at a delay of 400 ms. The target-to-distracter ratio was brought from infinity to +3 dB in about 1 s, and decreased gradually (0.5 dB/s) after that. In such cases (as the example shown in Fig. 3), data points are not only clustered around the target and distracter delay (at, respectively, 0 and 400 ms), but distributed more or less evenly between those values. This indicates a greater degree of uncertainty on the part of the participant; instead of strictly following the timing of either the distracter or the target, participants appear to be switching back and forth. Hence, gradually introducing delay leads to more predictable behavior on the part of the participants; if they are at all affected by the distracter, they are likely to attempt to synchronize to it.

2. Synchronization results for each tested condition

Mean normalized correlation coefficients of all conditions with fixed target-to-distracter ratios are given in Fig. 4.

Figure 4 shows that the correlation increases with target-to-distracter ratio. In other words, distracters are more disruptive if their level is higher. There appears to be a ceiling effect toward a target-to-distracter ratio of 0 and +6 dB; in order for the distracter to be disruptive, its level must be higher than the target sound level. Also, the correlation decreases (the distracter is more disruptive) as the delay increases.

Figure 5 gives similar results as shown in Fig. 4, but this time for conditions with a fixed distracter delay. Individual delayed auditory feedback, as used in classic experiments mentioned in Sec. I has an optimum effect at a delay time around 300 ms. There does not appear to be such a clear

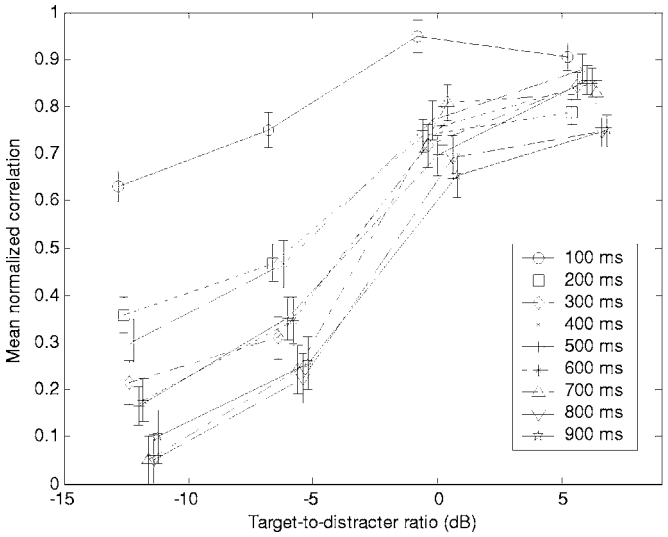


FIG. 4. Mean normalized correlation coefficient (correlation during test run divided by correlation without distracter) at fixed target-to-distracter ratios of −12, −6, 0, and 6 dB, averaged around instantaneous delays as indicated in the legend, ±50 ms. The error bars indicate the standard error (N=96).

optimum for the delay time in this case, although a local minimum in correlation may be present around 300–400 ms. In general however, and especially at lower target-to-distracter ratios, longer delays tend to be more disruptive.

A four-way analysis of variance (ANOVA; delay, target-to-distracter ratio, chant, repetition) was carried out on all data of Fig. 4 as well as Fig. 5. Both ANOVAs lead to the same results: significant ($p<0.001$) main effects for delay and target-to-distracter ratio, and also significant main effects ($p<0.05$) for chant and repetition. The significant effect for the type of chant can be understood by considering the difference in timing difficulty between chants. Different repetitions lead to different results because of learning and initialization effects: The first repetition shows consistently lower correlations, which can be understood by assuming that it takes subjects some time to get adjusted to the rhythm of the

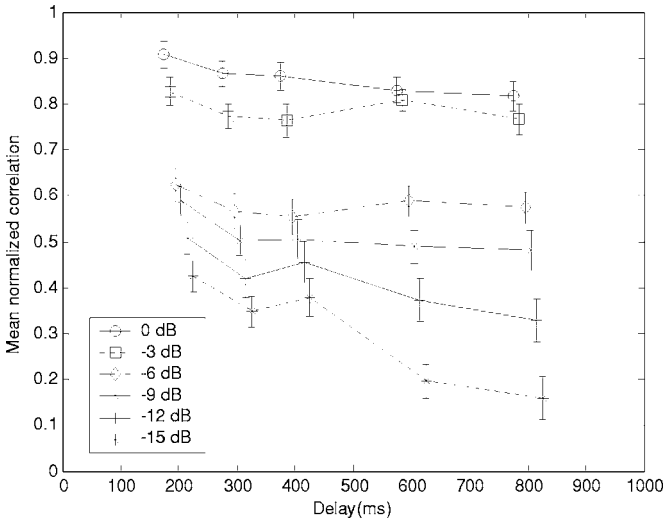


FIG. 5. Mean normalized correlation coefficient at fixed distracter delays of 200, 300, 400, 600, and 800 ms, averaged around instantaneous target-to-distracter ratios as indicated in the legend, ±1 dB. The error bars indicate the standard error (N=96).

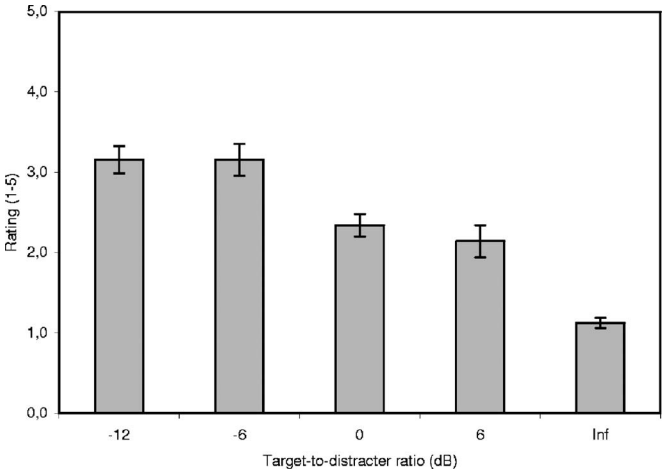


FIG. 6. Mean subjective degree of disruption experienced by participants, as a function of target-to-distracter ratio, rated on a scale from 1 (no disruption) to 5 (absolute disruption). Results are averaged across eight participants and four chants (N=32). The error bars indicate the standard error.

chant. Because the experiment was counterbalanced across repetitions and participants, the influence of this learning effect on the mean results for the other variables will be averaged out.

There is also a significant interaction ($p<0.001$) between delay and target-to-distracter ratio. This can be understood by inspecting Figs. 4 and 5: longer delays are more effective, but this effect also appears to be enhanced by distracter level. Another significant interaction ($p<0.05$) occurs between delay and chant. Chants that are more difficult to time are more profoundly affected by longer delays.

When comparing Figs. 4 and 5, it is noticed that any combination of delay and target-to-distracter ratio results in higher correlations in Fig. 5 than in Fig. 4. It appears that a gradually increasing delay is more detrimental to proper synchronization than increasing the distracter level at a fixed delay.

3. Subjective disruption ratings

Average disruption ratings are given in Figs. 6 (fixed target-to-distracter ratio) and 7 (fixed delay).

The subjective experience of disruption follows the same patterns as the objective synchronization data: Longer delays lead to more disruption, without a clear optimum delay time. Upon introduction of delayed feedback, the disruption rating increases from 1 (indicating no disruption) to an average of around 3, indicating moderate disruption. It should be kept in mind that this is an average rating, pooled across various (easy and hard) conditions. The hardest conditions received individual ratings of 4 or 5 by all participants.

D. Discussion

The results from experiment 1 show that timing and synchronization of an individual chanter is indeed adversely affected by a delayed contribution. Individual chanters also experience disruption subjectively. Longer delays tend to be more effective in disturbing synchronization; Delayed contri-

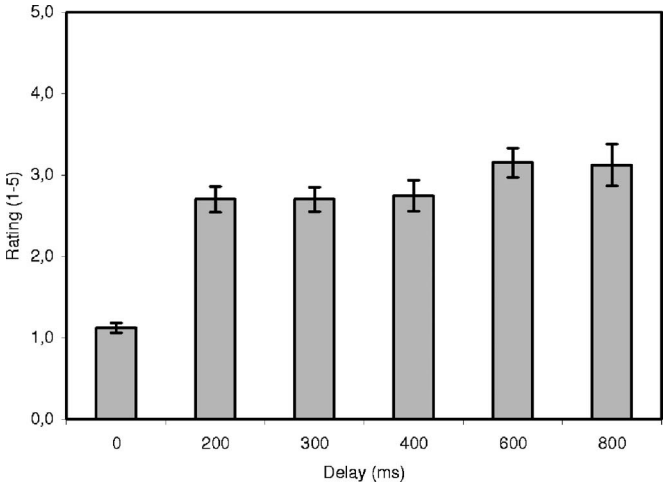


FIG. 7. Mean subjective degree of disruption experienced by participants, as a function of distracter delay, rated on a scale from 1 (no disruption) to 5 (absolute disruption). Results are averaged across eight participants and four chants ($N=32$). The error bars indicate the standard error.

butions must also have a sufficiently high level (higher than the original chant) to be disruptive. Timing is affected differently by a distracter that gradually increases in sound level, than by a distracter of which the delay is gradually increased. In the latter case, participants are likely to synchronize to the distracter signal instead of the original chant. If the delay is not introduced gradually (but instead the level increases gradually) the timing is more erratic and hard to predict. Overall, the effect of a gradually increasing delay is greater than the effect of a fixed delay at a gradually increasing distracter level.

III. EXPERIMENT 2: INFLUENCE OF DIRECTION AND SIGNAL FIDELITY

A. Goals of the experiment

In experiment 1 subjects were given very few cues to distinguish between target and distracter signals. In fact, since the direction and signal quality were the same, the main cue was the fact that the distracter was introduced later, after adaptation to the original chant.

If feedback signals are to be used for chant suppression in real stadiums, then there will always be additional cues for distinguishing between the original chant (sung "live" in a stadium) and artificial feedback signals. The effects of two such cues, direction of the sound sources and signal fidelity, were investigated in experiment 2.

B. Method

1. Participants

A group of eight male participants, none of whom also participated in experiment 1, took part. The participant group was matched to experiment 1 in terms of background and age. Scores on a set of nine reference conditions (see Sec. III B 3) that were identical between experiments 1 and 2 showed no significant between-group differences in the participant groups.

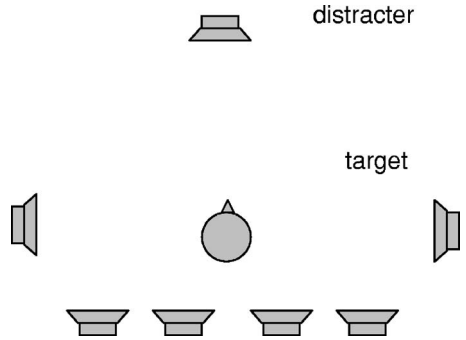


FIG. 8. Schematic representation of loudspeaker locations for presentation of the target chant in experiment 2.

2. Stimuli

The same recordings were used as in experiment 1. Additionally, a "low fidelity" version of the recorded chants was created by digitally introducing bandwidth limiting and nonlinear distortion components similar to the characteristics of a specific horn-type loudspeaker (Philips LBC 3403) sometimes used in open-air public address systems. Calculated in octave bands, the frequency transfer is nearly flat in the 500, 1000, and 2000 Hz bands, and approximately 8 dB lower in the 4000 Hz band. Other octave bands do not contribute to the useful sound (i.e., produce no sound at all, or only distortion components).

3. Apparatus and procedure

In experiment 1, the target chant as well as the distracter chant were always presented using a loudspeaker in front of the participant. In experiment 2 (described in this section), the distracter chant was still produced by that same loudspeaker, straight in front of the participant. However, the way that the target chant was presented was differed between conditions. A group of seven loudspeakers was used to create various versions of the target chant (see Fig. 8).

The loudspeakers configuration of Fig. 8 was used in three different ways: (1) the front speaker was used for target as well as distracter (replicating experiment 1); (2) the loudspeaker off to the right-hand side of the participant was used to present the target chant; and (3) a different recording of a single chanter was presented over each of the seven loudspeakers, surrounding the participant with voices of simulated fellow chanters.

Condition (3) comes closest to what one might expect in a real-life setting. Condition (1) replicates experiment 1, albeit for just a subset of the conditions tests in experiment 1. The overlap between experiments 1 and 2 consists of a group of nine "reference conditions," which allow for comparison between the participant groups. Four conditions featured a fixed target-to-distracter ratio (6, 0, −6, and −12 dB) at a varying delay. For five conditions, the delay was fixed (200, 300, 400, 600, and 800 ms), and the target-to-distracter ratio varied. In all cases, only chant 1 was used. Post-hoc tests did not reveal statistical differences between the participant groups in these nine reference conditions.

The conditions included in experiment 2 are given in Table II.

TABLE II. Conditions included in experiment 2. Each condition was repeated three times per subject. The presentation order was counterbalanced across participants and repetitions, in order to cancel out any effects of presentation order on the mean results.

	Chant	Delay (ms)	Target-to-distracter ratio (dB)	Target loudspeaker	Distracter fidelity	Number of conditions
Baseline	1	0	-	Front	High	1
Reference conditions Fixed Delay	1	200, 300, 400, 600, 800	−12 to 3	Front	High	5
Reference conditions Fixed target-to-distracter ratio	1	0–900	6, 0, −6, −12	Front	High	4
Low fidelity Fixed delay	1	200, 300, 400, 600, 800	−12 to 3	Front	Low	5
Low fidelity Fixed target-to-distracter ratio	1	0–900	6, 0, −6, −12	Front	Low	4
Side loudspeaker Fixed delay	1	200, 300, 400, 600, 800	−15 to 3	Side (90°)	High	5
Side loudspeaker Fixed target-to-distracter ratio	1	0–900	6, 0, −6, −12	Side (90°)	High	4
All loudspeakers Fixed delay	1	200, 300, 400, 600, 800	−15 to 3	All	High	5
All loudspeakers Fixed target-to-distracter ratio	1	0–900	6, 0, −6, −12	All	High	4

C. Results

1. Effects of signal quality

Correlation results for the different signal quality conditions (low fidelity and high fidelity) are given in Fig. 9. Results were subjected to ANOVA analysis, similar to the results of experiment 1.

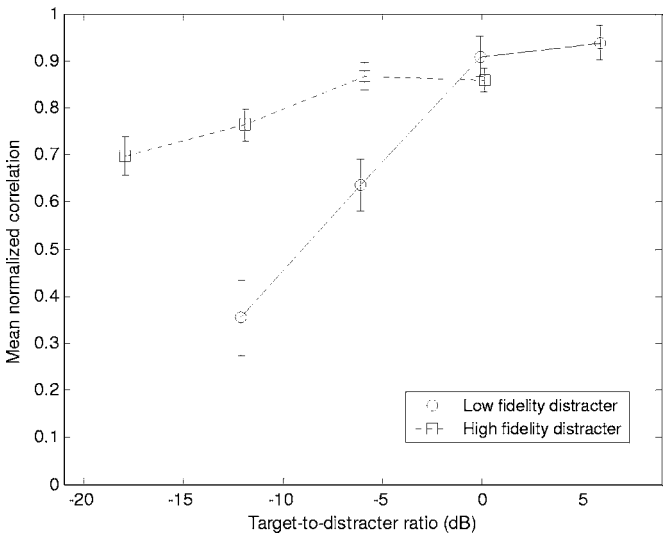


FIG. 9. Mean normalized correlation coefficient at fixed target-to-distracter ratios of −18, −12, −6, 0, and 6 dB, averaged across instantaneous delays between 200 and 800 ms. The error bars indicate the standard error (N =24).

Figure 9 shows that the distracter signal is much less disruptive if it can be distinguished from the target signal by differences in signal quality, although there is a significant effect of target-to-distracter ratio on synchronization ($p < 0.05$) even in the low-fidelity condition. The distorted signal leads to a correlation value at −18 dB that is similar to the correlation for the undegraded signal at 6 dB. In other words, the degraded distracter has to have 12 dB higher level to be equally effective. In reality, the difference may be smaller; ambient stadium acoustics, as well as interactions between speech signals from many chanters, will reduce the perceived quality of all chant signals compared to the "hifi" signal used in the laboratory experiment. The difference between target and distracter in the laboratory experiment is artificially large. However, the fact that signal quality cues may play an important role is clearly shown by this experiment.

2. Effects of sound direction

Conditions at various (fixed) target-to-distracter ratios were again considered, now with the target chant originating from different sources (in front, to the side, and surrounding the participant). Results are given in Fig. 10.

Again, synchronization is significantly affected by raising the distracter signal level. Clearly, directional cues derived from the target signal are helpful in distinguishing between target and distracter. The "group" condition with seven loudspeakers comes closest to the situation that one would

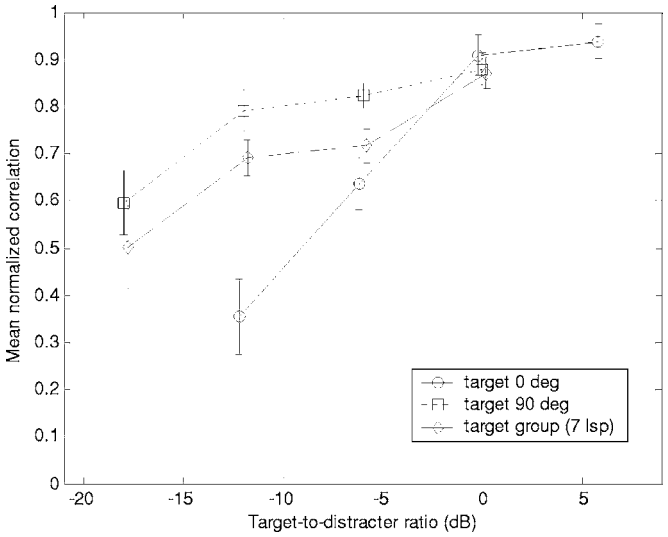


FIG. 10. Mean normalized correlation coefficient at fixed target-to-distracter ratios of −18, −12, −6, 0, and 6 dB, averaged across instantaneous delays between 200 and 800 ms. The error bars indicate the standard error (N = 24).

find in a true stadium setting, simulating a row of chanters behind the participant, as well as chanters to the side and in front. This condition allows for a less effective disruption by a delayed distracter than the conditions tested in experiment 1.

D. Discussion

Whereas experiment 1 showed that delayed feedback can be an effective disruptor of synchronization, experiment 2 makes it clear that real-life limitations, such as directional restraints and limits on signal quality, have a degrading effect on the effectiveness of the proposed technique of undermining sports chants. The extent to which the effect would be reduced in a real stadium setting cannot be determined from this experiment; the stylized laboratory experiments are too artificial compared to the situation in a stadium. Also, it is unclear how various cues differentiating between target and distracter would combine into an overall effect. However, the fact that directional cues and signal quality cues will play a noticeable role is adequately demonstrated.

IV. FIELD TRIALS

A. Need for field experiments

The above-described laboratory experiments have shown that chant timing may be disrupted by playing back delayed versions of a chant, and that the effectiveness of this approach depends on (at least) delay time, signal level, signal quality, and sound direction. However, the question whether a system based on delayed feedback can be effective against real-life chants, in real stadiums using large-scale sound systems, remains unanswered. Next to perceptual and (electro-) acoustic considerations, psychological aspects (such as adaptive group behavior) will determine whether this is the case. Only large-scale field testing will provide a definitive answer.

Translation of the feedback approach into a real system will involve an electro-acoustic setup, involving directional microphones (for picking up the original chant signal from the crowd) and directional loudspeakers. The sound cannot be fed back to the exact microphone position, or feedback distortion (howling, whistling) will occur. The simplest way to tackle this complication is to use multiple feedback loops, and to point each loudspeaker away from its own microphone. Each loudspeaker targets a separate section of the crowd, using the source signal from another section. If this works as planned, then each section may be presented with a different version of the distracter chant. However, since all of these feedback loops are placed in a common acoustic space, feedback distortion is still an issue; this is partly due to surface reflections and partly due to technical limitations to the directivity of microphones and loudspeakers. In practice, this means that the feedback signal cannot have a higher level than a certain maximum, which may or may not be high enough for the distracter chant to be sufficiently disruptive.

B. Results from preliminary field trials

While full-scale validation in field experiments was beyond of the scope of the study reported here, two preliminary field trials were organized to gain insight into complications that arise in practice. Primarily, these trials aimed to answer the question whether simple, commonly used sound systems can be used to produce a sufficiently disruptive feedback signal in the environment of a real stadium. Limits imposed on the distracter level due to feedback distortion were expected to be a major issue.

The first trial used a group of 350 participants, divided into three sections (associated with three feedback loops). Sound from the first section was presented to the second (adjacent) section, sound from the second section to the third, and sound from the third section to the first section. Various delay times (including fluctuating delays) were tested, as well as various loudspeaker setups. A single Bose 802b loudspeaker was used for each feedback loop; hence one loudspeaker was used for each section. Various microphone types (ranging from omnidirectional to highly directional) were tested. The participants (students, averaging 17 years of age) were given freedom to select chants at will.

The second trial was similar in setup, but with a group of only 115 participants (students, average age 19), divided into only two sections. In this case, two loudspeakers were used for each section (one in front of the crowd, and one behind the crowd). Also, a greater degree of structure was imposed on the trial (fixed chant lyrics, synchronized start of the chant, each participant standing exactly at a fixed position).

Both trials led to the same conclusions. It was found that the timing of subjects could be affected to such a degree that the intelligibility of the chant as a whole was subjectively somewhat reduced, according to the participants' own interview responses. However, participants could not be confused into discontinuing their chanting, and could not be split up into distinct groups sticking to different timing standards. A probe microphone, placed at ear-height in the middle of the

crowd, recorded A-weighted sound pressure levels ranging between 96 and 103 dB during chanting. At best, the level of the feedback signal in the middle of the crowd was high enough to achieve a target-to-distracter ratio of around 0 dB; higher levels resulted in such a degree of feedback distortion that this itself becomes a strong cue to differentiate between target and distracter sound.

Depending mostly on the loudspeaker configuration, up to 32% of participants in the 115 participant trial reported to experience disruption (indicated, according to instructions given beforehand, by hand-raising upon finishing the chant). These were consistently the subjects closest to the loudspeakers.

C. Conclusions from the preliminary field trials

The simple setup tested in the preliminary field trials will not be sufficiently effective to suppress chants in practice. The maximum levels of feedback sound that can be achieved are simply not high enough. Based on the laboratory results, the sound level of the feedback level should exceed the level of the original chant. This cannot be achieved with the type of sound system (in particular, the type of loudspeaker) used in the field trials. A more complex system, featuring highly directional loudspeakers and so-called electronic feedback destroyers, may result in higher maximum levels of the distracter signal. Whether this would allow for a more effective disruption of unwanted chants remains to be proven.

V. GENERAL DISCUSSION AND CONCLUSIONS

We showed that the timing of individuals joining in with sports chants can be severely disrupted by also presenting an artificially delayed version of this chant (called distracter): sometimes individuals synchronize to the distracter, but in other cases their timing is poorly correlated with the original chant as well as the distracter. If the distracter sound level is sufficiently high, and the delay is long enough, synchronization is affected. In addition to this (objective) effect, subjects also subjectively find the chants significantly harder to sustain. However, these effects are considerably reduced if the individual is given more cues (direction, fidelity) to differentiate between original chant and distracter.

Delayed auditory feedback is classically found to be optimally disruptive at delay times around 300 ms, and less effective at longer delays. Although our data suggest a local optimum around 300–400 ms in some cases, increasing the delay beyond 400 ms makes the feedback signal even more disruptive. In the case described here, the production-perception feedback loop is not actually closed within the individual, since the confusing, delayed signal stems from an external source. The difference in optimum delay (or rather, the lack thereof in the investigated range of delays) can be seen as an indication of differences in processing during individual sound production versus sound production in a group setting.

Although the laboratory experiments did yield sufficient proof of the perceptual principles exploited by the proposed chant disruption approach, the system tested in a set of preliminary field trials was not sufficiently effective. At least one reason for this could be identified: The maximum sound level of the distracter signal that could be achieved without unacceptable feedback distortion (i.e., feedback distortion becoming an additional, strong cue to differentiate between target and distracter) was not high enough. More advanced sound systems are likely to allow higher levels of feedback sound. Whether this will be enough to effectively suppress chants will also depend on other factors that have not yet been studied, such as (mostly nonacoustic) interactions between individuals in groups, and their effectiveness in devising—yet unidentified—counter-strategies.

APPENDIX: LYRICS OF FOOTBALL CHANTS USED IN EXPERIMENTS 1 AND 2 (IN DUTCH)

Chant 1

Hij is een onbenul
Hij is een onbenul
Hij is een onbenul
Hij is een onbenul

Chant 2

Alles of niets olé olé
Alles of niets olé olé
alles of niets
alles of niets
Alles of niets olé olé

Chant 3

Come on PSV
(repeated continuously)

Chant 4

Heerenveen maak die goal
Maak die goal
Maak die goaloaloal

Armstrong, G., and Young, M. (**1999**). "Fanatical football chants: Creating and controlling the carnival." Sport in Society **2**(3), 173–211.

Bromberger, C. (**1993**). "Fireworks and the ass," in *The Passion and Fashion: Football Fandom in the New Europe*, edited by S. Redhead (Aldershot).

Chafe, C., Gurevich, M., Leslie, G., and Tyan, S. (**2004**). "Effect of time delay on ensemble accuracy," *Proceedings of the International Symposium on Musical Acoustics (ISMA2004)*, Nara, Japan.

Jungers, M. K., Palmer, C., and Speer, S. R. (**2002**). "Time after time: The coordinating influence of tempo in music and speech," Cognitive Processing **1-2**, 21–35.

Kuniszyk-Jözkowiak W. (**1996**). "A comparison of speech envelopes of stutterers and non-stutterers," J. Acoust. Soc. Am. **100**(2), 1105–1110.

Lee, B. S. (**1950**). "Effects of delayed speech feedback," J. Acoust. Soc. Am. **22**(6), 824–826.

McRobb Calder, R. (**2002**). "Method of crowd control," UK Patent application GB 2 368 958 A.

Néda, Z., Ravasz, A., Brechet, Y., Vicsek, T., and Barabási, A-L. (**2000**). "Self-organizing processes: The sound of many hands clapping," Nature (London), **403**, 849–850.

Pfordresher, P. Q., and Palmer, C. (**2002**). "Effects of delayed auditory feedback on timing of music performance," Psychol. Res. **16**, 71–79.

Tye-Marray, N. (**1992**). "Young cochlear implant users' response to delayed auditory feedback," J. Acoust. Soc. Am. **91**(6), 3483–3486.
Van Wijngaarden, S. J., and Van Leeuwen, D. A. (**2006**). "Method and means for counteracting undesired sound utterances of a group," International Patent Application PCT/NL2005/000668.
Van Wijngaarden, S. J., and Van Balken, J. S. (**2006**). "Suppression of sports chants through delayed feedback of sound," J. Acoust. Soc. Am. **119**, 3332.
Yates, A. J. (**1963**). "Delayed auditory feedback," Psychol. Bull. **60**(3), 213–232.

Gap duration discrimination for frequency-asymmetric gap markers: Psychophysical and electrophysiological findings

John H. Grose,[a] Joseph W. Hall III, and Emily Buss
Department of Otolaryngology-Head & Neck Surgery, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599-7070

(Received 3 November 2006; revised 2 April 2007; accepted 3 April 2007)

This study investigated gap duration discrimination (GDD) for frequency-asymmetric gap markers, where one marker was a two-tone complex consisting of a primary tone and a secondary tone, and the other marker was the primary tone alone. Three experiments were undertaken to examine the order effect wherein performance is better when the two-tone marker is the leading marker than when it is the trailing marker. Experiment 1 demonstrated that GDD for frequency-asymmetric markers is intermediate between the boundaries of within-frequency-channel versus across-frequency-channel processing. Experiment 2 compared psychophysical performance with auditory brainstem responses (ABRs) elicited by the same stimuli. Whereas GDD thresholds were elevated for a complex trailing marker relative to a within-frequency-channel baseline, ABRs elicited by the complex marker were more robust. Experiment 3 tested the hypothesis that poor GDD performance with frequency-asymmetric markers is due to some form of nonenergetic, or informational, masking. The results did not support a role for informational masking conferred by synthetic listening; however, informational masking conferred by the occurrence of novel spectral events provided a parsimonious account. One possible interpretation is that the capacity to accurately encode a gap is undermined by the occurrence of novel spectral events that engage limited attentional resources. © *2007 Acoustical Society of America.* [DOI: 10.1121/1.2735106]

PACS number(s): 43.66.Mk, 43.66.Lj [AJO] Pages: 446–457

I. INTRODUCTION

The detection and discrimination of temporal gaps between two discrete stimuli is dependent in part upon the spectral similarity of the two stimuli. Sensitivity to gaps is most acute when the leading and trailing markers are spectrally similar, and declines when the spectral profiles of the markers diverge (e.g., Phillips *et al.*, 1997; Formby *et al.*, 1998a; Formby *et al.*, 1998b; Grose *et al.*, 2001; Phillips and Hall, 2002). In general, performance is poor when there is no spectral overlap between the markers, but is also elevated in cases where the differing bandwidths of the two markers allow the spectral content of the wider bandwidth marker to subsume the spectral content of the narrower bandwidth marker. Gap markers in this latter configuration, characterized by dissimilar but overlapping frequency content, are referred to as frequency-asymmetric markers. The study by Formby *et al.* (1998b) found an intriguing effect with frequency-asymmetric markers. When the leading marker was a two-tone complex consisting of a primary and secondary tone, and the trailing marker was the single primary tone alone, gap detection performance was significantly better than in the reverse configuration where the leading marker was the single primary tone and the trailing marker was the two-tone complex. This order effect was not subtle: When the two-tone marker was the trailing marker, thresholds were approximately five times greater than when the two-tone marker was the leading marker. Moreover, when the two-tone marker was the leading marker, thresholds were similar to the baseline where both leading and trailing markers were the same single pure tone. This order effect is not restricted to acoustic hearing. van Wieringen and Wouters (1999) tested configurations of electrode channel asymmetries in cochlear implant listeners. When the leading marker was carried on a single electrode channel and the trailing marker was carried simultaneously on two or three electrode channels, gap thresholds were significantly elevated compared to the reverse configuration where the leading marker was carried on multiple electrode channels and the trailing marker was carried on a single electrode channel.

The reason for this order effect is unclear, and forms the motivation for this investigation. Formby *et al.* (1998b) pointed out that the pattern of results is counterintuitive to a "cueing" explanation. That is, a single-tone leading marker might be expected to cue the listener as to which frequency channel to attend to—or "hear out"—in the trailing marker in order to achieve an optimal within-frequency-channel gap detection strategy (i.e., ignoring the remote frequency in the trailing two-tone marker). This would predict better performance in the configuration with the single-tone leading marker than with the two-tone leading marker where such a strategy is not possible, yet the converse is observed. van Wieringen and Wouters (1999) conjectured that the multichannel trailing marker might partially mask the single-channel leading marker, making this configuration more difficult for gap detection. However, they noted that the "backward masking" might be cognitive and not physiological in nature. For example, the timbral change brought about by the abrupt onset of additional channels in the trailing

[a]Author to whom correspondence should be addressed. Electronic mail: jhg@med.unc.edu

 0001-4966/2007/122(1)/446/12/$23.00

marker might interfere in some way with the memory trace of the leading marker. (It is not immediately clear why this notion would not apply equally to the converse "forward masking" configuration.) Oxenham (2000) also offered a suggestion for the order effect that focused on the role of onsets in gap detection. He raised the possibility that the simultaneous onset of the secondary tone in the trailing marker might stimulate neural units that would otherwise respond to the onset of the primary tone, thus diminishing the available within-channel information concerning the presence of the gap.

The notion that the presence of the secondary tone in the two-tone trailing marker in some way obscures information regarding the onset of the primary tone, and that this accounts for the decline in performance with the complex trailing marker, has an implicit premise. Specifically, it assumes that gap detection, even in the optimal within-channel configuration, is determined by the onset of the trailing marker. This premise appears reasonable given the plethora of onset-sensitive neurons in the auditory system coupled with the perceptual dominance of stimulus onsets (Heil, 1997; Fishbach *et al.*, 2001; Phillips *et al.*, 2002). However, the association of gap detection with onset detection (of the trailing marker) is not firmly established. On the one hand, both physiological (Barsz *et al.*, 1998) and behavioral (Ison *et al.*, 2002) work in mice has demonstrated that sensitivity to gaps is related to the abruptness of stimulus onset at the termination of the gap. On the other hand, work in humans has not found a benefit of abrupt onsets and offsets in gap detection (Allen *et al.*, 2002). In that study, gaps were introduced into noisebursts using different ramp shapes, but with the gap duration held constant at the half-amplitude points of the markers (constant "effective duration"). Results indicated that gap detectability was not related to the abruptness of the gap boundaries, with the caveat that this interpretation is somewhat dependent on the specification of "effective duration." Nevertheless, in that temporal gaps can be viewed as a special case of decrements in ongoing sounds, the absence of an onset effect is supported by Plack *et al.* (2006) who found that sensitivity to increments and decrements near threshold was better modeled by the *magnitude* of change in level rather than by the *rate* of change. That is, for a given duration of increment or decrement, the shape of the transition ramp had little effect—in particular, abrupt rise times did not result in better detection. It is important to note that these studies have employed frequency-symmetric stimuli; i.e., stimuli on either side of the gap or transition have been spectrally similar. In the case of frequency-asymmetric markers examined here, a discontinuity is always perceptible—even in the absence of an incremented gap—due to the presence of the remote secondary tone.

The purpose of this investigation was to determine the reliability of the order effect for frequency-asymmetric markers, and to test possible underlying mechanisms. The particular emphasis was on determining what factors account for the poor performance found with the complex trailing marker. Three experiments were undertaken. The preliminary experiment was designed to establish the order effect for a particular pairing of primary and secondary tones to be used throughout the investigation; this experiment had the specific focus of determining the magnitude of the order effect relative to the boundary conditions of within-frequency-channel versus across-frequency-channel processing. The second experiment was designed to test for onset effects by comparing psychophysical results with auditory brainstem responses (ABRs) evoked by the same stimuli. The third experiment was designed to determine whether some form of nonenergetic masking influences across-frequency gap discrimination.

II. EXPERIMENT 1. GAP DURATION DISCRIMINATION (GDD) FOR FREQUENCY-ASYMMETRIC MARKERS

The purpose of this experiment was to test the order effect using inharmonically related tone pairs lacking abrupt onsets. In addition to frequency-asymmetric markers, with dissimilar but overlapping frequency content, boundary conditions were also tested that employed frequency-symmetric markers (within-frequency-channel) and markers with no frequency overlap (across-frequency-channel).

A. Methods

1. Subjects

Six observers aged 19–39 years (mean=27 years) with normal hearing participated (observers, 1–6). All had audiometric thresholds $\leq$20 dB HL across the octave frequencies 250–8000 Hz (ANSI, 1996). All were given extensive training in the task by having each observer perform the entire experiment once for practice, including the collection of multiple replications per condition as stipulated below. The experiment was then repeated with retention of data.

2. Stimuli

The stimuli comprising the gap markers consisted of either a single primary tone or a two-tone complex (primary tone plus secondary tone). The frequency of the primary tone was 2 285 Hz and the frequency of the secondary tone was 4 000 Hz. This selection of frequencies was chosen because 2 285 Hz is greater than half an octave below 4 000 Hz, and gap detection is thought to become an across-frequency process beyond this interval (e.g., Formby *et al.*, 1996). Moreover, the 7/4 ratio between the two frequencies is essentially inharmonic. The nominal duration of each marker was 300 ms, including a 5 ms cosine-squared rise/fall ramp to reduce spectral splatter. The actual duration of each marker varied by ±20% of the nominal duration on a presentation-by-presentation basis in order to render the cue of overall stimulus duration (two markers plus gap) less reliable.

The gap was defined as the interval between the zero-voltage points separating the offset of the first marker and the onset of the second marker. The gap duration of the standard stimulus was always 0 ms. Note that because the 5 ms rise/fall ramps resulted in a perceptible temporal discontinuity even in the 0 ms standard gap configuration, the task was not one of gap detection but of gap duration discrimination (GDD). The stimuli were digitally generated at a rate of 20 kHz (TDT AP2), output through a DAC (TDT PD1), low-

pass filtered (Kemo VBF 10M, 8 kHz), attenuated (TDT PA4), and delivered to the left phone of a Sennheiser 580 headset at a pressure spectrum level of 60 dB/Hz.[1] In order to further diminish any confounding effects of spectral splatter, a continuous Gaussian background noise (0–8 kHz) was always present at a pressure spectrum level of 10 dB/Hz. An additional Gaussian noise low-passed at 750 Hz was added at 20 dB/Hz to provide further masking in the unlikely event that the complex marker generated an audible 2fl-f2 distortion product at 570 Hz.

3. Procedure

A three-alternative, forced-choice (3AFC) procedure was employed which incorporated a three-down, one-up adaptive rule that tracked the 79.4% correct point. Following three consecutive correct responses, the imposed increment in gap duration was decreased by a factor of 1.2; following one incorrect response, the imposed increment in gap duration was increased by the same factor. A threshold estimation run was terminated after ten reversals in gap duration, and the geometric mean of the gap duration at the final six reversal points was taken as the threshold estimation for that run. Four conditions were examined: (1) isofrequency GDD, where both gap markers consisted of the primary tone alone; (2) across-frequency GDD, where the leading marker consisted of the primary tone alone and the trailing marker consisted of the secondary tone alone; (3) frequency-asymmetric GDD, where the leading marker was the two-tone complex and the trailing marker was the single primary tone; and (4) frequency-asymmetric GDD, where the single primary tone was the leading marker and the two-tone complex was the trailing marker. For each observer, the conditions were blocked, in that all threshold estimates for a given condition were completed before the next condition was initiated, but the order of conditions was random across the observers. At least three replications, but occasionally four replications if time permitted, were collected per condition. Each observer's GDD threshold for a condition was taken as the geometric mean of all replications for that condition.

B. Results and discussion

The results are shown in Fig. 1 where the individual data (open symbols) cluster around the group means (solid symbols ±1 standard deviation). As expected, thresholds are lowest for the isofrequency condition. Thresholds for the across-frequency condition appear poorest. Thresholds for the frequency-asymmetric configurations are intermediate but, for five of the six listeners, are lower when the leading marker is the two-tone complex than when it is the single tone. In order to assess this pattern of results, the data were submitted to a repeated-measures analysis of variance (ANOVA). In this, and in all subsequent analyses, the analysis was performed on the log transforms of the data. The analysis indicated a significant effect of condition ($F_{3,15}=46.48$; $p<0.001$), and preplanned means comparisons for the three relevant comparisons confirmed the significance of the data pattern. That is, for the two frequency-asymmetric configurations, thresholds were significantly higher when the

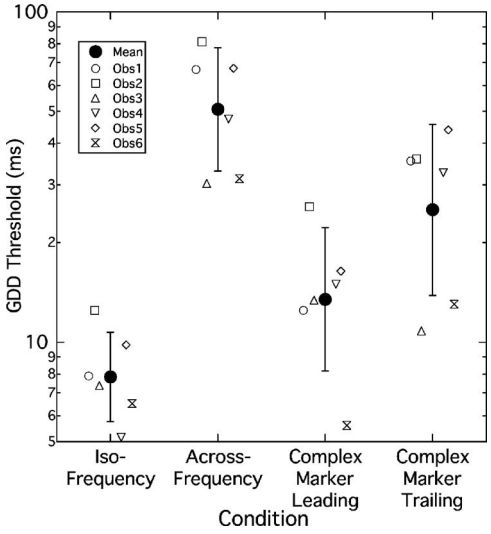


FIG. 1. GDD thresholds for individual observers (open symbols) and group mean (solid symbol) for the 4 conditions of experiment 1. Error bars are ±1 standard deviation.

leading marker was the single tone than when it was the two-tone complex ($t_5=3.183$; $p=0.024$). Nevertheless, thresholds for the configuration with the single-tone leading marker were not as poor as for the across-frequency configuration ($t_5=6.611$; $p=0.001$). Finally, thresholds for the two-tone leading marker configuration were significantly elevated relative to the isofrequency tonal baseline ($t_5=3.238$; $p=0.023$). Although these data support an order effect for frequency-asymmetric markers, they are less striking than those of Formby *et al.* (1998b) in two respects. First, thresholds for the two-tone trailing marker configuration were, on average, about twice those of the two-tone leading marker, whereas Formby *et al.* observed about a fivefold change. Second, this experiment found that thresholds for the two-tone leading marker were elevated relative to the isofrequency tonal baseline, whereas Formby *et al.* found a similar performance across these two configurations (i.e., no elevation for the two-tone leading marker). It is likely that these discrepancies are due to procedural differences across the two studies, particularly in terms of stimulus parameters. These differences include (1) the frequency range of the markers [2 000–3 100 Hz in Formby *et al.*; 2 285–4 000 Hz here], (2) the rise/fall times of the gap boundaries ["instantaneous" in Formby *et al.*; 5 ms here], and (3) the extent of marker duration randomization [100–900 ms in Formby *et al.*; 240–360 ms here]. Besides procedural differences across studies, it is also possible that individual differences in GDD performance for frequency-asymmetric markers may have played a role.

The results of this experiment confirm the order effect in that GDD for frequency-asymmetric markers is poorer for most listeners when the complex tone is the trailing marker than when it is the leading marker. The elevation in threshold for the two-tone trailing marker occurs even when the onset of the trailing marker transitions over a 5 ms interval. In order to examine the role of trailing marker onset further, a second experiment was undertaken that incorporated an electrophysiological test sensitive to stimulus onsets. At issue is

the deterioration in performance that occurs when a secondary tone is added to the trailing marker.

III. EXPERIMENT 2. ABR MEASURES OF FREQUENCY-ASYMMETRIC GDD

As noted earlier, it has been suggested that the order effect for frequency-asymmetric markers might be a reflection of the dominant role of stimulus onsets (in contrast to offsets) in auditory processing, such that the onset of the secondary tone in the trailing complex marker obscures the onset of the primary tone at some level of neural processing (Oxenham, 2000). There are several ways in which the primary tone onset might be obscured by the presence of the secondary tone, and two possible schemes will be considered here. Both are based on the assumption that the leading primary tone marker forward masks the trailing primary tone, but has essentially no effect on the trailing secondary tone. Within this context, it is further assumed that the cue for gap duration when both markers consist of the primary tone alone (within-frequency-channel) is linked to the degree to which the response to the trailing marker has recovered from forward masking; i.e., the onset response strength elicited by the trailing primary tone constitutes the gap duration cue. This suggestion is in line with traditional models of gap detection that are based on recovery from forward masking (e.g., Plomp, 1964). In the two schemes considered below, the response strength associated with the trailing primary tone onset is directly affected by the onset of the secondary tone. Both schemes also incorporate a central "onset detector" that is broadly tuned to onset information originating from different frequency regions.

The first possible mechanism for onset masking of the primary tone is a form of inhibition. Take, for example, the model of neural circuitry within the ventral cochlear nucleus proposed by Pressnitzer *et al.* (2001). In this model, the output from a particular cochlear location (here, the frequency channel associated with the primary tone) feeds two cell types in the cochlear nucleus. One, a narrow-band cell, is excited only by the primary tone frequency. The second cell type has a wide-band receptive field and responds not only to the primary tone but also to the secondary tone. The response of this wide-band cell is a fast-acting inhibition of the narrow-band cell response. The level of inhibitory input to the narrow-band cell responding to the primary tone is therefore greater when the secondary tone is present than when it is not. The interest here is not in testing this particular model; the model is described simply to illustrate that neural circuitry along these lines could conceivably provide a mechanism of onset masking in the context of GDD for frequency-asymmetric markers. Suppose that a mechanism of this sort resulted in mutual inhibitions between the two tones of the complex marker, where the magnitude of inhibition exerted by each tone on the other was proportional to their relative response strengths. For a short gap, where the primary tone response strength is diminished by forward masking, the inhibition exerted by the secondary tone onset onto the primary tone onset is predominant and, correspondingly, the primary tone exerts minimal inhibition of the secondary tone. For a long gap, where forward masking effects have mitigated, the mutual inhibition of the response onsets by the two tones is more equivalent. From the point of view of a central broadly tuned onset detector, two outcomes are conceivable. First, the onset information derived from the primary tone channel is diminished in the presence of the secondary tone (i.e., onset masking has occurred). Second, the registration of onset across frequency may not necessarily be more robust with increases in gap duration in the two-tone marker case since, for short gaps, the secondary tone onset is not inhibited, while for long gaps, the primary and secondary tone onsets are mutually inhibited.

The second scheme is simpler in that it does not involve inhibition. Instead, the central onset detector responds proportionally to the input it receives and is dominated by the most robust input. For a short gap, the onset response from the secondary tone is more robust than the onset response from the primary tone (which is in a state of forward masking), and therefore the overall response of the onset detector is dominated by the secondary tone onset at the expense of the primary tone onset. This constitutes onset masking. For a long gap, the onset responses from both tones are equally robust and so the overall response of the onset detector is greater than for a short gap, but it is not dominated by one tone onset over the other. As the scheme is presented here, no prediction is made regarding whether frequency-specific onset information is available to the listener, or whether onset information is irrevocably pooled across frequency.

In summary, two hypothetical schemes have been presented that both result in onset masking. In one, the masking is effected by diminishing the onset response through inhibition; in the other, the masking is effected by allowing the dominant onset to "capture" the whole response. Given this focus on onset responses, it was considered informative to explore their role using an electrophysiological analog of GDD. Experiment 2 therefore undertook to compare ABRs evoked by isofrequency pure-tone markers with those evoked by frequency-asymmetric markers having a pure-tone leading marker and a two-tone trailing marker. Because the ABR is essentially an onset response, where the magnitude of the response corresponds to the degree of synchronous neural firing at stimulus onset (Hecox *et al.*, 1976; Folsom and Aurich, 1987), this measure is sensitive to how well the onset of the trailing marker *as a whole* is encoded. In addition, experiment 2 included psychophysical GDD assessment for the same stimuli used to elicit the ABRs in order to enable direct comparison between electrophysiological and psychophysical performance.

In order to facilitate complementary measures of psychophysical GDD and ABRs evoked by the same stimuli, two changes were made to the stimulus configurations used in experiment 1. First, the markers were reduced in duration since the use of nominal durations of 300 ms made the overall stimulus duration (two markers plus gap) untenable for reasonable ABR recordings. Second, the frequencies of the primary and secondary tones were switched. The rationale for stipulating 4 000 Hz as the primary tone here was that any effect of adding the secondary tone (now 2 285 Hz) could not be attributed to a more basal cochlear shift of the site of stimulation. The ABR for wideband stimuli is known

to be dominated by the heightened neural synchrony associated with the base of the cochlea (Eggermont and Don, 1980). A basalward shift in the site of stimulation would be expected to result in a reduced response latency. Two marker configurations were tested: (1) a baseline within-frequency-channel configuration where both the leading and trailing markers were the primary tone alone; and (2) a frequency-asymmetric configuration where the leading marker was the primary tone alone and the trailing marker was the two-tone complex.

A. Method

1. Subjects

Six observers aged 23–43 years (mean=29.3 years) with normal hearing participated (observers 4, 7–11). All had audiometric thresholds ≤20 dB HL across the octave frequencies 250–8000 Hz (ANSI, 1996). Only observer 4 had participated in experiment 1.

2. Stimuli

The leading marker was always the 4000 Hz primary tone presented in isolation. In the baseline condition, the trailing marker also consisted of the primary tone presented in isolation. In a second condition, the trailing marker consisted of the primary and secondary tones presented simultaneously. Both the leading and trailing markers had nominal durations of 100 ms, but the actual duration of each marker varied by ±50% on each and every presentation. The actual duration included cosine-squared rise/fall ramps that were 2 ms in duration for the rise and 8 ms in duration for the fall. These rise/fall times were selected to facilitate synchronous neural responses to the marker onsets but not the offsets in the ABR procedure (see below, Fig. 3). The markers were generated digitally at a rate of 20 kHz (TDT AP2), output through one channel of a 16 bit DAC (TDT PD1), and low pass filtered at 8 kHz (Kemo VBF-8). For the ABR procedure, a TTL trigger pulse was simultaneously output through a second DAC channel. This pulse was synchronous with the onset of the trailing marker and was used to externally trigger the ABR recording system (Bio-logic Systems Traveler, 8 bit analog-to-digital converter). The average interval between successive triggers was approximately 300 ms for a 0 ms gap. (This trigger interval, of course, varied with the random marker duration.) This meant that, given the average marker duration of 100 ms, the interval between the offset of the trailing marker in one marker pair and the onset of the leading marker in the next following marker pair was about 100 ms. Markers were presented to the left ear at a spectrum level of 65 dB SPL through an Etymotic ER-2 insert phone.[2] No background masker was employed.

3. Psychophysical GDD procedure

Psychophysical GDD was measured as in experiment 1. For each of the two conditions, listeners practiced the task until performance was stable. At this point, data collection was initiated, and at least four replications of threshold estimate were collected for each condition. The final threshold for that condition was taken as the geometric mean of all estimates collected.

4. ABR procedure

Each listener was prepared for single-channel ABR recordings by placing surface electrodes on the high forehead (*Fz*, noninverting), left earlobe (*A*1, inverting), and low forehead (*Fpz*, ground). Electrode impedances were maintained at <2 kΩ throughout the recording sessions. Listeners reclined in a chair in a darkened, sound-treated room and were encouraged to sleep. The raw EEG was bandpass filtered from 100–3 000 Hz, and a gain of X150 000 applied. The recording window was 20 ms in duration, and 1 500 sweeps were averaged for a single recording. For each condition, at least three replicate traces were recorded. Given that each stimulus consisted of a fixed-duration gap bounded by two markers with nominal durations of 100 ms (see below), the recording time for a single 1 500-sweep trace exceeded nine minutes. A single recording session lasted no longer than two hours, and multiple sessions were required on different days to complete the data collection.

ABR recordings were time locked to the onset of the trailing marker. In a control condition, no leading marker was presented and the "trailing" marker consisted of the 4 000 Hz primary tone alone. This condition provided a reference for the ABR elicited by a 4 000 Hz tone with a nominal duration of 100 ms presented at 65 dB SPL in the absence of forward masking. In the within-frequency-channel condition set, the trailing marker continued to be the primary tone alone, but now the leading marker was reintroduced. Here, ABRs were recorded for fixed gap durations between the two primary tones of 0, 4, 8, 16, 32, and 64 ms. In this set of conditions, the ABRs can be considered to be responses to a 4 000 Hz tone that is forward masked by a preceding 4 000 Hz tone. In a third set of conditions, the trailing marker consisted of the primary and secondary tone pair. The ABRs were recorded for fixed gap durations between the leading and trailing markers of 0 and 64 ms.[3] It is important to note that, because the actual durations of both leading and trailing markers varied randomly on a presentation-by-presentation basis (100 ms±50%), and because the trigger pulse synchronized the recording to the onset of the trailing marker, the recording was not time locked to the onset of the leading marker.

B. Results and discussion

The results of the psychophysical GDD task are shown in Fig. 2 where the individual data (open symbols) cluster around the means (solid symbols ±1 standard deviation). Thresholds in the baseline within-frequency-channel condition are higher than those observed in the respective baseline condition of experiment 1 (cf. Fig. 1) and this is likely due to a combination of a shorter nominal marker duration coupled with an increase in the degree of duration randomization allowed in experiment 2. It is known that gap detection performance can depend on overall marker duration (Schneider and Hamstra, 1999; Snell and Hu, 1999), and deteriorates

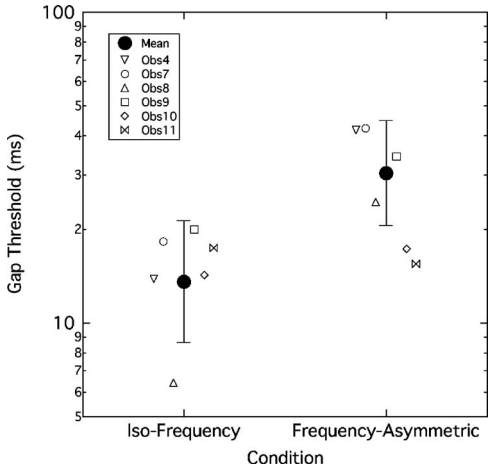


FIG. 2. GDD thresholds for individual observers (open symbols) and group mean (solid symbol) for the 2 conditions of experiment 2. Error bars are ±1 standard deviation.

with increases in marker duration variability (Penner, 1976). Although, on average, the thresholds are elevated in the frequency-asymmetric condition relative to the baseline, it is evident that for two of the six listeners (observers 10 and 11) there was little or no apparent effect of the addition of the secondary tone. Nevertheless, as a group, the difference between the two conditions was significant ($t_5=2.88$; $p=0.034$).

The ABR waveforms are portrayed in Figs. 3–5. A preliminary analysis was performed to confirm that the marker envelope shape was effective in eliciting an onset response, but not an offset response. In Fig. 3, ABR waveforms elicited by a single primary tone marker are shown for trigger pulses positioned either at the onset of the marker (left panels) or at the offset (right panels). The responses from three observers are shown in the lower rows: the individual wave- forms are shown overlaid on the left side of each pair of panels, and the weighted grand average of all collected sweeps on the right. The arrow marks the position of wave *V*. It can be seen that the 2 ms onset ramp generated sufficient neural synchrony to elicit a measurable response, whereas the 8 ms offset ramp did not. This result ensures that the ABR elicited by a trigger positioned at the onset of the trailing marker is indeed associated with that trailing marker onset and not with the offset of the leading marker.

Figure 4 shows a set of ABR results from one observer. The overlaid traces in the left column are individual sweeps and the right column shows the associated weighted grand average. The upper row depicts responses where the leading marker was absent. These reference traces indicate responses elicited by the primary tone "trailing" marker in the absence of any forward masking. The middle block of rows shows responses for the within-frequency-channel condition set, where the leading and trailing markers both consisted of the primary tone alone. Each row shows responses for a fixed gap duration. The lower block of rows shows responses for the frequency-asymmetric condition set. The vertical dashed line in the right column is positioned at the wave *V* latency for the unmasked reference condition. Two main findings are evident in this individual data set. First, for the within-frequency-channel condition set (middle block), wave *V* latency increased and waveform morphology deteriorated as the gap duration decreased. This pattern has been interpreted in terms of the forward masking of the trailing marker by the leading marker (e.g., Walton *et al.*, 1999). The second finding is that the presence of the secondary tone in the frequency-asymmetric trailing marker largely overcame the forward masking effect such that response latency did not shift appreciably with gap duration, and response morphology remained robust. Indeed, comparing the frequency-asymmetric traces (bottom) and the reference traces (top) in Fig. 4, it is apparent that the ABR evoked with the two-tone

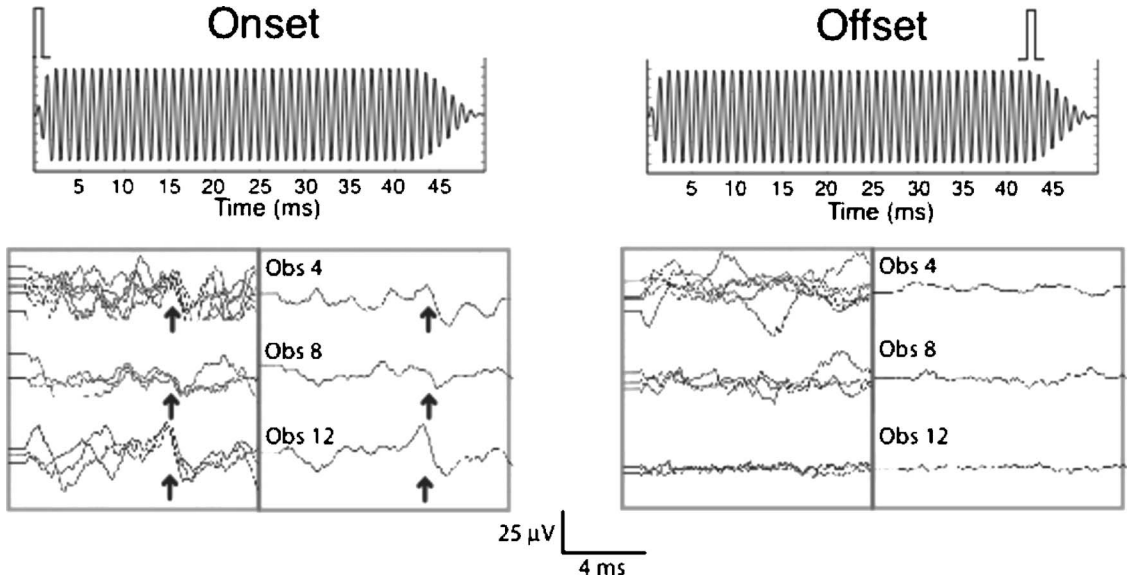


FIG. 3. Upper schematics illustrate the position of the trigger pulse relative to the 2 ms onset ramp (left) and the 8 ms offset ramp (right). The 50 ms stimulus, duration is for illustration purposes only. Lower panels show ABR waveforms from three observers elicited by stimulus onset (left pair of panels) and stimulus offset (right pair of panels). For each observer, the overlaid traces in the left-hand panel are individual sweeps, and the single trace in the right-hand panel is the weighted average of the individual sweeps. The arrows mark the position of wave *V*.

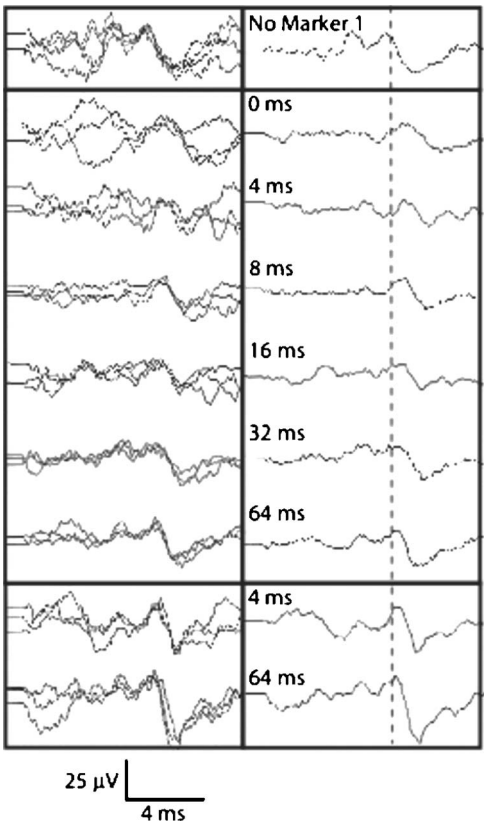


FIG. 4. ABR waveforms from one individual. The left column shows individual traces overlaid and the right column shows weighted averages. The upper row shows the reference condition where the leading marker was removed and only the primary tone trailing marker was present. The center block of rows shows the responses for the baseline within-frequency-channel condition set where the leading and trailing markers consisted of the primary tone alone. Each row depicts a fixed gap duration as stipulated in the inset. The bottom block of rows shows responses for the frequency-asymmetric condition set. The vertical line in the right column is positioned at the wave *V* latency for the reference condition.

trailing marker has a larger wave *V* amplitude than that evoked with a single-tone trailing marker, even in the absence of forward masking.

To demonstrate that this pattern was common across observers, Fig. 5 shows the weighted grand average responses from each of the six observers in separate columns. In general, the presence of the secondary tone in the frequency-asymmetric trailing marker leads to a larger wave *V* amplitude and nullifies any latency prolongation associated with the forward-masking effects of the leading marker, even for a gap duration of 0 ms and even though the secondary tone is lower in frequency than the primary tone. To quantify these latency shifts and amplitude changes, the wave *V* latency and amplitude for each grand average wave form was measured relative to the respective values in the unmasked reference.[4] The results are shown in Fig. 6 where latency shift (upper panel) and amplitude change (lower panel) are plotted as a function of gap duration. The thick solid line is the mean for the within-frequency-channel condition set and the dashed line is the mean for the frequency-asymmetric condition set. The individual data are plotted as filled and open symbols for the two condition sets, respectively. A repeated-measures ANOVA was undertaken on the latency shifts for gap durations of 0 ms and 64 ms for the two different condition sets. The analysis indicated that the main effects of gap duration and condition were significant ($F_{1,5}=61.44$; $p=0.001$ and $F_{1,5}=31.63$; $p=0.002$, respectively), as was the interaction between these factors ($F_{1,5}=9.71$; $p=0.026$). The significant interaction arose because the difference in latency shifts between condition sets was greater for a gap duration of 0 ms than for 64 ms. However, even for a gap duration of 64 ms the difference in latency shifts was still significant, as indicated by a Bonferroni-corrected means comparison ($t_5=3.63$; $p=0.015$). No significant latency shift was observed

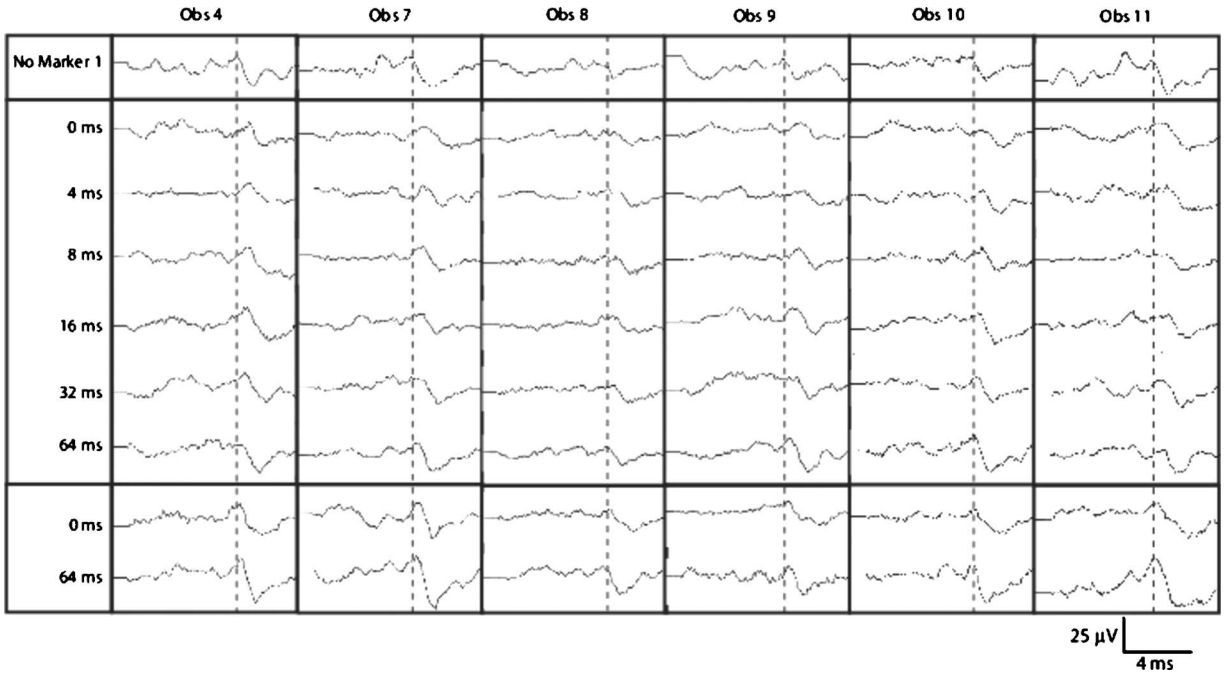


FIG. 5. Weighted average ABR traces from six observers. The upper row denotes the reference condition with no forward masking, with the dashed line indicating the latency of wave *V*. The center block of traces are elicited by the isofrequency markers with fixed gap durations as indicated in the left column. The lower block of traces are elicited by frequency-asymmetric markers at two fixed gap durations.

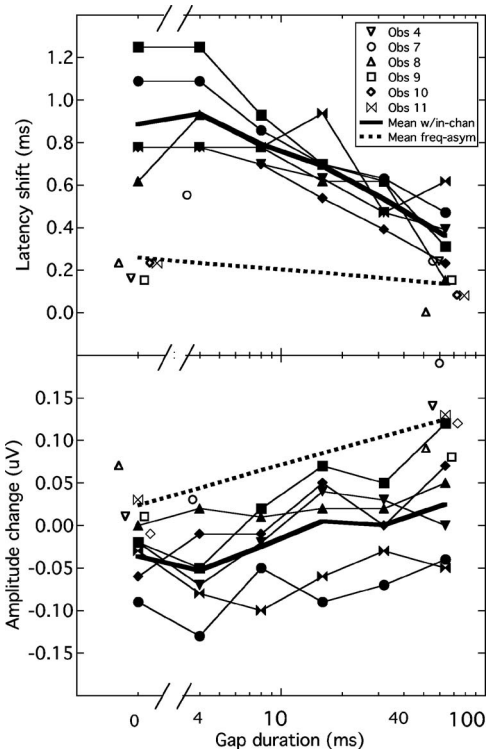


FIG. 6. Wave *V* latency shift (upper panel) and amplitude change (lower pannel) from baseline as a function of gap duration. Filled symbols connected by lines are individual data for the isofrequency condition set. The thick solid line indicates the group mean. Unfilled symbols are individual data for the frequency-asymmetric condition set. The dashed line indicates the group mean.

for the frequency-asymmetric condition set between a gap duration of 0 ms and 64 ms ($t_5=2.15$; $p=0.084$). The corresponding analysis for the amplitude changes indicated significant main effects of gap duration and condition ($F_{1,5}=27.1$; $p=0.003$ and $F_{1,5}=9.9$; $p=0.025$, respectively). The interaction between these factors was not significant ($F_{1,5}=1.38$; $p=0.29$).

The ABR data clearly indicate that the evoked potentials elicited by a complex trailing marker are more robust than those elicited by a trailing pure-tone marker, presumably due largely to the fact that the secondary tone is not being forward masked by the leading primary tone. This suggests that the elevated thresholds seen in the psychophysical data are not due to some sort of physiological onset masking wherein mutual inhibitions between the two tones of the complex marker result in the onset of the complex marker being less well encoded relative to that of the single primary tone marker alone. Rather, the results suggest that the onset of the complex trailing marker is dominated by the onset of the secondary tone (which is free of forward masking by the primary tone). If this onset dominance translates to a perceptual dominance, then the psychophysical GDD task becomes predominantly one of across-frequency-channel temporal processing rather than within-frequency-channel temporal processing. That is, the perceptual dominance of the secondary tone biases the GDD task towards across-frequency-channel temporal processing and, because across-frequency-channel GDD is poorer than within-frequency-channel GDD (see experiment 1), this leads to an elevation in threshold.

The notion of perceptual dominance provides a framework to account for the psychophysical order effect seen for frequency-asymmetric markers. The occurrence of novel spectral events associated with the complex trailing marker dominates the percept and biases the GDD task to one of across-frequency-channel processing. In the reverse configuration with the pure-tone trailing marker there are no novel spectral events associated with the trailing marker and therefore the temporal task is less biased towards across-frequency-channel processing. Another way of expressing this is in terms of a form of informational masking: the secondary tone in the complex trailing marker exerts informational masking over the primary tone because of its perceptual dominance, and this biases the task towards across-frequency-channel temporal processing, undermining performance. Less informational masking occurs when the secondary tone is part of the leading marker because it is not as perceptually dominant (the primary tone is not being forward masked). The interpretation of the order effect for frequency-asymmetric markers in terms of the perceptual dominance of the secondary tone suggests that if this dominance could be reduced, then the order effect would diminish. That is, if the amount of informational masking exerted by the secondary tone is attenuated, the elevated GDD thresholds should recover. The final experiment was designed to test this possibility.

IV. EXPERIMENT 3. INFORMATIONAL MASKING AND GDD FOR FREQUENCY-ASYMMETRIC MARKERS

The premise of this experiment is that informational masking exerted by the secondary tone over the primary tone undermines the listener's ability to make use of optimal within-frequency-channel cues. This informational masking has been described above in terms of the perceptual dominance of the secondary tone rendering the primary tone less salient. There are actually at least two—not mutually exclusive—ways in which this reduced saliency could come about. The first is that the perceptual dominance of the secondary tone might "capture the attention" of the listener, effectively diverting auditory processing power away from the primary tone. The second way is that the secondary tone might promote synthetic listening (in contrast to analytic listening) rendering the primary tone less salient because it is perceptually fused with the tonal complex. These distinctions will be revisited later but, to begin with, informational masking will be treated generically as a form of nonenergetic masking reflecting perceptual dominance of the secondary tone. The purpose of this experiment was to test the hypothesis that reducing the potential for informational masking leads to a recovery in GDD performance. Two stimulus manipulations were implemented that were designed to reduce informational masking by promoting perceptual segregation between the primary and secondary tones, thus enhancing primary tone saliency. The first manipulation (test 1) varied the mode of secondary tone presentation such that it was delivered either to the ear contralateral to the primary tone (dichotic) or to both ears simultaneously (diotic); the reference condition was the monaural presentation of both primary and secondary tones. The rationale here is that infor-

mational masking using simple tonal sequences is largely constrained to configurations where both target and masking tone sequences are presented to the same ear; contralateral presentation of the masking tone sequence typically has minimal effect (Kidd *et al.*, 2003). The second manipulation (test 2) involved placing the primary tone between *two* secondary tones such that the frequency spacing of the three-tone complex was uniform on a log scale. In addition, the flanking secondary tones could be frequency modulated. The rationale here is that perceptual fusion between a single component embedded in a complex and the remaining components of the complex can be diminished if the tone does not share a harmonic relationship with the complex (Moore *et al.*, 1986). Furthermore, a primary component can be perceptually isolated (i.e., it will "pop out") if the secondary components vary in frequency while the primary remains invariant (McAdams, 1989; Kidd *et al.*, 1994).

A. Method

1. Subjects

Different crews of normal-hearing adults participated in the two tests. The six observers from experiment 1 (observers 1–6) participated in test 1. Seven observers (observers 4, 12–17) participated in test 2. They ranged in age from 21–48 years (mean=33 years) and included one listener who had participated in both experiments 1 and 2.

2. Stimuli

For test 1, the frequencies of the primary and secondary tones, as well as their temporal parameters, were the same as described in experiment 1. Here, the frequency-asymmetric marker conditions of experiment 1 were repeated, but with the secondary tone presented either to both ears (diotic) or only to the ear contralateral to the ear receiving the primary tone (dichotic). For test 2, the frequency of the primary tone remained 2 285 Hz, but several other stimulus parameters were changed relative to experiment 1. First, two secondary tones were added to the primary tone to form the complex marker: the higher of these secondary tones was 4 000 Hz (the same frequency as used in the previous experiments) and the lower was 1 306 Hz. These three frequencies are equally spaced on a log scale. The second stimulus modification was that, in some conditions, frequency modulation (FM) was applied to the flanking secondary tones (but never to the primary tone) according to the equation

$$x(t) = A \sin\left[\left(2\pi \int_c t + \theta\right) + m \sin\left(2\pi \int_m t\right)\right].$$

The rate of FM (f_m), when present, was 10 Hz with a modulation index (m) of 10, and the starting phase of the modulator was always 0 rad; the starting phase (θ) of each secondary tone (f_c) was randomly selected on a presentation-by-presentation basis. In order to make marker duration an unreliable cue, while at the same time retaining complete 100 ms cycles of FM, two fixed marker durations were permitted, 200 ms and 300 ms, and the duration selection for each marker in a stimulus was random on a presentation-by-presentation basis. The stimuli in test 2 were presented at a sound pressure level of 65 dB/Hz in a broadband background noise (low passed at 8 000 Hz) having a pressure spectrum level of 10 dB/Hz.

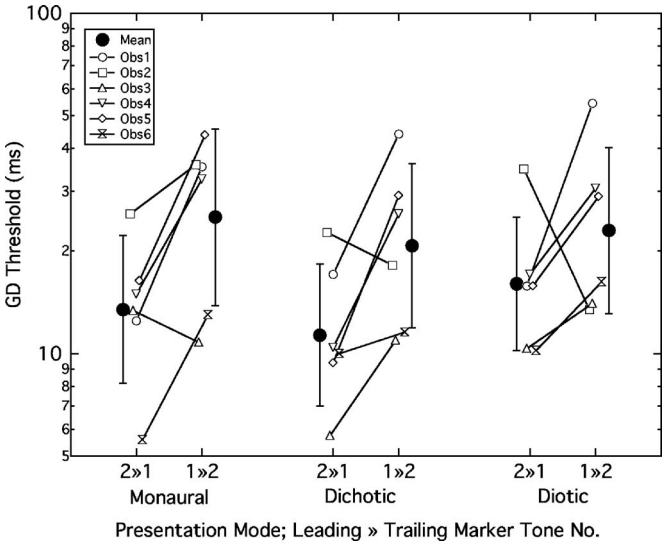


FIG. 7. GDD thresholds for test 1 of experiment 3. For each mode of presentation (monaural, dichotic, diotic), individual data for the two-tone leading marker ($2 \gg 1$) and the two-tone trailing marker ($1 \gg 2$) are shown as connected open symbols. The solid circles are mean data, with ±1 standard deviation error bars.

3. Procedure

The procedure was the same as that employed in experiment 1. For the observers in test 1 at least three, but occasionally four, replications were obtained per condition, and the final estimate was taken as the geometric mean of all replications obtained. For the observers in test 2, at least four, but often five if time permitted, replications were obtained per condition, and the final estimate was taken as the geometric mean of all replications obtained. All observers repeated their entire assigned conditions, with a stipulated number of replications, once for practice and once for data retention.

B. Results and discussion

The results for test 1, where the secondary tone was presented dichotically or diotically, are shown in Fig. 7. Because the same listeners participated in experiment 1, the reference monaural frequency-asymmetric data are replicated from Fig. 1. The order of the frequency-asymmetric markers is shown by the number of tones in the leading and trailing markers (leading≫trailing). For each mode of secondary tone presentation, the individual data are shown as open symbols connected by lines. The mean data are shown as solid circles with error bars. It is evident that the order effect seen in the monaural conditions of experiment 1 was retained here for most listeners in both the dichotic and diotic conditions of test 1; however, exceptions exist (observer 3 in monaural and observer 2 in dichotic and diotic). A two-factor repeated-measures ANOVA was undertaken to assess the effect of marker order (two levels) and mode of secondary tone presentation (three levels).

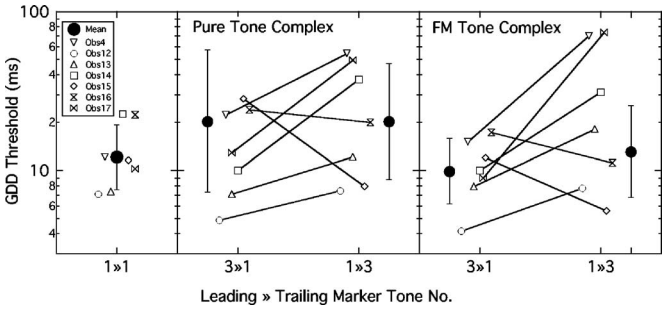


FIG. 8. GDD thresholds for test 2 of experiment 3. The left panel shows baseline isofrequency thresholds. The center panel shows thresholds for the frequency-asymmetric marker pair where the secondary tones were pure tones. The right panel shows thresholds for the frequency-asymmetric marker pair where the secondary tones were FM tones. In each panel, individual data for the three-tone leading marker ($3 \gg 1$) and the three-tone trailing marker ($1 \gg 3$) are shown as connected open symbols. The solid circles are mean data, with ±1 standard deviation error bars.

The analysis indicated a significant effect of marker order ($F_{1,5}=6.881$; $p=0.047$), but no effect of mode of presentation and no significant interaction between these two factors. This result indicates that, as a group, performance was worse when the trailing gap marker was the two-tone complex than when it was the single tone, irrespective of whether the secondary tone was presented to the ipsilateral ear, contralateral ear, or simultaneously to both ears. The continued observance of the order effect, even when the secondary tone is presented to the contralateral ear, is further evidence against a peripherally based onset masking mechanism.

The results for test 2 are shown in Fig. 8. The left panel shows the baseline isofrequency condition where both the leading and trailing markers were the single primary tone. The center panel shows the pair of frequency-asymmetric conditions where the secondary tones were pure tones. The right panel shows the frequency-asymmetric conditions where the secondary tones were FM tones. The order of the frequency-asymmetric markers is indicated by the number of tones in the leading and trailing markers (leading $\gg$ trailing). Individual data are shown as open symbols connected by lines; mean data are shown as solid circles. Although the majority of listeners exhibited the expected order effect for frequency-asymmetric markers, wherein performance is poorer for a complex trailing marker than a complex leading marker, there were noticeable exceptions (observers 15 and 16). In addition, the pattern of data is grossly similar across the pure-tone and FM-tone marker complexes. The data for the frequency-asymmetric markers were submitted to a two-factor repeated-measures ANOVA to assess the effects of complex marker type (pure-tone, FM-tone) and order (two levels). The analysis indicated no effect of either the FM factor or the order factor ($F_{1,6}=1.639$; $p=0.248$ and $F_{1,6}=2.559$; $p=0.161$, respectively). The lack of an order effect is not surprising given the marked individual variability in the data patterns. (If the data of observers 15 and 16 are omitted, the analysis now reveals a significant effect of marker order [$F_{1,4}=25.04$; $p=0.007$], but the factor of FM remains insignificant [$F_{1,4}=0.043$; $p=0.846$]).

The results of this experiment reinforce the general observation made in experiments 1 and 2 that the order effect for frequency-asymmetric markers is characteristic of most listeners, but is not without exception. That is, whereas most listeners find GDD to be more challenging when the complex marker is the trailing marker, a few observers exhibit either no effect or actually find the complex leading marker to be the more difficult configuration. (Recall that one of the four cochlear implant listeners in the van Wieringen and Wouters (1999) study also showed no order effect.) The presence of marked variability across listeners itself constitutes indirect support for the role of informational masking in the frequency-asymmetric marker effect: Individual variability is characteristic of informational masking, where some listeners appear to be more susceptible than others to nonenergetic masking (Neff and Callaghan, 1988; Durlach *et al.*, 2003).

If a role for informational masking in GDD for frequency-asymmetric markers is to be entertained, the results of experiment 3 shed light on how this informational masking should be understood. As noted earlier, one possible form of informational masking is that the secondary tone promotes synthetic listening and this renders the primary tone less salient because it is perceptually fused with the tonal complex. If this is the case, then it would be expected that stimulus manipulations that weakened perceptual fusion should diminish the effect. However, the results of test 1 strongly suggest that failure of analytic listening is not the key: the order effect remained intact even with contralateral presentation of the secondary tone. Kidd *et al.* (2003) have shown that contralateral presentation of informational maskers largely negates their masking effectiveness. In test 2, the imposition of FM on the secondary tones also did not affect the performance. McAdams (1989) has shown that differential FM can lead to parsing of complex sounds; thus the test 2 results also do not support an "unmasking" due to release from synthetic listening. Furthermore, the notion of informational masking conferred by synthetic listening in frequency-asymmetric GDD is not supported by the original work of Formby *et al.* (1998b). They found that performance with the two-tone trailing marker remained poor even when a temporal asynchrony of several tens of milliseconds was introduced between the primary and secondary tones. Most studies of informational masking observe a release from masking in the presence of temporal asynchrony between the target and interfering tones (e.g., Kidd *et al.*, 1994; Hall *et al.*, 2005).[5] The results of this experiment therefore do not support a role of informational masking conferred by synthetic listening.

A second possible form of informational masking is a type of distraction effect wherein the perceptual dominance of the secondary tone draws the listener's attention away from the primary tone, thus reducing the saliency of the primary tone. The perceptual dominance of the secondary tone in the complex trailing marker is likely due in part to its spectral novelty and in part to the fact that the primary tone is in a state of recovery from forward masking by the leading primary tone marker. The ABR data of experiment 2 indicate that the primary tone is still undergoing some degree of forward masking at the gap durations corresponding to discrimination thresholds. In other words, the primary tone is internally represented with less fidelity than the secondary

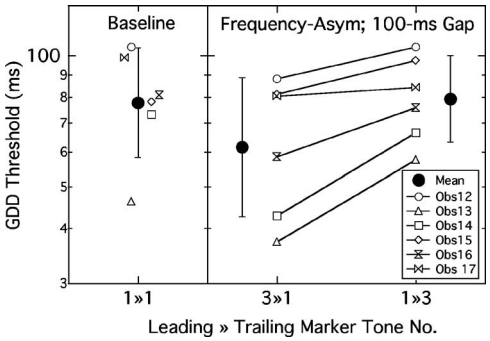


FIG. 9. GDD thresholds for a standard gap duration of 100 ms. The left panel shows baseline isofrequency thresholds. The right panel shows thresholds for the frequency-asymmetric markers plotted as a function of the number of tones in the leading and trailing markers. Solid circles are mean data; ±1 standard deviation, and individual data are shown as open symbols.

tones (which are not being forward masked). This combination of unmasked secondary tones and partially masked primary tone might contribute to the perceptual dominance of the secondary tones. If this hypothesis is correct, then it is predicted that the perceptual dominance of the secondary tones in the complex trailing marker should be partially mitigated by introducing a longer recovery interval between the leading and trailing markers. That is, if the primary tone in the trailing marker is no longer being forward masked by the leading primary tone marker its relative salience should recover.

To test this hypothesis, a supplementary set of conditions was undertaken on six of the seven observers who participated in test 2. Here, the reference and pure tone frequency-asymmetric conditions of test 2 were repeated but with a simple prolongation of the standard gap to 100 ms. All other details of the procedure remained unchanged. The results of the supplementary conditions are shown in Fig. 9. As expected, the baseline thresholds were elevated with respect to the 0 ms standard gap (Fig. 8) since GDD is proportional to the standard gap duration (e.g., Abel, 1972; Penner, 1976; Grose *et al.*, 2006). The key finding, however, is that for most listeners the order effect remained intact. This was confirmed with a repeated-measures ANOVA on the three conditions.

The analysis indicated a significant effect of condition ($F_{2,10}=8.267$; $p=0.008$), with post-hoc contrasts indicating that thresholds for the complex leading marker were lower than for the complex trailing marker ($F_{1,5}=15.967$; $p=0.01$). Remarkably, this effect was evident for the two listeners (observers 15 and 16) who showed a reversed order effect for the 0 ms standard gap duration. The results of the supplementary conditions therefore indicate that informational masking due to the perceptual dominance of the secondary tones in the complex trailing marker is unlikely to be due to the primary tone being in a state of partial forward masking. Rather, the results are consistent with an interpretation that the perceptual dominance of the secondary tones is due to their spectral novelty.[6]

It is of interest to conjecture why spectral novelty may result in a form of informational masking in this paradigm. One possibility already mentioned is that spectral novelty confers a perceptual dominance, and this perceptual dominance translates to a spectral biasing that makes the GDD task more of an across-frequency-channel temporal process than a within-frequency-channel process. Another possibility is that the occurrence of a novel event—in this case spectral—entails a compulsory allocation of limited attentional resources to the novel event, with the result that less capacity is available for the coding of other aspects of the stimulus—in this case the gap duration. This notion of limited attentional resources has been proposed in a number of contexts to model deficits in temporal processing in complex listening tasks (e.g., Fitzgibbons *et al.*, 1974; Phillips *et al.*, 1997; Vanneste and Pouthas, 1999). If the onset of a novel spectral event engages attentional resources that would otherwise be devoted to registering the characteristics of the silent interval, it is possible that the fidelity with which the interval is encoded is compromised. This reduced fidelity could apply to any or all aspects of the interval (i.e., offset, elapsed time, or onset). This possibility is reminiscent of the suggestion by van Wieringen and Wouters (1999) that the abrupt perceptual change associated with the onset of the complex trailing marker might interfere with the memory trace of the leading marker.

V. SUMMARY AND CONCLUSION

The three experiments in this study were designed to examine the nature of the order effect for frequency-asymmetric markers wherein GDD performance is poorer when the complex marker is the trailing marker than when it is the leading marker. Experiment 1 confirmed the order effect and demonstrated that it resulted in thresholds that are intermediate between within-frequency-channel processing and across-frequency-channel processing. However, the effect was less striking and more variable than the report by Formby *et al.* (1998b). Experiment 2 demonstrated that, whereas the addition of a secondary tone to the trailing marker led to a decline in GDD performance, the ABR elicited by this stimulus was more robust than in the absence of the secondary tone. Experiment 3 indicated that informational masking associated with synthetic listening is unlikely to underlie the order effect. Instead, informational masking resulting from the occurrence of novel spectral events in the complex trailing marker provides an account that is consistent with the pattern of results. The wide range of individual performance indicates that susceptibility to informational masking in this paradigm is observer-dependent. In conclusion, this study demonstrated that the order effect associated with frequency-asymmetric gap markers in the GDD paradigm is typical of most, but not all, listeners and is likely due to a form of informational masking.

ACKNOWLEDGMENTS

This work was supported by NIDCD R01-DC01507. We thank Andrew Oxenham, Bob Burkard, and an anonymous reviewer for their helpful comments on a previous version of this paper.

[1]Stimuli in all psychophysical testing were transduced by Sennheiser 580 headphones. These were calibrated using a flat-plate coupler seated on a Bruel & Kjaer 4135 Artificial Ear. The coupler housed a Larson-Davis 0.5 in. microphone (Model 2559), and the microphone output was read by a Larson-Davis sound level meter (Model 800B) set to a linear (unweighted) scale.

[2]Stimuli transduced with the ER-2 insert phones in the ABR study were calibrated using a GRAS 2 cc coupler (IEC 126, Type RA0038). The same microphone and sound level meter were used as noted previously.

[3]The minimum gap duration for observer 7 in this condition set was 4 ms, not 0 ms, but for the purposes of analysis this datum was treated as 0 ms.

[4]For the purposes of measurement, wave *V* was identified as the peak in the 6.5–8.4 ms latency window. Its amplitude was measured as the voltage difference between this peak and the following trough (V'). For each grand average wave form, positioning of the actual *V* and V' markings was determined visually based on the pattern of overlap of the multiple individual traces making up the grand average. Three judges, all experienced in ABR testing, contributed to the wave form markings. Although no formal cross checking across judges was undertaken, informal comparisons indicated agreement among judges.

[5]In fact, the listeners in experiment 1 also undertook a condition where, for the complex leading marker, the secondary tone began 75 ms *after* the onset of the primary tone (both tones were coterminous at the gap onset). There was no "masking release" effect observed for this temporal asynchrony ($t_5=0.956$; $p=0.383$), again suggesting that informational masking conferred by synthetic listening was not a key factor.

[6]Analysis of the supplementary conditions also revealed the unexpected finding that performance for the complex leading marker condition was better than for the baseline isofrequency condition ($F_{1,5}=9.059$; $p=0.03$). There does not appear to be any precedence for this finding in the literature; that is, GDD thresholds for complementary conditions with minimal standard gaps are either similar to (Formby *et al.*, 1998b), or elevated with respect to (e.g., Grose *et al.*, 2001; Pichora-Fuller *et al.*, 2006), the isofrequency baseline. It is possible that the presence of multiple frequency components enables some listeners to improve the coding of the offset of the leading marker, thereby enhancing the gap discrimination performance, but further work is required to clarify this.

Abel, S. M. (**1972**). "Discrimination of temporal gaps," J. Acoust. Soc. Am. **52**, 519–524.

Allen, P. D., Virag, T. M., and Ison, J. R. (**2002**). "Humans detect gaps in broadband noise according to effective gap duration without additional cues from abrupt envelope changes," J. Acoust. Soc. Am. **112**, 2967–2974.

ANSI (**1996**). *American National Standards Specification for Audiometers (ANSI S3-1996)* (American National Standards Institute, New York).

Barsz, K., Benson, P. K., and Walton, J. P. (**1998**). "Gap encoding by inferior collicular neurons is altered by minimal changes in signal envelope," Hear. Res. **115**, 13–26.

Durlach, N. I., Mason, C. R., Shinn-Cunningham, B. G., Arbogast, T. L., Colburn, H. S., and Kidd, G., Jr. (**2003**). "Informational masking: Counteracting the effects of stimulus uncertainty by decreasing target-masker similarity," J. Acoust. Soc. Am. **114**, 368–379.

Eggermont, J. J., and Don, M. (**1980**). "Analysis of the click-evoked brainstem potentials in humans using high-pass noise masking. II. Effect of click intensity," J. Acoust. Soc. Am. **68**, 1671–1675.

Fishbach, A., Nelken, I., and Yeshurun, Y. (**2001**). "Auditory edge detection: A neural model for physiological and psychoacoustical responses to amplitude transients," J. Neurophysiol. **85**, 2303–2323.

Fitzgibbons, P. J., Pollatsek, A., and Thomas, I. B. (**1974**). "Detection of temporal gaps within and between perceptual tonal groups," Percept. Psychophys. **16**, 522–528.

Folsom, R. C., and Aurich, C. D. (**1987**). "Auditory brainstem responses from human adults and infants: Influence of stimulus onset," Audiology **26**, 117–122.

Formby, C., Gerber, M. J., Sherlock, L. P., and Magder, L. S. (**1998a**). "Evidence for an across-frequency, between-channel process in asymptotic monaural temporal gap detection," J. Acoust. Soc. Am. **103**, 3554–3560.

Formby, C., Sherlock, L. P., and Forrest, T. G. (**1996**). "An asymmetric roex filter model for describing detection of silent temporal gaps in sinusoidal markers," Aud. Neurosci. **3**, 1–20.

Formby, C., Sherlock, L. P., and Li, S. (**1998b**). "Temporal gap detection measured with multiple sinusoidal markers: Effects of marker number, frequency, and temporal position," J. Acoust. Soc. Am. **104**, 984–998.

Grose, J. H., Hall, J. W. III, and Buss, E. (**2006**). "Temporal processing deficits in the pre-senescent auditory system," J. Acoust. Soc. Am. **119**, 2305–2315.

Grose, J. H., Hall, J. W. III, Buss, E., and Hatch, D. (**2001**). "Gap detection for similar and dissimilar gap markers," J. Acoust. Soc. Am. **109**, 1587–1595.

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.

Hecox, K., Squires, N., and Galambos, R. (**1976**). "Brainstem auditory evoked reponses in man. I. Effect of stimulus risefall time and duration," J. Acoust. Soc. Am. **60**, 1187–1192.

Heil, P. (**1997**). "Auditory cortical onset responses revisited. I. First-spike timing," J. Neurophysiol. **77**, 2616–2141.

Ison, J. R., Castro, J., Allen, P., Virag, T. M., and Walton, J. P. (**2002**). "The relative detectability for mice of gaps having different ramp durations at their onset and offset boundaries," J. Acoust. Soc. Am. **112**, 740–747.

Kidd, G., Jr., Mason, C. R., Arbogast, T. L., Brungart, D. S., and Simpson, B. D. (**2003**). "Informational masking caused by contralateral stimulation," J. Acoust. Soc. Am. **113**, 1594–1603.

Kidd, G., Jr., Mason, C. R., Deliwala, P. S., Woods, W. S., and Colburn, H. S. (**1994**). "Reducing informational masking by sound segregation," J. Acoust. Soc. Am. **95**, 3475–3480.

McAdams, S. (**1989**). "Segregation of concurrent sounds. I: Effects of frequency modulation coherence," J. Acoust. Soc. Am. **86**, 2148–2159.

Moore, B. C. J., Glasberg, B. R., and Peters, R. W. (**1986**). "Thresholds for hearing mistuned partials as separate tones in harmonic complexes," J. Acoust. Soc. Am. **80**, 479–483.

Neff, D. L., and Callaghan, B. P. (**1988**). "Effective properties of multicomponent simultaneous maskers under conditions of uncertainty," J. Acoust. Soc. Am. **83**, 1833–1838.

Oxenham, A. J. (**2000**). "Influence of spatial and temporal coding on auditory gap detection," J. Acoust. Soc. Am. **107**, 2215–2223.

Penner, M. J. (**1976**). "The effect of marker variability on the discrimination of temporal intervals," Percept. Psychophys. **19**, 466–469.

Phillips, D. P., and Hall, S. E. (**2002**). "Auditory temporal gap detection for noise markers with partially overlapping and non-overlapping spectra," Hear. Res. **174**, 133–141.

Phillips, D. P., Hall, S. E., and Boehnke, S. E. (**2002**). "Central auditory onset responses, and temporal asymmetries in auditory perception," Hear. Res. **167**, 192–205.

Phillips, D. P., Taylor, T. L., Hall, S. E., Carr, M. M., and Mossop, J. E. (**1997**). "Detection of silent intervals between noises activating different perceptual channels: Some properties of 'central' auditory gap detection," J. Acoust. Soc. Am. **101**, 3694–3705.

Pichora-Fuller, M. K., Schneider, B. A., Benson, N. J., Hamstra, S. J., and Storzer, E. (**2006**). "Effect of age on detection of gaps in speech and nonspeech markers varying in duration and spectral symmetry," J. Acoust. Soc. Am. **119**, 1143–1155.

Plack, C. J., Gallun, F. J., Hafter, E. R., and Raimond, A. (**2006**). "The detection of increments and decrements is not facilitated by abrupt onsets or offsets," J. Acoust. Soc. Am. **119**, 3950–3959.

Plomp, R. (**1964**). "Rate of decay of auditory sensation," J. Acoust. Soc. Am. **36**, 277–282.

Pressnitzer, D., Meddis, R., Delahaye, R., and Winter, I. M. (**2001**). "Physiological correlates of comodulation masking release in the mammalian ventral cochlear nucleus," J. Neurosci. **21**, 6377–6386.

Schneider, B. A., and Hamstra, S. J. (**1999**). "Gap detection thresholds as a function of tonal duration for younger and older listeners," J. Acoust. Soc. Am. **106**, 371–380.

Snell, K. B., and Hu, H. L. (**1999**). "The effect of temporal placement on gap detectability," J. Acoust. Soc. Am. **106**, 3571–3577.

van Wieringen, A., and Wouters, J. (**1999**). "Gap detection in single- and multi-channel stimuli by LAURA cochlear implantees," J. Acoust. Soc. Am. **106**, 1925–1939.

Vanneste, S., and Pouthas, V. (**1999**). "Timing in aging: The role of attention," Exp. Aging Res. **25**, 49–67.

Walton, J., Orlando, M., and Burkard, R. (**1999**). "Auditory brainstem response forward-masking recovery functions in older humans with normal hearing," Hear. Res. **127**, 86–94.

Age-related differences in discrimination of an interval separating onsets of successive tone bursts as a function of interval duration

Peter J. Fitzgibbons
Department of Hearing, Speech, and Language Sciences, Gallaudet University, Washington, DC 20002

Sandra Gordon-Salant and Jessica Barrett
Department of Hearing and Speech Sciences, University of Maryland, College Park, Maryland 20742

(Received 6 September 2006; revised 19 April 2007; accepted 20 April 2007)

The study measured listener sensitivity to increments in the inter-onset interval (IOI) separating pairs of successive 20-ms 4000-Hz tone pulses. A silent interval between the tone pulses was adjusted across conditions to create reference tonal IOI values of 25–600 ms. For each condition, a duration DL for increments of the tonal IOI was measured in listeners comprised of young normal-hearing adults and two groups of older adults with and without high-frequency hearing loss. Discrimination performance of all listeners was poorest for the shorter reference IOIs, and improved to stable levels for longer reference intervals exceeding about 200 ms. Temporal sensitivity of the young listeners was significantly better than that of the elderly listeners in each condition, with the largest age-related differences observed for the shortest reference interval. Age-related differences were also observed for duration DLs measured using single 4000-Hz tone bursts set to three reference durations in the range 50–200 ms. The tone DLs of all listeners were smaller than the corresponding tone-pair IOI DLs, particularly for the shorter reference stimulus durations. There were no significant performance differences observed between the older listeners with and without hearing loss for either discrimination task.
[DOI: 10.1121/1.2739409]

PACS number(s): 43.66.Mk, 43.66.Sr [JHG] Pages: 458–466

I. INTRODUCTION

There is growing awareness among auditory researchers that aging is associated with a general decline in auditory temporal processing, one that appears to be largely independent of factors related to audiometric hearing loss. Observations of the age-related processing decline came from earlier studies revealing that older listeners had particular difficulty understanding speech that was temporally altered in some manner by rate alteration or reverberation (Bergman, 1980; Wingfield *et al.*, 1985; Gordon-Salant and Fitzgibbons, 1993; Vaughan and Letowski, 1997). The altered speech signals are acoustically complex, but feature a number of correlated changes in the temporal properties of the signals that are thought to underlie some of the processing difficulties among older adult listeners. Some of these changes influence the overall timing and prosodic characteristics of speech, while other changes occur at segmental levels and alter the relative durations of bursts, formant transitions, and silent intervals that can serve as acoustic cues to phoneme identity. The possibility that older listeners exhibit a decline in sensitivity for some of these speech temporal cues prompted an interest in studying the temporal processing abilities of older listeners using controlled samples of both speech and nonspeech sounds.

The predominant emphasis of investigation has involved use of psychophysical experiments to measure listeners' ability to detect or discriminate very brief time intervals inserted between successive stimulus markers, defined usually by pairs of tones or noise bursts. The measured detection thresholds for these temporal gaps in older listeners are observed in several studies to be larger than those measured in younger listeners (Schneider *et al.*, 1994; Snell, 1997; Strouse *et al.*, 1998). Additionally, some related findings indicate that discrimination thresholds for brief temporal gaps are elevated among older listeners (Grose *et al.*, 2006), while other findings indicate that gap discrimination difficulties among older listeners become more pronounced when measured using spectrally disparate stimulus markers consisting of either speech or nonspeech sounds (Lister *et al.*, 2002; Lister and Tarver, 2004; Pichora-Fuller *et al.*, 2006; Grose *et al.*, 2006). Other research findings indicate that the limitations in temporal gap resolution among older listeners are most apparent for shorter duration stimuli (Schneider and Hamstra, 1999), or for temporal gaps inserted near the onsets of longer signals (He *et al.*, 1999). Each of the studies concludes that the age effects observed for the detection or discrimination of brief temporal gaps are largely independent of factors associated with age-related changes in hearing sensitivity.

A smaller number of studies on aging and duration discrimination have been conducted using stimuli that span a longer range of reference durations, as defined by the extent of tone or noise signals, or an interval of silence inserted between a pair of acoustic signals. One such study by Abel *et al.* (1990) used filtered noise-burst stimuli and observed that older listeners exhibited diminished sensitivity to changes in the duration of relatively brief 20-ms reference signals, but not longer signals of 200-ms duration. Later, Fitzgibbons and

 0001-4966/2007/122(1)/458/9/$23.00

Gordon-Salant (1994) observed reduced abilities among older listeners to discriminate changes in the duration of either a 250-ms tone burst or a silent interval of equivalent duration inserted between a pair of 250-ms tonal markers. More recently, Grose *et al.* (2006) showed that age-related declines in temporal sensitivity are evident at earlier stages of aging than considered previously. This discrimination study found that older (65–83 years) and middle-aged (40–55 years) listeners exhibited similar and significantly poorer sensitivity than young listeners (18–27 years) to changes in the duration of an intertone silent interval, with the age effects observed for reference silent intervals of 0, 35, and 250 ms. Grose *et al.* also observed that age-related deficits in temporal sensitivity sometimes could be exacerbated by increases in task complexity, a result that was also observed in our earlier discrimination experiments conducted with multi-tone sequences that featured varying degrees of stimulus complexity and task demands (Fitzgibbons and Gordon-Salant, 1995, 2001). Each of the duration discrimination studies conducted with the longer reference stimuli, like those conducted with minimum-duration temporal gaps, reports that the observed age-related differences in discrimination performance are not significantly influenced by the presence or absence of hearing loss in some of older listeners.

One finding that emerges from several of the earlier experiments on aging and temporal discrimination concerns the relative importance of stimulus duration. That is, some studies report that the predominant age-related difficulties associated with temporal discriminations are restricted to the processing of relatively brief sounds. For example, Schneider and Hamstra (1999) reported that the largest age-related differences in temporal gap detection occur for stimulus durations less than about 250 ms. Gap detection results reported by Muchnik *et al.* (1985) and He *et al.* (1999) also support the conclusion that gap detection deficits among older listeners are more pronounced when signals bounding a temporal gap are short in duration. For experiments on duration discrimination, the influence of the reference stimulus duration is less clear, because most of the available discrimination results have come from testing with a small number of stimulus conditions. One notable exception is a study on aging and duration discrimination conducted by Bergeson *et al.* (2001), who tested groups of younger and older listeners using a 2-kHz tonal signal that was set to different reference durations within the broad range of 1.5–1000 ms. These results with the tonal stimuli revealed large discrimination deficits among older listeners for brief tones less than about 20–40 ms, but little or no age-related performance differences for tones of longer reference duration. These findings for duration discrimination, like some of those reported for temporal gap detection, suggest that the reduction in temporal sensitivity among older listeners is primarily associated with the processing of relatively brief stimuli.

Less is known about the manner in which the discrimination of silent intervals by older listeners is influenced by the magnitude of the reference interval. Collective evidence from earlier studies conducted with trained young listeners indicates that the Weber fraction associated with the discrimination of silent intervals bounded by pairs of stimulus markers becomes progressively larger as the duration of the reference interval is reduced systematically below about 100–200 ms (Abel, 1972b; Getty, 1975; Penner, 1976). Corresponding results for older listeners are currently restricted to a smaller sample of reference silent intervals (e.g., Grose *et al.*, 2006), and do not indicate that age-related declines in gap discrimination become disproportionately larger for shorter reference intervals, as appears to be the case for tonal stimuli (Bergeson *et al.*, 2001). However, there are some reasons to anticipate that older listeners might be particularly disadvantaged in discriminating changes in the duration of brief temporal intervals. For example, most theoretical accounts that are relevant to the processing of stimulus duration invoke the operation of central timing mechanisms that are thought to act as counters to sum neural firings during stimulation to code signal duration (e.g., Creelman, 1962; Abel, 1972b; Divenyi and Danner, 1977). These accounts also postulate a high degree of precision in the sensory coding of stimulus onsets and offsets in order to accurately mark stimulus boundaries. It would seem to follow that any age-related changes within the auditory system that degrade the sensory response to stimulus onsets and offsets could influence listeners' temporal discrimination performance.

These considerations prompted us to measure the abilities of younger and older listeners to discriminate changes in the interval separating the onsets of brief tone pulses over a range of reference durations. The tone pulses within stimulus pairs were created to have rapid onsets in order to emphasize the onset-to-onset interval as the relevant timing cue for the discrimination measures, as suggested also by Penner (1976) and Divenyi and Danner (1977). If aging of the auditory system is associated with impoverished coding of signal onsets, a loss in sensitivity for the onset-to-onset interval is expected for older listeners, particularly for brief intervals with closely spaced stimulus onsets. Some preliminary support for this hypothesis emerged from results of our earlier experiments (Fitzgibbons and Gordon-Salant, 2001) with multi-tone isochronous stimulus sequences in which older listeners exhibited diminished sensitivity to changes in sequence tempo, or rhythm, particularly for stimulus sequences that featured relatively short tonal onset-to-onset intervals (e.g., 100 ms). However, interpretation of the earlier tempo discrimination results for multi-tone sequences is complicated by the current lack of corresponding discrimination data for single stimulus intervals over an extended range of reference durations. The present experiments are designed to collect these measurements.

In addition to the main discrimination experiments conducted with the successive tone-pair stimuli, the experiment also included a smaller number of conditions that measured duration discrimination for a single tonal stimulus set to three reference durations in the range 50–200 ms. Initially, this range of durations was selected for its correspondence to a range of speech phoneme durations that we are currently investigating as temporal cues in distinguishing various word-pair contrasts such as *beat* versus *wheat*, which differ in initial consonant transition duration (approximately 50 ms), or *wheat* versus *weed*, which differ in vowel duration

TABLE I. Mean pure tone air conduction thresholds and standard deviations (in dB HL, *re*: ANSI, 2004) across frequency for the three listener groups.

	Pure tone frequency (Hz)					
	250	500	1000	2000	4000	8000
Younger normal hearing						
Mean	8.0	6.0	1.0	2.0	1.5	5.5
Standard deviation	5.9	5.7	5.6	6.7	5.8	5.5
Older normal hearing						
Mean	15.0	11.5	12.0	16.0	20.5	48.0
Standard deviation	7.1	5.3	6.7	8.1	4.6	21.8
Older hearing loss						
Mean	16.5	20.0	25.0	38.0	51.0	71.7
Standard deviation	7.8	10.3	11.0	11.4	6.6	9.0

(approximately 200 ms). Also, the discrimination measures for samples of shorter reference tone durations were collected because our initial observation of an age-related difference in tonal duration discrimination was restricted to a single longer tone of 250 ms (Fitzgibbons and Gordon-Salant, 1994). Moreover, the results collected with tonal stimuli can provide some needed comparison data for a range of reference stimulus durations that appeared to show little or no age-related performance differences in previous duration discrimination studies (Abel *et al.*, 1990; Bergeson *et al.*, 2001).

In all stimulus conditions of the experiment, discrimination performance was examined in groups of younger and older listeners. Testing in each condition was restricted to a high-frequency region, and the potential effects of hearing loss were examined by comparing discrimination performance between groups of older listeners with and without hearing loss in the frequency region of the test stimuli.

II. METHOD

A. Subjects

Listeners in the experiments included 30 subjects assigned to three groups of ten each according to age and hearing status. One group (younger normal hearing) included younger listeners ranging in age from 18 to 30 years (mean =21.4 years) with mean pure-tone thresholds $\leq$20 dB HL (*re*: ANSI, 2004) across the octave frequencies 250–8000 Hz. Another group (older normal hearing) included older listeners ranging in age from 67 to 78 years (mean =72.4 years) with normal mean pure-tone thresholds $\leq$20 dB HL from 250 to 4000 Hz. For these listeners, hearing loss varied in degree at 8000 Hz, exhibiting a mean value of 48 dB HL, with thresholds $\leq$35 dB HL for four of the listeners and thresholds $\geq$45 dB HL for six of the listeners. The third group included older listeners (older hearing loss) ranging in age from 67 to 78 years (M=73.0 years) with bilateral mild-to-moderate sloping high-frequency sensorineural hearing losses from 250 to 8000 Hz. These listeners had a negative history of otologic disease, noise exposure, and family history of hearing loss. The probable etiology of hearing loss in the older listeners was presbycusis. Table I presents the mean tone thresholds in dB HL for the test ears of the listeners in each group of subjects. Additional criteria for subject selection included monosyllabic word recognition scores in quiet $\geq$80% (Northwestern University Auditory Test No. 6), normal middle ear function as assessed by tympanometry, and acoustic reflex thresholds that were within the 90th percentile for a given pure tone threshold (Gelfand *et al.*, 1990). Also, the listeners were tested using transient-evoked otoacoustic emissions (TEOAEs). Click stimuli were presented at 80 dB peak-equivalent SPL using the ILO88 OAE system. Band reproducibility (percent) and band signal-to-noise ratio (SNR) were analyzed within frequency bands centered from 1 through 4 kHz. Criteria for the presence of TEOAEs was band reproducibility $>$70% and SNR $>$6 dB (Robinette and Glattke, 2007). All listeners with normal hearing (except one of the older normal-hearing listeners) demonstrated TEOAEs that met these criteria. None of the listeners in the older hearing loss group met the criteria to document the presence of TEOAEs. The absence of TEOAEs and the presence of acoustic reflex thresholds at expected levels confirmed a cochlear site of lesion in listeners with hearing loss. All listeners were in general good health, with no history of stroke or neurological impairment and possessed sufficient motor skills to provide responses using a computer keyboard. Additionally, all listeners passed a screening test for general cognitive awareness (Pfeiffer, 1977). The listeners had not participated previously as subjects in psychoacoustic experiments and were paid for their services in the study.

B. Stimuli

The tonal stimuli for the experiments were generated using an inverse fast Fourier transform (FFT) procedure with a digital signal processing board (Tucker-Davis Technologies, AP2) and 16-bit D/A converter (Tucker-Davis Technologies DD1, 20-kHz sampling rate) that was followed by low-pass filtering (Frequency Devices 901F, 6000-Hz cutoff, 90 dB/oct). All testing was conducted using a stimulus frequency that was selected to coincide with a spectral region of sensitivity loss in the listeners with hearing impairment. The tone-pair stimuli were constructed using 4000-Hz tone bursts separated in time by a silent interval. Each tone burst within a pair had a fixed duration of 20 ms that included 2.5-ms cosine-squared rise/fall envelopes, with all tone and silent-interval durations specified between zero-voltage points on the electrical waveforms. Within each stimulus pair, the silent interval between the tones was adjusted to establish the desired reference tonal inter-onset interval (IOI), an interval that included the duration of the leading tone burst and the intertone silent interval. Six discrimination conditions were evaluated with the stimulus tone pairs, using reference IOI values of 25, 50, 100, 200, 400, and 600 ms. For each condition, comparison tone pairs used for the discrimination trials were the same as the reference pairs, but featured a longer IOI that was varied across trials to measure a duration DL for increments of the IOI. Figure 1 displays a sample waveform for a tone pair featuring an IOI of 25 ms (panel a), along with the corresponding amplitude spectrum (panel b). Also shown in the figure is a waveform and corresponding amplitude spectrum for a tone pair with an IOI of 30 ms in panels

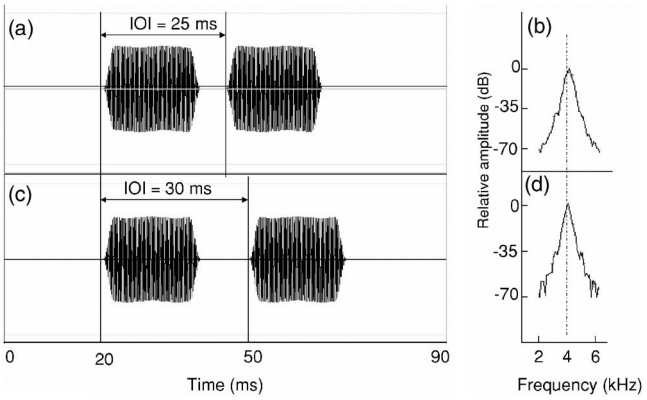


FIG. 1. Waveform and amplitude spectrum of tone-pair stimuli with IOI =25 ms [panels (a) and (b), respectively], and waveform and amplitude spectrum of tone-pair stimuli with IOI=30 ms [panels (c) and (d), respectively].

c and d, respectively. As the figure reveals, the amplitude spectra of the two tone-pair stimuli are approximately the same. For each tone-pair waveform, the spread of spectral energy in the 4-kHz region is determined primarily by the duration of the tonal markers, with the intertone silent interval affecting details of spectral side-band structure. Thus, with all stimulus pairs in the experiment having fixed tone-burst durations, changes in the tonal IOI do not alter the spectral spread of energy in a manner that is likely to serve as a reliable discrimination cue for listeners.

Three additional duration discrimination conditions were conducted using single 4000-Hz tone bursts set to reference durations of 50, 100, and 200 ms, values that included the 2.5-ms rise/fall envelopes described above. For each reference value, a similar duration DL for increments was measured by varying the duration of a comparison tone across a series of discrimination trials. For these tonal stimuli, increments in duration will produce a correlated narrowing of the spectral energy distribution, a result that could provide a potential spectral cue for listeners discriminating a change in stimulus duration. However, these potential spectral cues in tonal duration discrimination have been shown previously to influence listeners' discrimination performance only for very brief stimulus durations, less than 5–10 ms (Small and Campbell, 1962; Abel, 1972a). In the present experiment, all testing with the tonal stimuli was restricted to a range of longer reference durations, where spectral factors do not exert a significant influence on listeners' duration discrimination performance.

C. Procedure

For the stimulus tone pairs, the measurement of DLs for the tonal inter-onset interval was obtained using an adaptive three-interval, two-alternative forced-choice discrimination procedure. Each discrimination trial contained three observation intervals spaced 500 ms apart. The first listening interval of each trial contained a sample of the reference stimulus pair, with the second and third intervals containing samples of the reference and comparison stimulus pairs in either order selected randomly across listening trials. For each of the six IOI conditions, the reference and comparison tone pairs of a listening trial differed only by the duration of the intertone silent interval, which was always longer in the comparison stimulus pair. Listeners used a keyboard to respond to the comparison stimulus in the second or third observation interval of each trial. All listening intervals were marked by a visual display that also provided correct-interval feedback for each trial.

Estimates of the duration DLs were obtained using an adaptive rule for varying the comparison tonal IOI such that the IOI decreased in magnitude following two consecutive correct responses by the listener and increased in magnitude following each incorrect response. Threshold estimates derived by this adaptive rule corresponded to values associated with 70.7% correct discrimination (Levitt, 1971). Testing in each condition was conducted in 50-trial blocks with an IOI starting value 1.4 times the reference value, and a step size for IOI change that decreased logarithmically over trials to produce rapid convergence on threshold values. Following the first three reversals in direction of IOI change, a threshold estimate was calculated by averaging reversal-point IOI values associated with remaining even-numbered reversals. The same procedures were used to measure duration DLs for the simple tonal signals for each of the three reference duration values. For these measurements, the duration of the comparison tone on each listening trial was varied adaptively to measure the DL for a duration increment. An average of six threshold estimates was used to calculate a final DL for each listener in each discrimination condition. Prior to data collection, each listener received eight to ten practice blocks of trials in each condition, with each listener showing performance stability after three to five blocks of trials in each condition.

The listeners were tested individually in a sound-treated booth. The nine discrimination conditions (six reference tone-pair IOIs, and three reference durations for the single tone) were tested in a different randomly selected order for each listener. Stimulus levels were 85–90 dB SPL in order to provide adequate audibility for the older listeners with high-frequency hearing loss. Each of the older listeners participated in an audibility screening using a Bekesy tracking procedures to insure that stimulus audibility for the 20-ms tone bursts corresponded to minimum sensation levels of 25–30 dB at 4000 Hz. The stimuli were delivered to listeners through an insert earphone (Etymotic ER-3A) that was calibrated in a 2-cm^3 coupler (B & K, DB0138). Testing was monaural in the better ear of listeners with hearing loss, and in the preferred ear of listeners with normal hearing. Listening was conducted in 2-h sessions over the course of several weeks. Total test time (not including practice sessions) varied across listeners, but averaged 7–8 hours.

III. RESULTS

For the purpose of analysis and comparison of results across stimulus conditions, all duration DLs collected with tone-pair stimuli were converted to relative values expressed as a fraction of the reference tonal IOI (i.e., the Weber fraction). Table II displays these mean relative DLs for each reference IOI value for each of the younger and older listener

TABLE II. Mean relative difference limens (DLs) in each of the tonal IOI conditions for the three listener groups. Standard errors of the means are shown also.

	Tonal IOI (ms)					
	25	50	100	200	400	600
Younger normal hearing						
Mean	0.25	0.18	0.18	0.11	0.07	0.06
Standard error	0.02	0.02	0.02	0.01	0.01	0.01
Older normal hearing						
Mean	0.62	0.44	0.35	0.20	0.13	0.14
Standard error	0.04	0.02	0.02	0.02	0.01	0.01
Older hearing loss						
Mean	0.60	0.43	0.36	0.21	0.14	0.13
Standard error	0.05	0.02	0.02	0.01	0.01	0.01

groups, with the standard errors of the means also shown. Performance variability among listeners in each of the older groups was about double that of the younger listeners for the 25-ms IOI reference, but there was relatively little difference in performance variability among listener groups across the range of longer IOI reference values. An analysis of variance (ANOVA) was conducted on the individual relative DL values using a repeated-measures design with one within-subjects variable (reference IOI) and one between-subjects variable (listener group). Results of the analysis revealed significant main effects of the reference IOI $[F(5,135)=176.83, p<0.01]$ and listener group $[F(2,27)=76.33, p<0.01]$, and a significant interaction between IOI and listener group $[F(10,135)=12.03, p<0.01]$. *Posthoc* analysis of simple group effects in the data and subsequent multiple comparison tests (Scheffe) revealed that the performance of the younger listeners was significantly better than that of either group of older listeners ($p<0.05$) for each IOI value, and there was no significant performance difference between the two groups of older listeners with normal hearing and with hearing loss at each IOI value. Additional *posthoc* analysis of simple IOI effects revealed that, for each listener group, discrimination thresholds did not vary significantly across the range of longer reference IOIs of 200–600 ms ($p>0.05$). For the older listeners, thresholds increased progressively as the reference IOI decreased from 100 to 25 ms, with each relative DL in this range of shorter reference IOIs being significantly larger than those observed for the longer IOI reference intervals ($p<0.01$, each comparison). Discrimination thresholds of the young listeners were also elevated for the shorter reference IOIs, but only the thresholds for the shortest 25-ms IOI proved to be significantly larger ($p<0.01$) than thresholds measured for the longer IOIs in the 200–600-ms range.

Because audiometric hearing loss among the older listeners produced no significant effects in discrimination performance, results for the two older listener groups were subsequently collapsed and compared to those for the younger listeners. These results are displayed in Fig. 2, which shows the mean relative DLs in percent as a function of the tonal IOI(ms) for the younger and older listeners, with error bars in the figure representing standard errors of the means. Results in the figure reveal that the relative DLs of the younger and older listeners decrease from larger values for the shorter IOIs to smaller relatively stable values for the longer IOIs. For each of the reference IOIs, the mean relative DLs of the older listeners are larger than corresponding values of the young listeners, with the greatest age-related performance difference observed for the shortest tonal IOI.

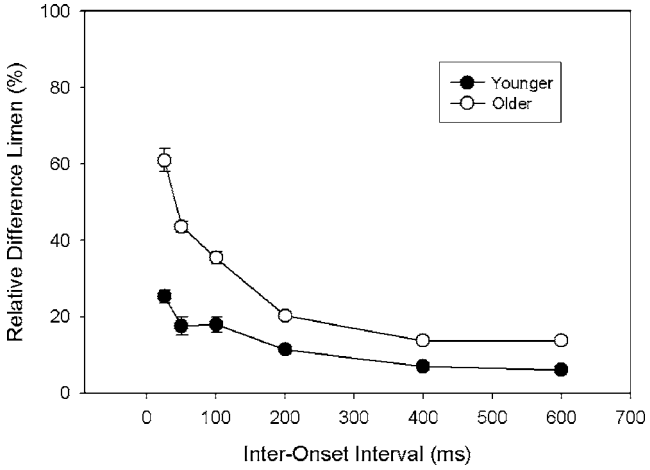


FIG. 2. Mean relative difference limens in percent as a function of the tonal inter-onset interval (ms) for the younger listeners (filled symbols) and older listeners with and without hearing loss (unfilled symbols). Error bars represent standard errors of the means.

The mean relative tone DLs measured for the three reference durations tested are shown in Table III for each group of listeners, with values representing standard errors of the means also displayed. For these tone discrimination results, a separate ANOVA was conducted on the individual relative tone DLs using a repeated-measures design with one within-subjects variable (reference tone duration) and one between-subjects factor (listener group). The analysis revealed significant main effects of reference duration $[F(2,54)=24.8, p<0.01]$ and listener group $[F(2,27)=315.1, p<0.01]$, with no significant interaction between the two variables. Multiple comparison testing (Scheffe) on group means revealed that the younger listeners performed significantly better than the older listeners for each reference tone duration ($p<0.01$), with no significant performance differences observed between the two groups of older listeners with normal hearing

TABLE III. Mean relative difference limens (DLs) for each reference tone duration for the three listener groups. Standard errors of means are shown also.

	Reference duration (ms)		
	50	100	200
Younger normal hearing			
Mean	0.15	0.11	0.09
Standard error	0.01	0.01	0.01
Older normal hearing			
Mean	0.33	0.29	0.21
Standard error	0.02	0.02	0.01
Older hearing loss			
Mean	0.34	0.30	0.24
Standard error	0.02	0.02	0.01

TABLE IV. Results of stepwise multiple regression analyses conducted with each DL value in the various discrimination conditions serving as the criterion variable, and age, 4 kHz threshold, and 8 kHz threshold as predictor variables. R^2 entries indicate the variance accounted for (VAF) by the derived regression equations, with associated F values and significance levels. Significant predictor variables are indicated by *.

Criterion variable (DL)	R^2 VAF	F	Age	4 kHz threshold	8 kHz threshold
IOI 25 ms	68.4%	58.4, $p<0.01$	$t=2.6$, $p<0.01$*	$t=0.6$, $p>0.05$	$t=0.2$, $p>0.05$
IOI 50 ms	76.6%	88.3, $p<0.01$	$t=9.4$, $p<0.01$*	$t=0.5$, $p>0.05$	$t=0.9$, $p>0.05$
IOI 100 ms	58.8%	38.5, $p<0.01$	$t=6.2$, $p<0.01$*	$t=-0.5$, $p>0.05$	$t=-0.4$, $p>0.05$
IOI 200 ms	47.7%	24.9, $p<0.01$	$t=5.0$, $p<0.01$*	$t=-0.5$, $p>0.05$	$t=-1.3$, $p>0.05$
IOI 400 ms	53.6%	31.1, $p<0.01$	$t=5.6$, $p<0.01$*	$t=-1.2$, $p>0.05$	$t=-1.1$, $p>0.05$
IOI 600 ms	76.5%	87.7, $p<0.01$	$t=9.4$, $p<0.01$*	$t=1.0$, $p>0.05$	$t=0.3$, $p>0.05$
tone 50 ms	79.7%	106.2, $p<0.01$	$t=10.3$, $p<0.01$*	$t=0.8$, $p>0.05$	$t=-0.1$, $p>0.05$
tone 100 ms	72.1%	69.8, $p<0.01$	$t=8.4$, $p<0.01$*	$t=-1.3$, $p>0.05$	$t=0.1$, $p>0.05$
tone 200 ms	84.5%	147.6, $p<0.01$	$t=12.1$, $p<0.01$*	$t=-2.2$, $p<0.05$	$t=-0.9$, $p>0.05$

and hearing loss. The main effect of reference duration was also analyzed using multiple comparison testing (Scheffe). The results showed that all listeners exhibited significantly larger DLs for the 50-ms tones than for the 200-ms tones ($p<0.01$). There were no significant differences in tone DLs between the 100 ms tone and either the 50- or 200-ms tones.

Subsequent multiple regression analyses were conducted on the discrimination results for the tone-pair stimuli and the single-tone stimulus conditions, to examine further the contributions of age and high-frequency hearing loss on the relative DL values. These analyses were motivated by the observation that the younger and older listeners with normal hearing appear to have different pure-tone thresholds, despite both groups satisfying the criterion of hearing within normal limits through 4 kHz. To that end, separate multiple regression analyses (stepwise method) were conducted on data from all 30 participants with the relative DL value in each condition serving as the criterion variable, and age, 4-kHz threshold, and 8-kHz threshold serving as the predictor variables. The results of these analyses are shown in Table IV. The table shows that the variance accounted for (R^2) by the regression equations ranged from 47.7% to 84.5%, and each was significant, across the different conditions. The results also show that the variable age was the primary significant predictor of the discrimination performance in each condition, whereas the listeners' detection thresholds at 4 and 8 kHz generally were not significant predictors of performance.

IV. DISCUSSION

The experiments were designed primarily to compare the abilities of younger and older listeners to discriminate changes in the duration of an interval separating the onsets of two brief tone bursts, for a range of shorter and longer reference intervals. Duration discrimination was also assessed for a smaller sample of reference durations using pure-tone stimuli. Results of the measurements indicated that auditory sensitivity to changes of stimulus duration depends on the age of listener, magnitude of the reference interval, and type of stimulus defining the reference interval.

A. Younger listeners

Results collected from the young adult listeners in the experiments exhibit trends that are similar in several ways to those reported in earlier studies. Several of the earlier studies on duration discrimination found that the Weber fraction remains fairly stable over a range of stimulus reference durations exceeding about 200 ms, but increases progressively as the reference duration is reduced within the range of shorter durations (Creelman, 1962; Small and Campbell, 1962; Abel, 1972a, b; Getty, 1975). In a similar manner, the results of our young listeners for tonal IOI discrimination showed only small variation in the relative DLs for reference IOI durations in the 200–600-ms range, and progressively larger values as the reference IOI is reduced. Other comparable studies that used tone-pair stimuli like those of the present experiments observed that young listeners could discriminate increments in tonal onset-to-onset intervals of about 6%–10%, with performance being fairly stable for reference intervals exceeding about 200 ms (e.g., Divenyi and Danner, 1977; Hirsh *et al.*, 1990; Drake and Botte, 1993). Young adult listeners in the present study produced relative DLs for increments of tonal IOI in the range of 6%–11% for the reference durations of 200 ms or longer, values that agree closely with the previously reported estimates.

The data collected from the younger listeners in the three duration discrimination conditions that used single tonal stimuli revealed mean relative DLs that shifted from about 9% to 14.5% as the reference tone duration decreased from 200 to 50 ms. This performance shift is similar in magnitude to that observed for the IOI DLs across the same range of reference durations, indicating that the two DL measures are closely related. This impression was confirmed with correlation analysis, which showed a high correlation between the IOI DLs and tone DLs at each of the three corresponding reference durations ($r>0.66$, $p<0.01$). The tone DLs of the younger listeners were somewhat smaller than the IOI DLs for corresponding reference durations of 50 and 100 ms, but the two DL measures were equivalent for the reference duration of 200 ms. These comparison results for the single-tone and tone-pair stimuli are similar in some respects to those reported in earlier duration discrimination

studies in which tone DLs were compared to the DLs for silent intervals, or gaps, inserted between a pair of acoustic markers. Some of these earlier results indicated that duration DLs for single tones were consistently smaller than those for temporal gaps of equal reference duration (e.g., Abel, 1972a; Rammsayer and Lima, 1991), although our earlier discrimination measurements using longer 250-ms reference duration found that duration DLs for tones and gaps were approximately the same in younger listeners (Fitzgibbons and Gordon-Salant, 1994). However, it should be noted that straightforward comparisons between tone DLs and gap DLs in some of the earlier experiments are complicated by findings that show that gap discrimination thresholds can be influenced by parameters of the acoustic markers surrounding the gap, particularly the duration of the leading marker (Penner, 1976; Grose *et al.*, 2001, 2006). These marker duration effects are less relevant to the present comparison of DLs across stimulus types, as the tonal IOI is used as the reference interval for all tone-pair discrimination testing.

The discrimination measures collected for the tone-pair stimuli show essentially the same trends observed previously in our temporal discrimination experiment conducted using sequences of five brief tone bursts separated equally by silent intervals (Fitzgibbons and Gordon-Salant, 2001). Listeners in this earlier study were asked to discriminate small changes in sequence presentation rate that were implemented by simultaneous variation of the tone-burst IOIs at different baseline sequence rates. Results with the tone sequences, like those of the present experiments, revealed nearly equivalent rate discrimination performance for sequences having IOIs in the 200–600-ms range, with somewhat poorer temporal sensitivity seen for a fastest sequence rate featuring an IOI of 100 ms. However, results from the two studies differ in one important respect. That is, all relative DLs measured in the present experiment with the tone-pair stimuli are about twice the magnitude of those measured previously with the five-tone sequences, for corresponding values of the reference tonal inter-onset interval. These performance comparisons suggest that the repetition of a fixed stimulus interval in multi-tone sequences may lead to improved discrimination performance, and that the duration DLs measured here for single reference intervals might not reflect the absolute limits in listeners' temporal sensitivity for tonal IOI.

B. Older listeners

For the older listeners, the magnitude of the relative IOI DLs with the tone-pair stimuli was fairly stable across the range of longer reference durations 200–600 ms, and became progressively larger as the reference duration was reduced. Across the range of reference durations tested, the mean relative DL for increments of tonal IOI for the older listeners shifted from about 14% at the longer durations to about 61% at the shortest reference duration of 25 ms. For each reference tonal IOI, the discrimination performance of the older listeners was found to be poorer than that of the younger listeners. The size of the age-related performance difference was greatest at the shortest reference IOI duration, where the mean relative DL of the older listeners was observed to be about 2.5 times that of the younger listeners. The age-related differences were smaller for the longer IOI values of 100–600 ms, but the relative DLs of the older listeners were still about twice those of the younger listeners across this range of tonal IOI. The performance differences among listener groups are primarily attributed to age, rather than shifts in hearing sensitivity in the two older groups, as the multiple regression analysis revealed that age was the primary factor that accounted for most of the variance in the discrimination DLs.

For the single-tone discrimination conditions, the mean relative DLs of the older listeners shifted progressively from 22.5% to 33.5% across conditions of decreasing reference tone duration from 200 to 50 ms. These relative tone DLs were found to be significantly larger than those of the younger listeners at each corresponding reference duration. However, unlike the discrimination results for the tone-pair stimuli, there was no evidence indicating that the magnitude of the age-related difference in tone DLs was disproportionately larger for the shortest reference tone duration. This particular outcome may simply reflect the limited range of reference durations tested with the single-tone stimuli. Like the younger listeners, the older listeners exhibited smaller tone DLs compared to IOI DLs for the shorter reference durations of 50 and 100 ms, but not for the longer 200-ms reference duration. For the shortest comparable reference duration of 50 ms, the mean IOI DLs were larger than tone DLs by about 10% for the older listeners, and about 3% for the younger listeners. Thus, it appears that for this shorter reference duration, the older listeners, and to lesser extent, the younger listeners exhibit greater relative difficulty discriminating changes in a tonal onset-to-onset interval compared to an equivalent onset-to-offset interval that defines the duration of a single tone.

One purpose of the experiments was to extend the examination of aging and duration discrimination to a broader range of reference intervals than examined in most of the previous investigations. Additionally, a specific goal of the testing was to examine the hypothesis that discrimination difficulties among older listeners become more pronounced for shorter reference stimulus durations. Support for this hypothesis is found in the tone-pair discrimination data, which shows larger age-related performance differences for stimulus durations less than about 100 ms, and smaller stable discrimination differences for longer reference durations. Bergeson *et al.* (2001) also observed exaggerated age-related deficits in duration discrimination for their shorter duration tonal stimuli of less than 20 ms, but age effects in these data were reported to be negligible across a broad range of longer reference durations exceeding 40–80 ms. However, Bergeson *et al.* also reported a high degree of performance variability among their older listeners, a situation that obscures interpretation of possible age effects in some of their results collected with longer reference stimuli. For example, although Bergeson *et al.* do not examine the statistical significance of age effects in their data analysis, they do report that the mean relative DLs for duration increments in their older and younger listeners were 60% and 30%, respectively, for a reference tone duration of 200 ms. These mean DL values

are much larger than corresponding estimates in the present study, but they show agreement with our findings in revealing that the mean discrimination thresholds of older listeners are about twice those of the younger listeners for the longer reference stimuli. This outcome describes the present discrimination results for our reference tonal stimulus of 200 ms, and each of the tone-pair conditions with IOIs in the range 100–600 ms.

Overall, the results of the experiments provide additional evidence for the existence of age-related temporal processing deficits that are largely independent of factors associated with audiometric hearing loss in older listeners. However, the sources of diminished temporal sensitivity among the older listeners remain unclear. It is possible that some aspects of the processing difficulties can be attributed to age-related changes in the central timing mechanisms that are thought to be implicated in discrimination tasks involving judgments about stimulus duration (e.g., Creelman, 1962). As discussed previously, the postulated timing mechanism is presumed to function as a counter that accumulates neural pulses during stimulation to code duration. However, the density of neural pulses feeding such a counter could become diminished simply as a consequence of an age-related reduction in the population of nerve fibers (Willott, 1990). In this case, longer stimulus increments would be required to discriminate duration differences by older listeners, compared to young listeners. This account would seem most applicable in explaining the smaller and relatively uniform discrimination deficits observed for the older listeners across the range of longer reference stimulus intervals examined in the present investigation.

The decrease in discrimination performance among older listeners observed for the tone-pair stimuli of relatively brief duration would seem to implicate a different process, perhaps one involved with the coding of signal onsets. For young listeners, the coding of stimulus boundaries is usually presumed to depend primarily on signal audibility, and Divenyi and Danner (1977) reported that levels of about 25 dB SL are sufficient to minimize uncertainty in registering signal onsets, a criterion that was satisfied in the present experiments for both younger and older listeners. However, in recent years, some investigators (e.g., Schneider and Pichora-Fuller, 2000) have pointed to physiological evidence from animal studies on aging indicating a possible age-related loss of synchrony in the nerve-fiber response patterns associated with stimulus onsets (e.g., Hellstrom and Schmiedt, 1990; Boettcher *et al.*, 1996). Other animal studies on aging in the neurosciences have observed age-related reductions in the number of subcortical neural units that are uniquely sensitive to brief time-separated auditory stimuli (Walton *et al.*, 1998; Frisina, 2001). These types of evidence point to a number of possible age-related changes within the auditory system that could influence accuracy in the timing of stimulus boundaries. In psychophysical experiments, any imprecision in the coding of signal onsets would be expected to influence temporal discriminations, particularly for signals of brief duration, or sequential stimuli with closely spaced onsets of the type used in the present experiments.

As mentioned earlier, many of the psychophysical studies on aging and auditory temporal processing were broadly motivated by a desire to learn if diminished temporal sensitivity was a principal source of the speech understanding difficulties demonstrated by many older listeners. For example, some of the earlier studies with older listeners applied correlation analysis techniques to their data and found significant relationships between speech recognition performance and specific psychophysical measures of temporal resolution, such as gap detection (Lutman, 1991) or gap duration discrimination (Gordon-Salant and Fitzgibbons, 1993). However, other investigators reported only weak correlations between measures of speech recognition and temporal sensitivity in their older listeners (Abel *et al.*, 1990; Humes, 1996). More recent investigations have addressed the issue by targeting specific temporal cues found in speech that are known to be used by listeners to distinguish phoneme categories. One such speech cue that is relevant to the present experiment is the brief silent interval found in spoken word samples containing unvoiced stop consonants (e.g., stay, stew, ditch), but which is absent in corresponding word-pair counterparts without the stop consonant (e.g., say, sue, dish). Two recent studies on aging reported that the magnitude of the silent interval at the perceptual boundary between word samples in a pair was significantly larger in older versus younger listeners (Gordon-Salant and Fitzgibbons, 2006; Grose *et al.*, 2006). These studies also collected discrimination measures for silent intervals and found that duration DLs among the older listeners were larger than those of the younger listeners. Additionally, the age-related difficulties in discriminating changes in the duration of brief tonal signals, as observed in the present experiments, are similar to some of our speech results, which show that older listeners required extended consonant glide transitions in order to distinguish word pairs such as *beat-wheat*. Thus, it appears that the age-related differences in the temporal discrimination, as observed in the earlier and present experiments, have relevance to speech processing as well.

ACKNOWLEDGMENTS

This research was supported by an individual research grant (R37AG09191) from the National Institute on Aging. The authors are grateful to Michele Spencer for her assistance in data collection.

Abel, S. M. (**1972a**). "Duration discrimination of noise and tone bursts," J. Acoust. Soc. Am. **51**, 1219–1223.

Abel, S. M. (**1972b**). "Discrimination of temporal gaps," J. Acoust. Soc. Am. **52**, 519–524.

Abel, S. M., Krever, E. M., and Alberti, P. W. (**1990**). "Auditory detection, discrimination, and speech processing in ageing, noise-sensitive and hearing-impaired listeners," Scand. Audiol. **19**, 43–54.

ANSI. (**2004**). ANSI S3.6-2004, "American National Standard Specification for Audiometers," (American National Standards Institute, New York).

Bergeson, T. R., Schneider, B. A., and Hamstra, S. J. (**2001**). "Duration discrimination in younger and older adults," Can. Acoust. **29**, 3–9.

Bergman, M. (**1980**). *Aging and the Perception of Speech* (University Park, Baltimore, MD).

Boettcher, F. A., Mills, J. H., Swerdloff, J. L., and Holley, B. L. (**1996**). "Auditory evoked potentials in aged gerbils: responses elicited by noises separated by a silent gap," Hear. Res. **102**, 167–178.

Creelman, C. D. (**1962**). "Human discrimination of auditory duration," J.

Acoust. Soc. Am. **34**, 582–593.
Divenyi, P. L., and Danner, W. F. (**1977**). "Discrimination of time intervals marked by brief acoustic pulses of various intensities and spectra," Percept. Psychophys. **21**, 125–142.
Drake, C., and Botte, M.-C. (**1993**). "Tempo sensitivity in auditory sequences: Evidence for a multiple-look model," Percept. Psychophys. **54**, 277–286.
Fitzgibbons, P. J., and Gordon-Salant, S. (**1994**). "Age effects on measures of auditory duration discrimination," J. Speech Hear. Res. **37**, 662–670.
Fitzgibbons, P. J., and Gordon-Salant, S. (**1995**). "Age effects on duration discrimination with simple and complex stimuli," J. Acoust. Soc. Am. **98**, 3140–3145.
Fitzgibbons, P. J., and Gordon-Salant, S. (**2001**). "Aging and temporal discrimination in auditory sequences," J. Acoust. Soc. Am. **109**, 2955–2963.
Frisina, R. D. (**2001**). "Subcortical neural coding mechanisms for auditory temporal processing," Hear. Res. **158**, 1–35.
Gelfand, S., Schwander, T., and Silman, S. (**1990**). "Acoustic reflex thresholds in normal and cochlear-impaired ears: Effect of no-response rates on 90th percentiles in a large sample," J. Speech Hear Disord. **55**, 198–205.
Getty, D. J. (**1975**). "Discrimination of short temporal intervals: A comparison of two models," Percept. Psychophys. **18**, 1–8.
Gordon-Salant, S., and Fitzgibbons, P. J. (**1993**). "Temporal factors and speech recognition performance in young and elderly listeners," J. Speech Hear. Res. **36**, 1276–1285.
Gordon-Salant, S., and Fitzgibbons, P. J. (**2006**). "Age-related differences in identification and discrimination of temporal cues in speech segments," J. Acoust. Soc. Am. **119**, 2455–2466.
Grose, J. H., Hall, J. W., and Buss, E. (**2001**). "Gap duration discrimination in listeners with cochlear hearing loss: effects of gap and marker duration, frequency separation, and mode of presentation," J. Assoc. Res. Otolaryngol. **2**, 388–398.
Grose, J. H., Hall, J. W., and Buss, E. (**2006**). "Temporal processing deficits in the presenescent auditory system," J. Acoust. Soc. Am. **119**, 2305–2315.
He, N.-J., Horwitz, A. R., Dubno, J. R., and Mills, J. H. (**1999**). "Psychometric functions for gap detection in noise measured from young and aged subjects," J. Acoust. Soc. Am. **106**, 966–978.
Hellstrom, L. L., and Schmiedt, R. A. (**1990**). "Compound action potential input/output functions in young and quiet-aged gerbils," Hear. Res. **50**, 163–174.
Hirsh, I. J., Monahan, C. B., Grant, K. W., and Singh, P. G. (**1990**). "Studies in auditory timing: I. Simple patterns," Percept. Psychophys. **47**, 215–226.
Humes, L. E. (**1996**). "Speech understanding in the elderly," J. Am. Acad. Audiol **7**, 161–167.
Levitt, H. (**1971**). "Transformed up-down methods in psychoacoustics," J. Acoust. Soc. Am. **49**, 467–477.
Lister, J., Besing, J., and Koehnke, J. (**2002**). "Effects of age and frequency disparity on gap discrimination," J. Acoust. Soc. Am. **111**, 2793–2800.
Lister, J., and Tarver, K. (**2004**). "Effects of age on silent gap discrimination in synthetic speech stimuli," J. Speech Lang. Hear. Res. **47**, 257–268.
Lutman, M. E. (**1991**). "Degradations in frequency and temporal resolution with age and their impact on speech identification," Acta Oto-Laryngol., Suppl. **476**, 120–126.
Muchnik, C., Hildescheimer, M., Rubinstein, M., Sadeh, M., Shegter, Y., and Shibolet, B. (**1985**). "Minimal time interval in auditory temporal resolution," J. Aud Res. **25**, 239–246.
Penner, M. J. (**1976**). "The effect of marker variability on the discrimination of temporal intervals," Percept. Psychophys. **19**, 466–469.
Pfeiffer, E. (**1977**). "A short portable mental status questionnaire for the assessment of organic brain deficit in elderly patients," J. Am. Geriatr. Soc. **23**, 433–441.
Pichora-Fuller, M. K., Schneider, B. A., Benson, N. J., Hamstra, S. J., and Storzer, E. (**2006**). "Effect of age on detection of gaps in speech and nonspeech markers varying in duration and spectral complexity," J. Acoust. Soc. Am. **119**, 1143–1154.
Rammsayer, T. H., and Lima, S. D. (**1991**). "Duration discrimination of filled and empty auditory intervals: cognitive and perceptual factors," Percept. Psychophys. **50**, 565–574.
Robinette, M. S., and Glattke, T. J. (**2007**). *Otoacoustic Emissions: Clinical Applications* (Thieme, New York).
Schneider, B. A., Pichora-Fuller, M. K., Kowalchuk, D., and Lamb, M. (**1994**). "Gap detection and the precedence effect in young and old adults," J. Acoust. Soc. Am. **95**, 980–991.
Schneider, B. A., and Hamstra, S. J. (**1999**). "Gap detection thresholds as a function of tonal duration for younger and older listeners," J. Acoust. Soc. Am. **106**, 371–380.
Schneider, B. A., and Pichora-Fuller, M. K. (**2000**). "Implications of perceptual deterioration for cognitive aging research," in *The Handbook of Aging and Cognition*, 2nd ed., edited by F. I. Craik and T. A. Salthouse (Erlbaum, Mahwah, NJ), pp. 155–219.
Small, A. M., and Campbell, R. A. (**1962**). "Temporal differential sensitivity for auditory stimuli," Am. J. Psychol. **75**, 401–410.
Snell, K. B. (**1997**). "Age-related changes in temporal gap detection," J. Acoust. Soc. Am. **101**, 2214–2220.
Strouse, A., Ashmead, D. H., Ohde, R. N., and Grantham, D. W. (**1998**). "Temporal processing in the aging auditory system," J. Acoust. Soc. Am. **104**, 2385–2399.
Vaughan, N., and Letowski, T. (**1997**). "Effects of age, speech rate, and type of test on temporal auditory processing," J. Speech Lang. Hear. Res. **40**, 1192–1200.
Walton, J. P., Frisina, R. D., and O'Neill, W. E. (**1998**). "Age-related alteration in processing of temporal sound features in the auditory midbrain of the CBA mouse," J. Neurosci. **18**, 2764–2776.
Willott, J. F. (**1990**). *Aging and the Auditory System* (Singular, San Diego).
Wingfield, A., Poon, L. W., Lombardi, L., and Lowe, D. (**1985**). "Speed of processing normal aging: effects of speech rate, linguistic structure, and processing time," J. Gerontol. **40**, 579–585.

Frequency modulation detection: Effects of age, psychophysical method, and modulation waveform

Ning-ji He,[a] John H. Mills, and Judy R. Dubno
Department of Otolaryngology-Head and Neck Surgery, Medical University of South Carolina, 135 Rutledge Avenue, P.O. Box 250550, Charleston, South Carolina 29425

(Received 18 September 2006; revised 23 April 2007; accepted 25 April 2007)

As part of an ongoing study of auditory aging, detection of sinusoidal and quasitrapezoidal frequency modulation (FM) was measured with a 5-Hz modulation frequency and 500- and 4000-Hz carriers in two experiments. In Experiment 1, psychometric functions for FM detection were measured with several modulation waveform time patterns in younger adults with normal hearing. Detection of a three-cycle modulated signal improved when its duration was extended by a preceding unmodulated cycle, an effect similar to adding a modulated cycle. In Experiment 2, FM detection was measured for younger and older adults with normal hearing using two psychophysical methods. Similar to frequency discrimination, FM detection was poorer in older than younger subjects and age-related differences were larger at 500 Hz than at 4000 Hz, suggesting that FM detection with low modulation frequencies and frequency discrimination may share common underlying mechanisms. One mechanism is likely related to temporal information coded by neural phase locking which is strong at low frequencies and decreases with increasing frequency, as observed in animals. The frequency-dependent aging effect suggests that this temporal mechanism may be affected by age. The effect of psychophysical method was sizable and frequency dependent, whereas the effect of modulation waveform was minimal. © *2007 Acoustical Society of America.* [DOI: 10.1121/1.2741208]

PACS number(s): 43.66.Mk, 43.66.Sr, 43.66.Fe [JHG] Pages: 467–477

I. INTRODUCTION

Frequency discrimination and frequency modulation (FM) detection are signal paradigms commonly used to measure the smallest detectable frequency change, i.e., the frequency difference limens. Substantial variability has been observed both within and across studies measuring frequency difference limens, in contrast to intensity difference limens (see Jesteadt and Bilger, 1974; Jesteadt and Sims, 1975, for reviews). This variability reflects the influence of subjects' hearing sensitivity, ability to perform the task, and age, as well as experimental methods.

It is well documented that sensorineural hearing loss can lead to deficits in frequency discrimination (Turner and Nelson, 1982; Simon and Yund, 1993) and FM detection (Moore and Skrodzka, 2002; Buss *et al.*, 2004). Reduced ability to detect FM is also correlated with poorer speech recognition observed in some subjects with hearing loss (Buss *et al.*, 2004). This may reflect an association between the ability to use fine temporal information that is coded by neural phase locking and the ability to recognize speech. In a previous study (He *et al.*, 1998), frequency discrimination was measured for younger and older subjects with normal hearing. A frequency-dependent aging effect was observed, such that age-related differences in frequency discrimination were larger at lower frequencies than at higher frequencies. This trend could not be explained in terms of hearing sensitivity because thresholds of younger and older subjects were matched more closely at lower frequencies than at higher frequencies. We hypothesized that frequency discrimination at low frequencies was primarily determined by temporal cues which may be degraded with increasing age.

Limited data are available regarding the effect of age on FM detection at lower and higher frequencies. In an electrophysiological study, Boettcher *et al.* (2002) measured frequency-modulation following responses (FMFR) in younger and older subjects with normal hearing. Measured with a carrier frequency of 500 Hz, modulation frequencies of 4–38 Hz, and varying modulation depths (0–80%), the effect of age on the FMFR amplitude was dependent on the modulation rate and modulation depth. At a low modulation rate (4 Hz), response amplitudes were smaller for older than for younger subjects, whereas at higher modulation rates and modulation depths from 40% to 70%, response amplitudes were larger for older than for younger subjects. For older subjects, larger FMFR amplitudes were associated with smaller amplitudes of auditory brainstem responses. Because behavioral measures of FM detection were not measured in these subjects, the association between larger FMFR amplitudes and age-related declines in FM detection remains unknown.

Hartmann (1997) emphasized that the comparison between FM detection and frequency discrimination should be made only at very low modulation rates ($\leq$6 Hz), because underlying mechanisms involved in FM detection vary with modulation frequency. Several mechanisms have been proposed for frequency difference limens. Some were based on excitation patterns and the output of auditory filters and were referred to as the place mechanism (Zwicker, 1956, 1970; Siebert, 1968). Others were based on the temporal pattern

[a]Electronic mail: hening@musc.edu

(phase locking) of neural responses observed in laboratory animals (Rose *et al.*, 1971; Palm and Russell, 1986) and therefore were referred to as the temporal mechanism (Goldstein and Srulovicz, 1977; Wakefield and Nelson, 1985). In a series of studies, Sek and Moore (Sek and Moore, 1995; Moore and Sek, 1996) established that both place and temporal mechanisms played a role in FM detection, as in frequency discrimination, but these mechanisms were dependent on modulation and carrier frequency. Frequency discrimination was primarily determined by the temporal mechanism for frequencies lower than 4000 Hz and by the place mechanism for frequencies higher than 4000 Hz, where phase locking information was diminished. Similarly, the temporal mechanism was thought to be involved in FM detection but only for carrier frequencies below 4000 Hz and for very low modulation rates ($\leqslant$5 Hz). The place mechanism dominated FM detection for higher carrier frequencies, as well as for lower carrier frequencies when modulation rates were approximately 10 Hz or higher.

Moore and Sek (1995, 1996) further pointed out that the temporal mechanism involved in FM detection was inefficient and that FM detection was dependent on the time spent at the frequency extremes of the stimulus. To test this hypothesis, Sek and Moore (2000) measured detection of quasitrapezoidal FM and amplitude modulation. Each cycle of the quasitrapezoidal modulation waveform consisted of two steady-state time (T_{ss}) portions at the frequency or amplitude extremes and sinusoidal rise/fall transition portions. They found that, for modulation rates $<$10 Hz, FM detection improved when T_{ss} was increased beyond 0 ms (sinusoidal modulation), consistent with the assumption that increased time at the frequency extremes in each modulation cycle benefited FM detection. The improvement in FM detection as a function of T_{ss} was observable for a lower (1000 Hz) but not higher (6000 Hz) frequency carrier. This frequency-dependent effect was interpreted as support for the hypothesis that FM detection at a low modulation rate was, at least partially, mediated by temporal cues.

Psychophysical methods may also affect frequency difference limens. Method-related differences in frequency discrimination were systematically studied by Jesteadt and Bilger (1974), who reviewed the performance ratios for procedures involving one observation interval (such as a yes–no task) and two observation intervals [such as a two-interval forced choice (2IFC) task]. Signal detection theory predicts that thresholds measured in a 2IFC method will be better than thresholds measured in a yes–no method by a factor of $\sqrt{2}$ based on the standard deviations in the decision-making process associated with one-interval and two-interval procedures.[1] However, as observed by Jesteadt and Bilger (1974) for frequency discrimination, the performance level ratio between 2IFC and yes–no methods was greater than the $\sqrt{2}$, and was about 2. Jesteadt and Sims (1975) reported that the performance level ratio between 2IFC and yes–no methods was smaller for FM detection than for frequency discrimination (1.11 vs 2.65 for 1000 Hz). Note that these performance level ratios were not calculated from thresholds but from the slopes of psychometric functions relating d' to Δf (Hz). Therefore, if the steepness of the psychometric function for frequency discrimination and FM detection differs dramatically, the ratios are likely to be affected. In addition, an even larger ratio is expected when the comparison is made between yes–no and 3IFC methods.

The purpose of this study was to assess factors that underlie FM detection, such as age, modulation waveform, and psychophysical method. Among them, the effect of age on FM detection was the main focus, given that it is largely unknown. Comparing the current study of effects of age on FM detection to the previous study on frequency discrimination (He *et al.*, 1998), a sizable difference between results for the two signal paradigms was expected. Hartmann and Klein (1980) measured psychometric functions of frequency difference limens in a 2IFC task using both FM detection (modulation rate=4 Hz) and frequency discrimination paradigms at carrier frequencies of 1000 Hz and lower. Thresholds for sinusoidal FM detection were higher than for frequency discrimination and the ratio between the two measures was about 1.8. Studies examining the relationship between the two signal paradigms with a wider range of reference frequencies (Fastl, 1978; Sek and Moore, 1995) reported that FM detection thresholds and frequency discrimination thresholds changed in different ways as a function of the reference frequency, f_r (standard frequency or carrier frequency). For f_r below 4000 Hz, FM detection thresholds, expressed in $\Delta f/f_r$, varied little with increasing f_r, whereas thresholds for frequency discrimination were much lower than those of FM detection. At f_r of 4000 Hz and greater, the differences were dramatically reduced, with FM detection thresholds actually lower than frequency discrimination thresholds. Taken together, these results suggest that there are differences in the underlying mechanisms between FM detection and frequency discrimination.

Two experiments were conducted in this study. To provide a better understanding of the role of modulation waveforms in FM detection, detection of sinusoidal (T_{ss}=0 ms) and quasitrapezoidal (T_{ss}=80 ms) frequency modulation was measured at a low modulation rate (5 Hz). In Experiment 1, psychometric functions for FM detection from younger adults with normal hearing were measured while varying the time patterns of the modulation waveforms to determine the appropriate stimulus duration for Experiment 2.

In Experiment 2, FM detection thresholds were measured for younger and older adults with normal hearing. Based on the frequency-dependent age effect on frequency discrimination reported by He *et al.* (1998), and the effects of modulation waveform reported by Sek and Moore (2000), a frequency-dependent interaction between aging and modulation waveform (i.e., changing T_{ss}) was predicted. That is, relatively less improvement in FM detection with increasing T_{ss} was expected at 500 than at 4000 Hz for older subjects. FM detection thresholds were obtained with two psychophysical methods: (1) a method of constant stimuli with a one-interval, yes/no task, and (2) a three-interval, three-alternative forced-choice (3I, 3AFC) method. FM detection thresholds were predicted to be higher using the yes–no method than the 3AFC method for all subjects, especially for the lower carrier frequency.

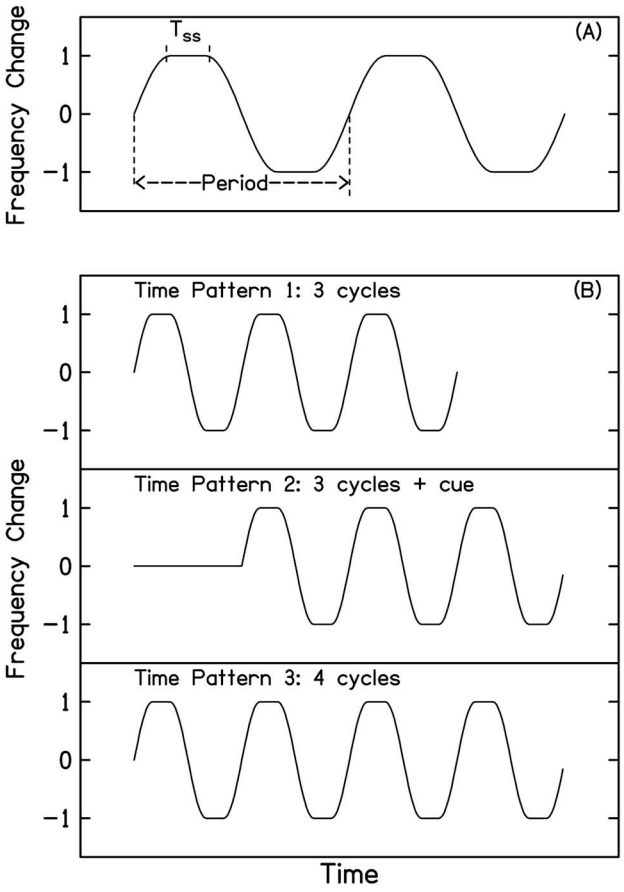


FIG. 1. Frequency modulation patterns, $Z(t)$. (A) A schematic of a quasi-trapezoidal FM waveform pattern when $T_{ss}>0$ ms. When $T_{ss}=0$ ms, the modulation pattern is a sine wave. (B) Three time patterns of FM waveforms used in Experiment 1.

II. GENERAL METHODS

A. Frequency-modulated signals

FM signals were generated using real-time LABVIEW software (National Instruments) developed in this lab and converted to analog using a 16-bit digital-to-analog converter (National Instruments, 6052E) with a sampling rate of 50 000 Hz. Carrier frequencies were 500 and 4000 Hz. The modulation frequency (f_m) was always 5 Hz, which resulted in a modulation cycle period of 200 ms. The time waveform of an FM stimulus was mathematically described as (Sek and Moore, 2000, Eq. 3)

$$x(t) = \sin 2\pi\left(f_c t + \int_0^t \Delta F(\lambda) d\lambda\right), \quad (1)$$

where f_c is the carrier frequency and ΔF is the instantaneous frequency change over time with a dummy variable λ. $\Delta F = \Delta f\, Z(t)$, where Δf is the peak frequency deviation from the carrier frequency and $Z(t)$ describes the time pattern of the modulation waveform. Figure 1(A) illustrates $Z(t)$ for a quasitrapezoidal FM waveform [Fig. 1(B) will be discussed in Sec. III]. Within each modulation period (P), the frequency stays constant at the extremes (1 or −1) over the steady-state time (T_{ss}). The transition time between the extremes has the form of a half-cycle cosine function. When $T_{ss}=0$ ms, the $Z(t)$ is simply a sinusoidal waveform with a period of P.

Unless otherwise specified, all signals were 600 ms in duration (i.e., three modulation cycles) with a cosine-square rise/fall time of 30 ms and presented at 75 dB SPL.[2] Two T_{ss} values were used, 0 and 80 ms, which resulted in a sinusoidal and a quasitrapezoidal modulation pattern, respectively. The generated signals were low-pass filtered (TDT PF1) with a cut-off frequency of 10 000 Hz and attenuated (TDT PA4). The stimuli were then sent through a headphone buffer (TDT HB5) to an insert earphone (Etymotic Research, ER2). The frequency response of the earphone was flat from 100 to 7000 Hz as measured in an ear simulator (B&K 4157).

B. Psychophysical methods

Two psychophysical methods were used to measure FM detection, a constant stimuli, one-interval, yes–no method and a 3I, 3AFC method. A one-interval maximum-likelihood yes–no method was used to measure absolute thresholds for pure-tone signals.

1. Constant stimuli/yes–no method

The method of constant stimuli was used with a yes/no task to measure psychometric functions, with no correct-answer feedback given to the subjects. For each psychometric function, 11 evenly spaced peak frequency deviation (Δf) values in hertz were included. The minimum Δf was always 0 Hz and the maximum varied depending on the carrier frequency and subject. Each Δf value was presented 50 times (Experiment 1) or 20 times (Experiment 2) in random order. Thus, each psychometric function was the result of 550 or 220 presentation trials. To minimize fatigue, the total number of trials was divided into several smaller blocks. Each block started with an orientation trial, during which a pair of tones was repeatedly presented every 2 s with no modulation ($\Delta f = 0$ Hz) in the first tone and the maximum modulation in the second tone. If the subject reported that modulation was not detected for the signal with the maximum Δf, the modulation Δf was increased and additional orientation trials were conducted. The orientation trial was terminated by the subject by pushing a button on a votebox (TDT RBOX) to initiate data collection. For data collection, a single signal was presented (single interval) and subjects were instructed to press the "Yes" button on the votebox if they detected frequency modulation or the "No" button if they did not.

Each measured psychometric function was then fitted minimizing rms errors with a logistic function adapted from Green (1993):

$$P(\text{yes}) = \alpha + (1-\alpha)/(1+e^{-k(X-m)}), \quad (2)$$

where P(yes) is the probability of a "yes" response for a given Δf, α is the false alarm rate, which is the probability of a "yes" response given for the minimum Δf, k is a dimensionless slope factor for the psychometric function, and X and m are Δf's (in hertz) with m corresponding to the midpoint of the psychometric function and X as the independent variable. In order for the slope (k) of the psychometric functions to be relatively uniform across conditions, both X and m were transformed into natural logarithmic units. Finally,

the estimated parameters of m, k, and α were used to calculate a threshold of $d'=1$ for each psychometric function (Swets, 1964, Table I in the Appendix).

2. Three-interval, three-alternative forced-choice method

Within each trial, three signals separated by 500 ms were presented sequentially and only one signal was frequency modulated; the position of the modulated signal within the sequence (1, 2, 3) was randomized. The subject's task was to press one of three buttons on the response box corresponding to the position of the modulated signal. Feedback was given after each response. The Δf was reduced after two consecutive correct responses and increased after one incorrect response, resulting in a threshold estimate corresponding to the 70.7% point of the psychometric function (Levitt, 1971) and a d' value of 1.26 (Hacker and Ratcliff, 1979). The initial step size was 2 Hz for $f_c=500$ Hz and 6 Hz for $f_c=4000$ Hz. After three reversals, the step size was reduced to 0.25 Hz for 500 Hz and 1 Hz for 4000 Hz. The measurement stopped after eight additional reversals and the average of the last four reversals was taken as the threshold. The final result was the mean of two threshold measurements. Note that in this experiment FM detection was measured first with the yes–no method and then with the 3AFC method. After completing measurements of four psychometric functions, subjects were well trained. Therefore, the two measurements of 3AFC thresholds were reasonably stable with associated standard deviations less than 10% of the means, averaging 0.056 (in % of carrier frequency) for older subjects and 0.032 for younger subjects.

3. Measurement of absolute thresholds for pure tones

A one-interval yes–no maximum-likelihood method without feedback (Green, 1993) was used to measure absolute thresholds for pure tones at frequencies of 250, 500, 1000, 2000, 3000, 4000, 6000, and 8000 Hz. The signal duration was 400 ms with 30-ms cosine-square rise/fall times. The slope factor (k) was 0.5 according to Green (1993). Each threshold was measured with 24 trials including 5 catch trials. During each catch trial, the minimum stimulus level was presented to more accurately estimate the false alarm rate (α). The minimum level was adjusted according to experimental conditions. The range of stimulus levels used was kept constant at 50 dB; minimum step size was 0.5 dB. Threshold was defined as the "sweet point" (Green, 1993) which was calculated based on the estimated m (the midpoint of the psychometric function) and α (the false alarm rate) after 24 trials.

C. Subjects

A total of twelve younger subjects (mean age 19.9 years) and eight older subjects (mean age 70.8 years) were tested; four younger subjects were tested in Experiment 1 and the remaining subjects were tested in Experiment 2. All subjects had absolute thresholds of 20 dB HL (ANSI, 1996) or better up to 4000 Hz. Figure 2 shows pure-tone thresholds in quiet

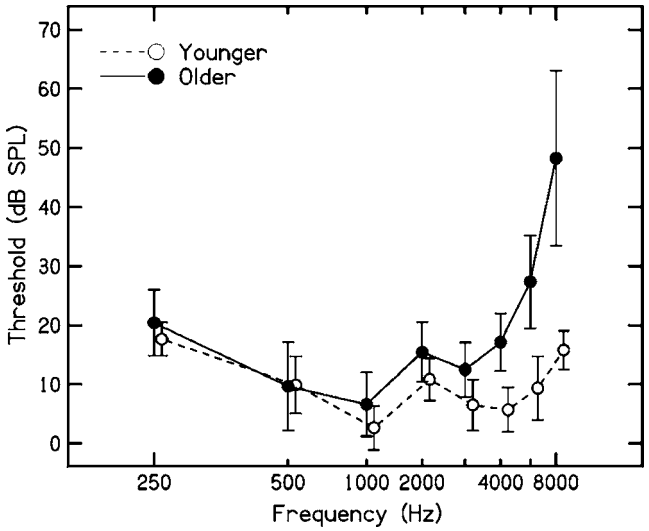


FIG. 2. Group means and standard deviations of absolute thresholds for pure tones for younger and older subjects.

for the eight younger subjects (open symbols and dotted lines) and eight older subjects (closed symbols and solid lines) who participated in Experiment 2. Differences in thresholds between younger and older subjects were small at lower frequencies, but increased as frequency increased above 3000 Hz. For the two carrier frequencies used in Experiment 2, thresholds for younger and older subjects were not significantly different at 500 Hz $[t(14)=0.07, p=0.94]$ but differed significantly at 4000 Hz $[t(14)=-5.24, p<0.001]$.

Each subject was trained on the FM detection task prior to data collection. During practice, a psychometric function with a small number (5) of repetitions was measured repeatedly until the function was stable with a reasonable sigmoidal shape and low false-alarm rate (<0.3).

III. EXPERIMENT 1: EFFECT OF TIME PATTERNS OF MODULATION WAVEFORMS

A. Rationale and methods

The objective of this experiment was to characterize psychometric functions for FM detection with different modulation waveforms. The waveforms differed either in the steady-state time (T_{ss}) within each modulation cycle or in the total duration. Effects of stimulus duration have been reported (Hartmann and Klein, 1980) such that FM detection improved as the number of modulation cycles (or duration) increased up to four cycles. Additional increases in the number of cycles generated no significant differences in FM detection. That is, the four cycles represented the breakpoint of the function that related FM detection and duration. However, it was not clear whether the improvement in FM detection was due to the increased number of cycles or to the increased duration. To answer this question, FM detection was measured with modulation patterns that were manipulated in the time domain as shown in Fig. 1(B). Time pattern 1 was 600 ms in duration and consisted of three modulation cycles. Time pattern 2 had three modulation cycles plus a 200-ms leading unmodulated portion ("cue"), so that the to-

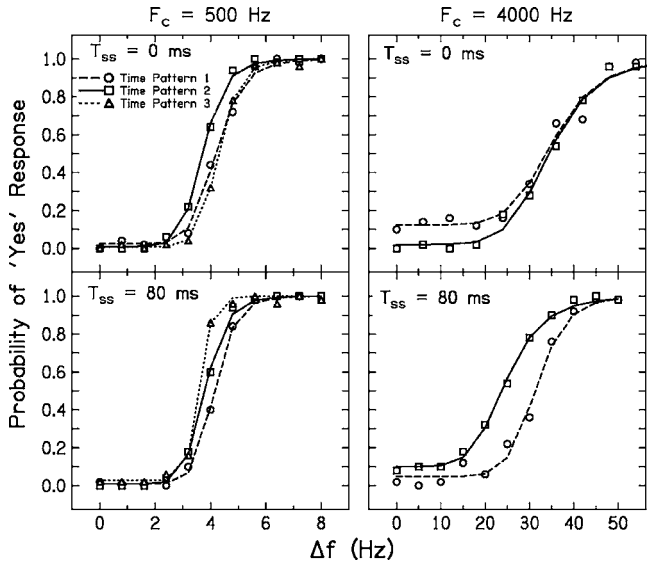


FIG. 3. Psychometric functions of FM detection measured from a younger normal-hearing subject with signals described in Fig. 1(B). Data for 500-Hz and 4000-Hz carrier frequencies are presented in the left and right columns, respectively. The top and bottom rows present data for T_{ss}=0 ms and T_{ss} =80 ms, respectively.

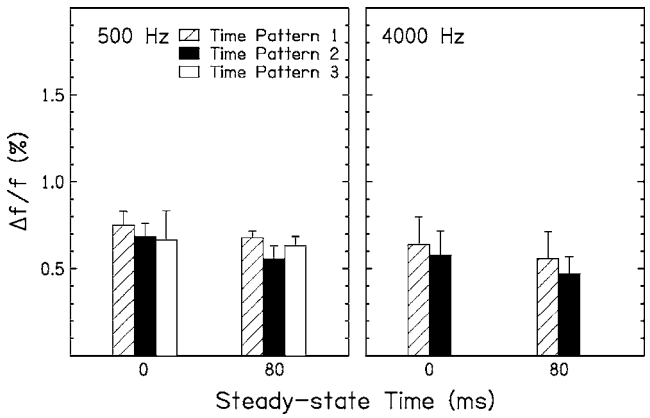


FIG. 4. Group means and standard deviations of FM detection thresholds (d'=1) for younger subjects with normal hearing from Experiment 1. Thresholds measured for time pattern 1 and 2 are presented by hatched and closed bars, respectively. FM detection thresholds for the four-cycle signal (time pattern 3) were collected only for the 500-Hz carrier frequency (open bars).

tal duration was 800 ms. Time pattern 3 was also 800 ms in duration but had four complete modulation cycles. Thus, FM detection can be compared for signals with the same number of modulation cycles but different durations (time pattern 1 vs time pattern 2) and for signals with different numbers of modulation cycles but the same duration (time pattern 2 vs time pattern 3).

Psychometric functions for time patterns 1 and 2 were measured first. Because no consistent improvement in FM detection was observed with the addition of a cue (time pattern 2) at 4000 Hz, FM detection with the four-cycle modulated stimulus (time pattern 3) was measured only for the 500-Hz carrier.

B. Results and discussion

Figure 3 shows typical results for one younger subject for measured (symbols) and fitted (lines) psychometric functions for FM detection for carrier frequencies of 500 Hz (left panels) and 4000 Hz (right panels). Functions for T_{ss}=0 ms are in the top panels and functions for T_{ss}=80 ms are in the bottom panels. The parameter is time pattern of the modulation waveform (time patterns 1, 2, or 3). With Δf transformed into natural logarithmic units, the measured psychometric functions can be precisely fit by a logistic function [Eq. (2)] with a steep slope and a relatively low false alarm rate (α), especially for the 500-Hz carrier. A steeper slope and lower false alarm rate indicate less variance for FM detection.

Slopes (k) for time patterns 1 and 2 were analyzed by a repeated measures ANOVA with T_{ss} (0 vs 80 ms), carrier frequency (500 vs 4000 Hz), and time patterns (1 vs 2) as repeated measures. On average, psychometric function slopes were generally steeper for T_{ss}=80 ms than for T_{ss} =0 ms (k=9.54 vs k=7.10) [$F(1,3)$=10.190, p=0.050]. Other effects and interactions were also not statistically significant. The overall mean slope (k) was 9.06 (s.d.=3.02).

As shown in Fig. 3, FM detection of a three-cycle modulated tone (time pattern 1, circles) was improved when the signal duration was increased by a preceding 200-ms unmodulated cue (time pattern 2, squares), as demonstrated by the leftward shift of the functions for time pattern 2, except at 4000 Hz for T_{ss}=0 ms. Figure 4 shows the mean (+1 s.d.) FM detection thresholds (d'=1) for two values of T_{ss} for the 500-Hz carrier (time patterns 1, 2, and 3; left panel) and the 4000-Hz carrier (time patterns 1 and 2; right panel). Thresholds were generally lower for T_{ss}=80 ms than for T_{ss}=0 ms, and thresholds for time pattern 1 were always higher than those for time patterns 2 and 3. This was confirmed by a repeated-measures ANOVA, showing a significant effect of T_{ss} (0 vs 80 ms) [$F(1,3)$=101.51, p=0.0021] and a significant effect of time patterns (pattern 1 vs pattern 2) [$F(1,3)$ =62.25, p=0.0042]. However, the effect of the carrier frequency (500 vs 4000 Hz) was not significant, nor were any interactions. In addition, thresholds for time patterns 2 and 3 for the 500-Hz carrier (left panel) were analyzed by a paired two-sample t-test for means, which showed that there were no significant differences in thresholds between time pattern 2 (three cycles, 800 ms, closed bars) and time pattern 3 (four cycles, 800 ms, open bars), suggesting that the addition of a cycle-long unmodulated portion (cue) was equivalent to adding a modulated cycle.

For most subjects, FM detection thresholds were lower for T_{ss}=80 ms than for T_{ss}=0 ms, except for the four-cycle signal (time pattern 3). The ratio of the differences between FM thresholds for the steady-state times (T_{ss}=0 ms vs T_{ss} =80 ms) averaged across frequency was 1.18, close to the ratio between sinusoidal and trapezoidal waveforms (1.16) calculated from data of Hartmann (1997, Fig. 19.8). As shown in Fig. 4, the T_{ss} effect was observable for both low and high carrier frequencies with similar threshold difference ratios, i.e., 1.10 and 1.14 for time pattern 1 for 500 and 4000 Hz, respectively, and 1.23 for time pattern 2 for both frequencies. This observation was not consistent with Sek

and Moore (2000), who reported that increasing T_{ss} improved FM detection at the lower frequency (1000 Hz) but not at the higher frequency (6000 Hz). However, the higher carrier frequency used in the present study (4000 Hz) was lower than that used in the study of Sek and Moore (6000 Hz); at 4000 Hz, temporal cues from phase locking may not be entirely eliminated, which may explain the difference in results for the two studies. This issue will be addressed further in Sec. V.

For the four-cycle signal (time pattern 3), the thresholds for the two T_{ss} conditions were nearly equal, that is, increasing the steady-state time did not affect FM detection for the four-cycle signal. Hartmann and Klein (1980) reported that no significant improvement in FM detection was observed as the number of modulation cycles increased from four cycles. These results were confirmed in the current study, such that, at the breakpoint, no significant improvement in FM detection thresholds was observed with additional information regarding frequency extremes introduced by extending the steady-state time. Note that the four-cycle signals corresponded to different durations in the two studies (800 vs 1000 ms) due to the use of different modulation frequencies (5 and 4 Hz). Because results for the four-cycle signals showed little effect of T_{ss}, the three-cycle signal (time pattern 1) was used in Experiment 2.

IV. EXPERIMENT 2: EFFECTS OF AGE AND PSYCHOPHYSICAL METHOD

A. Methods

FM detection was measured for eight younger and eight older subjects with normal hearing for two carrier frequencies (500 and 4000 Hz) and two steady-state times (T_{ss}=0 and 80 ms) using two psychophysical methods: constant stimuli, one-interval, yes–no task and 3I, 3AFC. For each subject, FM detection was measured first with the yes–no method and then with the 3AFC method. As in Experiment 1, subjects were trained on the FM detection task. Additional practice was given before data collection with the 3AFC task to assure that subjects understood the procedure. Because subjects were well trained in detecting frequency modulation following data collection with the yes–no method, this practice session was usually brief. The test order may have contributed to differences in FM detection for the two psychophysical methods, which will be discussed later.

B. Results

1. Psychometric functions and thresholds (constant stimuli, yes–no method)

For measuring each psychometric function for FM detection, eleven evenly spaced peak frequency deviation (Δf) values in hertz were included and repeated twenty times; the results were then fitted with a logistic function [Eq. (2)]. Figure 5 shows individual fitted psychometric functions as a function of Δf. The dotted lines represent data for eight younger subjects and solid lines for eight older subjects. For the 500-Hz carrier frequency, within-group variances were smaller for the younger subjects than the older subjects, and functions were well separated between the two age groups,

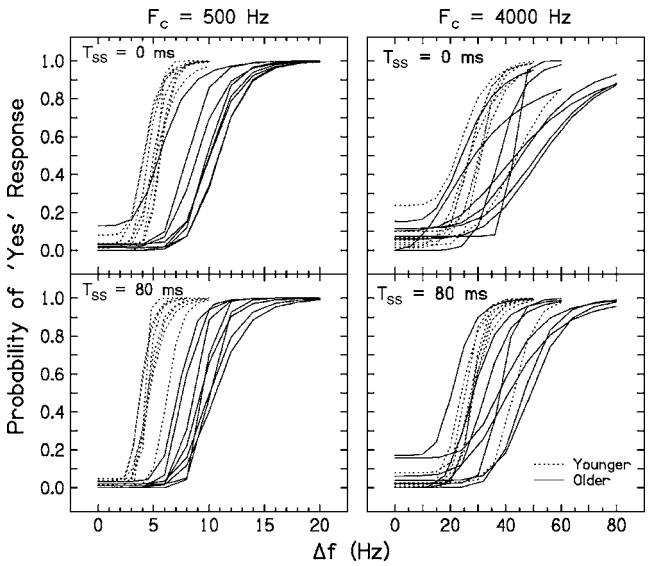


FIG. 5. Individual psychometric functions for younger (n=8) and older (n=8) subjects measured using a constant stimuli, one-interval, yes–no method. Results for 500- and 4000-Hz carriers are presented in the left and right columns, respectively. The top and bottom rows present results for T_{ss}=0 ms and T_{ss}=80 ms, respectively.

especially for T_{ss}=80 ms. For T_{ss}=0 ms, one psychometric function of an older subject fell within the range of the younger subjects; however, the false alarm rate was much higher than for the younger subjects (α=0.12). For the 4000-Hz carrier frequency (right panel), functions for the younger and older subjects showed increased intersubject variability and essentially overlapped, and they showed increased false alarm rates. The false alarm rate tended to be smaller for T_{ss}=80 ms than for T_{ss}=0 ms, especially for younger subjects and for the 4000-Hz carrier, suggesting that FM detection was improved by increasing the time at the frequency extremes.

Similar to the results of Experiment 1, a feature of the psychometric functions for FM detection was the relatively steep slope. The slope of the psychometric function was slightly steeper at 500 than at 4000 Hz, with the ratio of k equal to 1.29 for the younger subjects and 1.11 for the older subjects. However, as analyzed with a repeated measures ANOVA, differences in slope were not statistically significant between low and high carrier frequencies or between age groups, but were significant between T_{ss}=0 ms and T_{ss}=80 ms [$F(1,14)=6.38$, $p=0.024$]. This is consistent with the results of Experiment 1, suggesting that increasing the time at frequency extremes improved FM detection. Interactions were not statistically significant. Hence, the slope of the psychometric function for FM detection is independent of subject age and carrier frequency, with a mean slope for younger subjects of 9.06 (Experiment 1) or 9.02 (Experiment 2) and a mean slope for older subjects of 8.44 (Experiment 2).

Group means and standard deviations of FM thresholds (d'=1) calculated from the psychometric functions shown in Fig. 5 are presented in Fig. 6 (in percent of carrier frequency). Results for the 500- and 4000-Hz carriers are presented in the left and right panels, respectively. FM detection

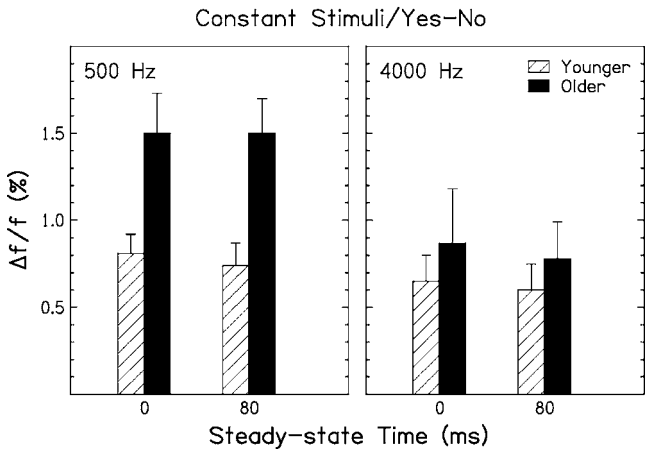


FIG. 6. Group means and standard deviations of FM detection thresholds ($d'=1$), calculated from psychometric functions in Fig. 5, for $T_{ss}=0$ ms and $T_{ss}=80$ ms. Thresholds for 500- and 4000-Hz carriers are presented in the left and right panels, respectively.

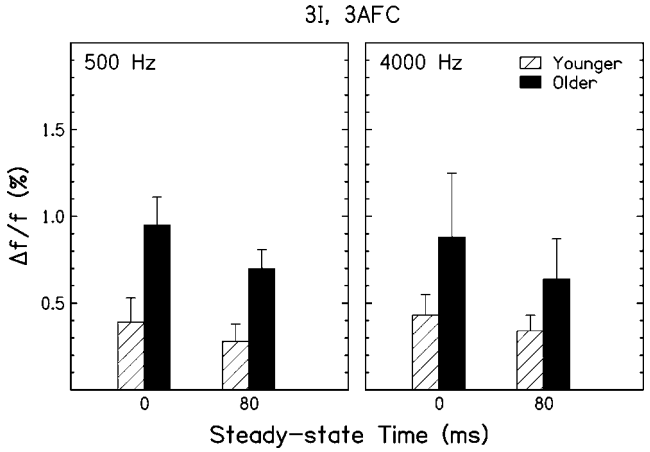


FIG. 7. Group means and standard deviations of FM detection thresholds measured with a 3I, 3AFC method, for $T_{ss}=0$ ms and $T_{ss}=80$ ms. Thresholds for 500- and 4000-Hz carriers are presented in the left and right panels, respectively.

thresholds were analyzed with a repeated measures ANOVA, with age as a grouping factor and T_{ss} (0 vs 80 ms) and carrier frequency, f_c (500 vs 4000 Hz) as repeated measures. The results showed a significant effect of age [$F(1,14)=38.05$, $p<0.0001$] and carrier frequency [$F(1,14)=325.33$, $p<0.0001$] with a significant carrier frequency by age interaction [$F(1,14)=5.73$, $p=0.031$]. The T_{ss} effect was not significant nor were the T_{ss} by age interaction or the T_{ss} by carrier frequency interaction. The significant carrier frequency by age interaction indicates a significant difference in the aging effect between carrier frequencies at 500 and 4000 Hz. As shown by a post hoc test, FM detection thresholds differed significantly between younger and older subjects 500 Hz [$F(1,14)=129.95$, $p<0.0001$] but not at 4000 Hz.

Thus, three main features are apparent in these results. First, FM detection thresholds for older subjects were significantly higher than thresholds for younger subjects at the lower carrier frequency but not at the higher carrier frequency. This is consistent with a previous study of frequency discrimination (He *et al.*, 1998) showing that age-related differences were larger at lower frequencies than at higher frequencies. Second, for both age groups and carrier frequencies, thresholds for the two steady-state times did not differ significantly. Finally, FM detection thresholds were significantly higher for the 500-Hz carrier than for the 4000-Hz carrier, and the frequency-related differences were larger for older than younger subjects. For the 500-Hz carrier, mean $d'=1$ thresholds (averaged across T_{ss}) were 0.77 for younger subjects and 1.51 for older subjects. For the 4000-Hz carrier, the age-related difference was significantly smaller; mean $d'=1$ thresholds (averaged across T_{ss}) were 0.63 and 0.82 for younger and older subjects, respectively.

2. 3I, 3AFC thresholds

Figure 7 shows the mean (+1 s.d.) FM detection thresholds (in % of carrier frequency) obtained using the 3AFC psychophysical method. Data are plotted in the same format as in Fig. 6. Thresholds for younger subjects with $T_{ss}=0$ ms are comparable to thresholds with a 5-Hz modulation rate reported by Sek and Moore (1995).

As expected, thresholds obtained using the 3AFC method were lower than those extracted from psychometric functions generated using a constant stimuli, yes–no method, especially for the 500-Hz carrier. FM detection thresholds using the 3AFC method were analyzed with a repeated-measures ANOVA, using the same factors as the previous ANOVA. The analysis showed a significant effect of age group [$F(1,14)=47.96$, $p<0.0001$], carrier frequency [$F(1,14)=401.25$, $p<0.0001$], and T_{ss} [$F(1,14)=78.56$, $p<0.0001$]; none of the interactions were significant. Thus, FM detection thresholds using the 3AFC method differed from those extracted from the psychometric functions in two respects. First, thresholds for the 3AFC procedure were significantly higher for the 0-ms steady-state time than for the 80-ms steady-state time. Second, although thresholds were significantly higher for older than for younger subjects, age-related differences did not differ significantly for the 500- and 4000-Hz carriers.

C. Discussion

In this experiment, effects of age, psychophysical method, and modulation waveform on FM detection were systematically investigated. In addition to statistical analyses, difference ratios of thresholds reported in the literature (e.g., Hartmann and Klein, 1980; or performance level ratio, Jesteadt and Bilger, 1974) provide a straightforward assessment of these factors and their relative effectiveness. Table I summarizes the threshold difference ratios in FM detection thresholds related to age (older vs younger), psychophysical method (constant stimuli/yes–no vs 3AFC), and modulation waveform ($T_{ss}=0$ ms vs $T_{ss}=80$ ms). A ratio of 1 indicates that the difference between thresholds is zero. As shown in Table I, although the ratios for each factor were dependent on other factors, age had the largest effect with a mean ratio of 1.93, followed by the effects of method (1.74) and modulation waveform (1.21). In addition, the effects of age and

psychophysical method were frequency dependent but the effect of modulation waveform was independent of carrier frequency.

A significant aging effect was observed for both psychophysical methods (Figs. 6 and 7). Although FM detection thresholds were consistently higher for older subjects than younger subjects, the younger/older difference ratio varied depending on carrier frequency and psychophysical method. As shown in Table I, the ratio was larger at 500 than at 4000 Hz, consistent with the frequency discrimination results from He *et al.* (1998) where the age-related ratios averaged 2.57 for 500 Hz and 1.61 for 4000 Hz. The aging effect observed here was also in the same direction as the one that can be inferred from the FMFR electrophysiological results of Boettcher *et al.* (2002) for modulation frequencies $\leqslant$5 Hz.

Thresholds measured with the 3AFC method were sizably lower than those with the constant stimuli/yes–no method and the differences were dependent on the carrier frequency. For younger subjects, the method-related difference ratios were greater than 2 for 500 Hz and less than 2 for 4000 Hz (Table I). As mentioned in Sec. II 3AFC thresholds corresponded to a higher d' value than yes–no thresholds ($d'=1.26$ vs $d'=1$). If threshold comparisons were made at the same d' value ($d'=1.26$), the difference ratios between the two methods would be even larger, because yes–no thresholds would be higher for $d'=1.26$ than for $d'=1$. The method-related difference ratios (Table I) were larger than the performance level ratios reported by Jesteadt and Sims (1975), but consistent with our prediction based on signal detection theory and the literature (see Jesteadt and Bilger, 1974). This method-related difference may be attributed to the use of feedback in the forced choice procedure, which may aid detection. Also, as observed in Experiment 1, FM detection can be improved by extending the stimulus duration with an unmodulated cycle (cue), which provides additional information regarding the carrier frequency, or the resting position for the frequency deviation (see Sec. V for a more detailed explanation).

Furthermore, as demonstrated in Table I, there is an interaction between aging and psychophysical method, such that age-related ratios were larger for the 3AFC method than for the yes–no method. This suggests that younger subjects benefited more than older subjects from the comparison among stimuli available in the 3AFC method. This benefit appeared to be reduced for the 4000-Hz carrier, where the ratios were smaller than those for the 500-Hz carrier, especially for the older subjects. At 4000 Hz, the older subjects had ratios either close to or less than 1, i.e., minimal differences between psychophysical methods. Otherwise, the method-related difference ratios were all greater than $\sqrt{2}$. The overall mean ratio was 2.11 at 500 Hz and 1.38 at 4000 Hz. The ratio for the 500-Hz carrier was quantitatively similar to that reported by Jesteadt and Bilger (1974), who found that the performance level ratio between 2IFC and yes–no methods for frequency discrimination at 1000 Hz was 2.10. The sizable method-related difference may be partly due to a training effect. As mentioned earlier, thresholds using the 3AFC method were obtained after the measurement of four psychometric functions using the yes–no method. However, this training effect cannot account for the frequency-dependent trend of the method-related differences, because a training effect should not differ for lower and higher carrier frequencies.

Compared to the other two factors, the effect of modulation waveform was substantially smaller and varied from measurement to measurement. Although, on average, thresholds for $T_{ss}=80$ ms were lower than those for $T_{ss}=0$ ms, the trend varied among subjects and from experiment to experiment. Therefore, a marginally significant T_{ss} effect was observed in Experiment 1 for thresholds measured using constant stimuli/yes–no, except for four-cycle data, whereas in Experiment 2, the T_{ss} effect was significant for 3AFC but not for constant stimuli/yes–no methods. Furthermore, unlike effects of aging and psychophysical method, no systematic differences in waveform-related ratios were observed between low and high carrier frequencies. This is inconsistent with results of Sek and Moore (2000) showing that increasing T_{ss} improved FM detection only at lower frequencies ($\leqslant$1000 Hz) but not at the higher frequency (6000 Hz). The discrepancy might be attributed to different carrier frequency ranges (4000 vs 6000 Hz) used in the two studies (see the following section).

V. GENERAL DISCUSSION

A. Psychometric functions for FM detection and frequency discrimination

As shown in Figs. 3 and 5, psychometric functions for sinusoidal (top panels) and quasitrapezoidal FM detection (bottom panels), fitted with the same logistic function [Eq. (2)], were generally similar in shape. The functions were characterized by a steep slope ($k\approx 9.0$), and the slope was generally uniform across parameters of carrier frequency and age, and differed only in their relative locations along the x axis (Δf) and/or the false alarm rates. These trends are consistent with those for frequency discrimination (Nelson and Freyman, 1986; He *et al.*, 1998), which also show a uniform

TABLE I. Difference ratios in FM detection thresholds for different factors.

Factor		Carrier Frequency 500 Hz	4000 Hz	
Age (older/younger)				
(yes–no)	$T_{ss}=0$ ms	1.88	1.33	
	$T_{ss}=80$ ms	2.04	1.29	
(3AFC)	$T_{ss}=0$ ms	2.43	2.08	
	$T_{ss}=80$ ms	2.51	1.87	Overall mean=1.93
Method (yes–no/3AFC)				
(Younger)	$T_{ss}=0$ ms	2.07	1.53	
	$T_{ss}=80$ ms	2.63	1.77	
(Older)	$T_{ss}=0$ ms	1.60	0.98	
	$T_{ss}=80$ ms	2.13	1.22	Overall mean=1.74
Modulation waveform ($T_{ss}=0$ ms/$T_{ss}=80$ ms)				
(yes–no)	Younger	1.09	1.08	
	Older	1.01	1.12	
(3AFC)	Younger	1.39	1.25	
	Older	1.34	1.39	Overall mean=1.21

slope (although much more shallow) that was independent of stimulus frequency and subjects' hearing status.

The method used to measure psychometric functions in Experiment 1 was similar to that used to measure psychometric functions for frequency discrimination in a previous study (He *et al.*, 1998). Data collection and analysis were similar for both studies, except that a single-interval task was used in the current study whereas a two-interval same/different task was used in the previous study. A comparison was made between data for $T_{ss}=0$ ms (sinusoidal FM) from Experiment 1 and the psychometric functions for frequency discrimination from the previous study. There were several differences between the results using the two paradigms. First, the slopes were quite different. Psychometric functions for FM detection were much steeper than for frequency discrimination, with a mean k of about 9.0 for FM detection and 3.0 for frequency discrimination. A steeper slope indicates a smaller measurement variance for the psychometric function. Second, false alarm rate (α) was much lower for FM detection, averaging 0.05 for FM detection but 0.26 for frequency discrimination. These differences in the slopes of psychometric functions and false alarm rates suggest that FM detection is a relatively easier task than frequency discrimination, which is consistent with an earlier report of smaller individual variability for FM detection than frequency discrimination (Fastl, 1978). The dramatic differences in slopes between frequency discrimination and FM detection may explain the differences in performance level ratios (2IFC/yes–no) between the two paradigms (Jesteadt and Sims, 1975).

A third difference is that the frequency difference limen measured with the FM detection paradigm was larger than that measured with the frequency discrimination paradigm. The mean $d'=1$ threshold in $\Delta f/f$ (%) was 0.38 for 500 Hz and 0.45 for 4000 Hz in the previous study of frequency discrimination, as compared to 0.72 and 0.60 for the current study of FM detection. The difference ratio between frequency discrimination and FM detection paradigms was about 1.89 at 500 Hz and 1.33 at 4000 Hz. This observation is consistent with the frequency-dependent trend reported by Fastl (1978) and Sek and Moore (1995), and is numerically close to data of Hartmann and Klein (1980) who reported a ratio of 1.8 for carrier frequencies of 1000 Hz and lower.

B. Frequency-dependent aging effect

As shown in Figs. 5–7, FM detection was consistently poorer for older than younger subjects and age-related differences were larger at low frequencies than at high frequencies. This frequency-dependent trend was consistent with that observed in frequency discrimination (He *et al.*, 1998), suggesting some common underlying mechanism for FM detection and frequency discrimination, at least in the aging process. Similar to the previous study (He *et al.*, 1998), differences in absolute thresholds between younger and older subjects were larger at 4000 than at 500 Hz (Fig. 1 in He *et al.*, 1998 and Fig. 2 in the current study) whereas age-related differences in FM detection were larger at 500 than at 4000 Hz. Therefore, the aging effect was not related to a difference in hearing sensitivity, but may relate to an age-related decline in the temporal mechanism involved in FM detection, as described in the following section.

C. Temporal mechanism for FM detection

As shown in Experiment 1, FM detection can be improved by extending the stimulus duration with a preceding unmodulated cycle (cue), an effect equivalent to increasing the duration with a modulated cycle. This suggests that both carrier frequency and frequency deviation are important information for FM detection. That is, information about the carrier frequency (i.e., the resting position of the frequency deviation) may be extracted when listening to FM signals and that information may be used to detect frequency modulation. Adding a preceding cue to the signal may benefit the extraction process and improve FM detection.

This process then requires some mechanism similar to that proposed by the sampling-differencing model (Hartmann and Klein, 1980, 1981), although in the model, the carrier frequency was disregarded and scaled down to zero for mathematical convenience. The basic scheme of the model is that, when listening to FM signals, the auditory system takes a time series of samples and estimates dynamic frequency changes from successive brief windows, most likely based on the pattern of phase locking. When the change exceeds a certain criterion, FM is detected. According to the model, the decision process is a cross correlation of the sampled modulation and a modulation pattern template stored in memory. The template is acquired in part from the experimental context (i.e., stimulus) and in part from a multistage template matching process (Hartmann and Klein, 1980, 1981). The model assumes that at threshold, where frequency deviation is small, modulation waveforms are not distinguishable from each other and the only information available is the periodicity of the modulation. Therefore, integration of the FM information is based on modulation cycles. This assumption is supported by results of the current experiment, wherein FM detection improved when signals were preceded by an unmodulated cycle (cues, time pattern 2, Figs. 3 and 4). In this case, the cue provided information regarding the resting position for the frequency modulation which was then used in the decision process in the following cycles. Furthermore, the observation that adding a cue was more effective at the lower carrier frequency than at the higher carrier frequency (Figs. 3 and 4) is consistent with the notion that extracting the carrier frequency information relies on the temporal information provided by phase locking of neural responses. While our observations support the temporal sampling model for FM detection (Hartman and Klein, 1980, 1981), a new challenge is placed on the model to account for the carrier frequency as providing information on resting position for the frequency modulation.

The importance of carrier frequency information in FM detection may provide an additional explanation for the large difference ratios between constant stimuli/yes–no and 3AFC thresholds, the frequency dependency, as well as the interaction between aging and frequency (Experiment 2). Other than different decision processes that may be involved in the

two psychophysical methods, more opportunities to listen to the carrier frequency, as in the case of the 3AFC method, may also be beneficial. In the 3AFC method, subjects had twice as many opportunities to listen to the carrier frequency than to listen to the modulated signal. In the one-interval yes–no method, the carrier frequency was not presented and information about the carrier frequency had to be extracted from the modulated cycles. Furthermore, as demonstrated in Table I, method-related difference ratios were larger for the 500-Hz carrier than for the 4000-Hz carrier, consistent again with the notion that extracting the carrier frequency information was carried out in the temporal domain. The observed interaction between aging and method differences suggested that both effects may well be related to the temporal mechanism described earlier.

The results of the current study did not, however, confirm the initial prediction of an interaction between aging and modulation waveform, which was based on a study by Sek and Moore (2000), showing a frequency-dependent effect of steady-state time. As demonstrated in Figs. 4, 6, and 7, the effect of modulation waveform was relatively small, as compared to effects of aging and psychophysical method. As shown in Table I, the T_{ss} effect was independent of carrier frequency and age. It should be noted that the current study used 4000 Hz as the higher-frequency carrier, instead of 6000 Hz as used by the previous study of Sek and Moore (2000), because pure-tone thresholds for our older subjects were elevated at frequencies $\geqslant$4000 Hz (Fig. 2). This frequency difference precluded a valid comparison of our data with those of the previous study. However, if the frequency-dependent trend for the effect of T_{ss} was mediated by some temporal mechanism (such as phase locking), as suggested by Sek and Moore, then a trend of gradual change should be observable as carrier frequency increased from 500 to 4000 Hz. As demonstrated in Table I, this was not the case and T_{ss}-related ratios for 500 and 4000 Hz differed randomly.

The effect of modulation waveform was also a focus of the sampling-differencing model proposed by Hartmann and Klein (1980) in which all FM waveforms were described by a frequency probability density of frequency deviations. The model was confirmed by experimental data showing that frequency difference limens measured with FM detection were lowest for the square waveform, close to those obtained with frequency discrimination, and increased in an orderly manner as the modulation waveform changed from square to trapezoidal, sine, and triangle. However, when equally detectable, these modulation waveforms had generally equal rms values, suggesting that the auditory system may perform rms averaging of frequency deviations over time during processing of FM signals. According to this scheme, the effect of steady-state time should not be frequency dependent.

The observed differences between sinusoidal and quasitrapezoidal FM detection in the current study support the rms averaging mechanism described by Hartmann and Klein (1980). Under the rms-averaging scheme, the difference ratio in detection thresholds between T_{ss}=0 ms and T_{ss}=80 ms should not be frequency dependent and should be close to 1.15, a ratio between sinusoidal and trapezoidal waveforms at thresholds calculated from data of Hartmann (1997). The current study showed that the mean ratio between sinusoidal and quasitrapezoidal modulation waveforms was 1.21 in Experiment 2 and 1.15 averaged across all conditions in Experiment 1.

In summary, several of the factors that significantly affected FM detection exhibited a consistent frequency-dependent trend, which has been interpreted as an indication of involvement of temporal mechanisms. However, the underlying mechanism for FM detection, as well as for frequency discrimination, is not fully understood, nor is the relationship between the two measurements. Nevertheless, the similar frequency-dependent effect of aging observed in FM detection (the current study) and frequency discrimination (He *et al.*, 1998) suggests that both measurements involve a temporal mechanism which may be affected by age.

VI. CONCLUSIONS

(1) The psychometric function for FM detection measured at a low modulation frequency (5 Hz) was characterized by a steep slope ($k \approx 9$). The slope was generally independent of subject age, carrier frequency, modulation waveform, and stimulus duration, so that psychometric functions obtained under different conditions differed only in their relative locations along the Δf axis and/or false alarm rates.

(2) FM detection was poorer for older than younger subjects with normal hearing. In general, age-related differences in FM detection were larger at lower carrier frequencies than at higher carrier frequencies, a trend consistent with that observed previously in frequency discrimination, suggesting that the two measurements may share a common underlying mechanism. The mechanism is likely related to the temporal pattern of neural responses, i.e., phase locking, which is strong at low frequencies and diminishes with increasing frequency. The consistent effect of aging observed in both FM detection and frequency discrimination further suggests that the underlying temporal mechanism may decline with age.

(3) In addition to subject age, frequency difference limens were affected by psychophysical method and stimulus paradigm. The effects were frequency dependent. As characterized by difference ratios of frequency difference limens, aging had the largest effect, followed by the effects of method (yes–no vs 3AFC) and stimulus paradigm (frequency discrimination vs FM detection).

(4) For carrier frequencies $\leqslant$4000 Hz, differences between sinusoidal and quasitrapezoidal FM detection were relatively small as compared to the age-related and method-related differences. The effect of the modulation waveforms (T_{ss}=0 ms vs T_{ss}=80 ms) was not dependent on carrier frequency, but followed an rms averaging model as described by Hartmann and Klein (1980).

(5) Consistent with observations of Hartmann and Klein (1980), which showed no significant improvement in FM detection for modulation cycles $\geqslant$4, the present study further demonstrated that FM detection with four modulation cycles was not improved with an increase in the steady-state time of the cycle. For modulation cycles

<4, FM detection was improved as steady-state time increased from 0 to 80 ms. In addition, FM detection improved as the duration (in the form of the number of cycles) increased, regardless of whether the increased time section was modulated or unmodulated. This suggests that information regarding carrier frequency is equally important to FM detection as that of frequency modulation. The auditory system may extract information about the carrier frequency, i.e., the resting position of the frequency modulation, and use this information to better detect frequency modulation. Explanation of results from the current study requires a mechanism similar to that proposed by the sampling-differencing model for FM detection (Hartmann and Klein, 1980) with an extension that can account for the importance of carrier frequency information.

ACKNOWLEDGMENTS

This work was supported (in part) by research Grant Nos. P50 DC00422 and R01 DC00184 from NIH/NIDCD and the MUSC General Clinical Research Center (M01 RR 01070). This investigation was conducted in a facility constructed with support from Research Facilities Improvement Program Grant No. C06 RR14516 from the National Center for Research Resources, National Institutes of Health. The authors thank Fu-Shing Lee for advice and assistance with data analysis, and Jayne B. Ahlstrom, Kelly C. Harris, and Stefanie R. Reed for their comments on an earlier draft of this paper. Our gratitude is also given to Associate Editor John Grose and two anonymous reviewers for their constructive comments that greatly improved this paper.

[1]There are other method-related effects, such as stepping rules in an up–down adaptive procedure (Levitt, 1971), which result in different threshold estimates corresponding to different percentage points of a psychometric function. In Jesteadt and Bilger (1974), this effect was controlled by converting the thresholds to d' values (Table I).

[2]The effect of stimulus level was not addressed in this study. However, in a previous study (He *et al.*, 1998), stimulus level (40 vs 80 dB SPL) had only a minimal effect on Δf for both younger and older subjects. This was likely due to the large variance associated with frequency difference limens in general.

American National Standards Institute (ANSI) (**1996**). *Specification for Audiometers*, ANSI S3.6-1996 (ANSI, New York).

Boettcher, F. A., Madhotra, D., Poth, E. A., and Mills, J. H. (**2002**). "The frequency-modulation following response in young and aged human subjects," Hear. Res. **165**, 10–18.

Buss, E., Hall, J. W., and Grose, J. H. (**2004**). "Temporal fine-structure cues to speech and pure tone modulation in observers with sensorineural hearing loss," Ear Hear. **25**, 242–250.

Fastl, H. (**1978**). "Frequency discrimination for pulsed versus modulated tones," J. Acoust. Soc. Am. **63**, 275–277.

Goldstein, J. L., and Srulovicz, P. (**1977**). "Auditory-nerve spike intervals as an adequate basis for aural frequency measurement," in *Psychophysics and Physiology of Hearing*, edited by E. F. Evans and J. P. Wilson (Academic, London), pp. 337–347.

Green, D. M. (**1993**). "A maximum-likelihood method for estimating thresholds in a yes–no task," J. Acoust. Soc. Am. **93**, 2096–2105.

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

Hartmann, W. M. (**1997**). *Signals, Sound, and Sensation* (Springer, New York), pp. 431–448.

Hartmann, W. M., and Klein, M. A. (**1980**). "Theory of frequency modulation detection for low modulation frequencies," J. Acoust. Soc. Am. **67**, 935–946.

Hartmann, W. M., and Klein, M. A. (**1981**). "The effect of uncertainty on the detection of frequency modulation at low modulation rates," Percept. Psychophys. **30**, 417–424.

He, N., Dubno, J. R., and Mills, J. H. (**1998**). "Frequency and intensity discrimination in a maximum-likelihood procedure from young and aged normal-hearing subjects," J. Acoust. Soc. Am. **103**, 553–565.

Jesteadt, W., and Bilger, R. C. (**1974**). "Intensity and frequency discrimination in one-and two-internal paradigms," J. Acoust. Soc. Am. **55**, 1266–1276.

Jesteadt, W., and Sims, S. L. (**1975**). "Decision processes in frequency discrimination," J. Acoust. Soc. Am. **57**, 1161–1168.

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

Moore, B. C. J., and Sek, A. (**1995**). "Effects of carrier frequency, modulation rate and modulation waveform on the detection of modulation and the discrimination of modulation type (amplitude modulation versus frequency modulation)," J. Acoust. Soc. Am. **97**, 2468–2478.

Moore, B. C. J., and Sek, A. (**1996**). "Detection of frequency modulation at low modulation rates: Evidence for a mechanism based on phase locking," J. Acoust. Soc. Am. **100**, 2320–2331.

Moore, B. C. J., and Skrodzka, E. (**2002**). "Detection of frequency modulation by hearing-impaired listeners: Effects of carrier frequency, modulation rate, and added amplitude modulation," J. Acoust. Soc. Am. **111**, 327–335.

Nelson, D. A., and Freyman, R. L. (**1986**). "Psychometric functions for frequency discrimination for listeners with sensorineural hearing loss," J. Acoust. Soc. Am. **79**, 799–805.

Palm, A. R., and Russell, I. J. (**1986**). "Phase-locking in the cochlear nerve of the guinea-pig and its relation to the receptor potential of inner hair-cells," Hear. Res. **24**, 1–15.

Rose, J. E., Hind, J. E., Anderson, D. J., and Brugge, J. F. (**1971**). "Some effects of stimulus intensity on response of auditory nerve fibers in squirrel monkey," J. Neurophysiol. **34**, 685–699.

Sek, A., and Moore, B. C. J. (**1995**). "Frequency discrimination as a function of frequency, measured in several ways," J. Acoust. Soc. Am. **97**, 2479–2486.

Sek, A., and Moore, B. C. J. (**2000**). "Detection of quasi-trapezoidal frequency and amplitude modulation," J. Acoust. Soc. Am. **107**, 1598–1604.

Siebert, V. M. (**1968**). "Stimulus transformations in the peripheral auditory system," in *Recognizing Patterns*, edited by P. A. Kollers and M. Eden (MIT, Cambridge), pp. 104–133.

Simon, H. J., and Yund, E. W. (**1993**). "Frequency discrimination in listeners with sensorineural hearing loss," Ear Hear. **14**, 190–201.

Swets, J. A. (**1964**). *Signal Detection and Recognition by Human Observers* (Wiley, New York).

Turner, C. W., and Nelson, D. A. (**1982**). "Frequency discrimination in regions of normal and impaired sensitivity," J. Speech Hear. Res. **25**, 34–41.

Wakefield, G. H., and Nelson, D. A. (**1985**). "Extension of a temporal model of frequency discrimination: Intensity effects in normal and hearing-impaired listeners," J. Acoust. Soc. Am. **77**, 613–619.

Zwicker, E. (**1956**). "Die elementaren Grundlagen zur Bestimmung der Informationskapazitat des Gehors (The elementary foundations for the determination of the information capacity of hearing)," Acustica **6**, 356–381.

Zwicker, E. (**1970**). "Masking and psychological excitation as consequences of the ear's frequency analysis," in *Frequency Analysis and Periodicity Detection in Hearing*, edited by R. Plomp and G. F. Smoorenburg (Sijthoff, Leiden), pp. 376–394.

Influence of fundamental frequency and source elevation on the vertical localization of complex tones and complex tone pairs

Densil Cabrera
Faculty of Architecture, Design and Planning, University of Sydney, New South Wales, 2006, Australia

Masayuki Morimoto
Environmental Acoustics Laboratory, Faculty of Engineering, Kobe University, Kobe, Japan

(Received 24 May 2005; revised 26 October 2006; accepted 13 April 2007)

This study investigates the vertical localization of single complex tones (monads) and simultaneous complex tone pairs (dyads), especially as it is affected by their fundamental frequency and source elevation. Two complex tone timbres are considered: one consisting of five low-order harmonics, and the other of all odd harmonics (a square wave). Sound sources were at −15, 0, 15, and 30 deg from the horizontal plane at ear height. For eight subjects, this source array was in the median plane, and for a further nine subjects, it was directly to the subject's left (lateral plane). The subjects localized the angle of the auditory image(s) of one or two complex tones around the vertical plane containing the sound sources. Mean responses for the five-harmonic complex tones show a systematic effect (referred to as Pratt's effect) of fundamental frequency on vertical localization—whereby high-frequency complex tones are localized to positions higher than low-frequency complex tones for equivalent source positions. For the square wave, the sound-source position dominates localization, although some effect of fundamental frequency is evident for median plane sources. © *2007 Acoustical Society of America.* [DOI: 10.1121/1.2736782]

PACS number(s): 43.66.Qp [RAL] Pages: 478–488

I. INTRODUCTION

This study investigates the auditory vertical localization of complex tone stimuli, particularly in terms of how the source elevation and fundamental frequency affect the image elevation. An effect of tone frequency on image elevation was first observed by Pratt (1930) for five octave-related pure tones (256 to 4096 Hz), presented in the median plane [earlier, Urbantschitch (1889) obtained similar results for tones presented via tubes to the ears]. Similar experiments by Trimble (1934) and Roffler and Butler (1968a) also showed image elevation of pure tones to correlate positively to their frequency, being independent of the source elevation (Roffler and Butler found this for nine median plane pure-tone frequencies from 250 to 7200 Hz). The present study investigates whether Pratt's effect is present for complex, rather than pure, tones. This may be of interest for two reasons. First, the harmonic components of complex tones may increase a listener's ability to resolve the sound-source location (especially when the harmonic frequencies are in the range of pinna-related spectral localization cues), so the extent to which Pratt's effect remains gives an indication of localization cue weighting. Second, if Pratt's effect is to have some use outside of the laboratory (for example, in auditory graphing or other forms of auditory display), it would be of more use if it could be exploited through a variety of timbres rather than being restricted to pure tones.

The studies of Pratt's effect mentioned above were conducted with sound sources in the median plane. Many auditory localization experiments have used the median plane experiment paradigm because this provides the greatest angular range over which binaural difference cues are approximately constant, allowing the investigation to focus on the role of spectral cues (related to the pinnae, head, and torso) in auditory localization. This contrasts with, for example, the horizontal plane, around which binaural difference cues (interaural time and level differences) play a powerful role in auditory localization. By rotating the horizontal plane 90 deg around the interaural axis, we have the lateral vertical plane (henceforth referred to as the lateral plane), which is characterized by similar binaural difference cues for azimuth as the horizontal plane (but is distinguished by different spectral cues). The present study investigates whether Pratt's effect exists for laterally presented stimuli, that is, the extent to which distinct binaural difference cues suppress the effect. This test of the robustness of the effect is again of interest in understanding localization cue weighting, as well as pointing to the potential and limitations of practical applications of the effect (that is, whether it could be effective in an environment where sound sources are not fixed to the median plane).

The simultaneous vertical localization of low- and high-frequency stimuli has not been widely studied. Best *et al.* (2003) show that listeners use the interaural time difference (ITD) cue to segregate two identical synchronous broadband sound sources, and so are unable to do so for sources on the median plane. Similarly, Morimoto *et al.* (2003) find that the single auditory image location from a pair of synchronous high- and low-passed white noise bands (crossover frequency of 4.8 kHz) presented in the median plane is determined by the high-frequency source position—with no localization effect from the low-frequency source. Ferguson and

 0001-4966/2007/122(1)/478/11/$23.00

Cabrera (2005) find that, while separate auditory images can be elicited for low- and high- passed synchronous noise bands from vertically arrayed sound sources, the low-frequency auditory image is unaffected by the low-frequency source location, and instead follows the high-frequency source at a lower elevation (with a greater offset when the high-frequency source is at high elevations). Furthermore, a large spectral gap between the low- and high-frequency bands (the largest tested was for noise bands <125 Hz and >8 kHz) yielded greater image separation than a small spectral gap (the smallest was no gap with a crossover at 1 kHz)—results were generally consistent with Pratt's effect. The present study examines the simultaneous localization of complex tone pairs (dyads) possessing different fundamental frequencies in terms of the effects of the low-frequency and high-frequency sound-source elevations, as well as the effects of the fundamental frequencies. Auditory segregation of the two dyad components may be due to a combination of pitch and spatial separation of the auditory images, and the spatial separation could be due to the physical separation of sound sources and/or a localization illusion such as Pratt's effect.

According to Blauert (1997), Pratt's effect is part of the directional bands phenomenon, a term that originates from Blauert's (1969/70) observation that for median plane sources, auditory images of 1/3-octave noise bands around 8 kHz are heard overhead, and that 1/3-octave bands centered on other frequencies tend to be localized to the front or rear. This is supported by the study of directional bands by Itoh *et al.* (2006), which shows a general increase in image elevation associated with 1/3- or 1/6-octave band center frequency up to about 8 kHz, after which the image usually descends (cf. Middlebrooks, 1992). These studies use narrow-band noise stimuli, rather than pure tones—and Pratt's effect appears to be evident in other studies examining the vertical localization of broader noise bands, such as that of Ferguson and Cabrera (2005) mentioned above, and of Cabrera and Tilley (2003), which used octave noise bands (125 Hz–8 kHz). For even broader-band stimuli, Gardner (1973), Morimoto and Aokata (1984), and Noble (1987) observed that low-frequency noise bands tend to yield images with low elevations. These observations, consistent with Pratt's effect for non-narrow-band noise stimuli, suggest that the effect is related to the spectral envelope of the signal, rather than being based on pitch perception.

The present paper does not seek to better explain Pratt's effect, but does present further evidence of how Pratt's effect behaves. Specifically, it investigates the following questions: (i) Does Pratt's effect occur for complex tones, and if so, how does the presence or absence of high-frequency content (in the region of pinna-related elevation cues) affect it? (ii) Is there a difference between median and lateral vertical plane presentation for complex tone stimuli (considering that binaural difference cues are constant in the median plane, and variable in the lateral plane)? and (iii) To what extent is Pratt's effect manifest for simultaneous complex tone pairs (dyads)?

II. METHOD

The study used a subjective localization experiment, for complex tonal stimuli presented in either the median or lateral planes.

A. Room and equipment configuration

The experiment was conducted in an anechoic room. The subject was seated in a chair, and their head position was carefully aligned to and firmly restrained in a standard position using a slim-profile padded clamp. Loudspeakers were at a distance of 1.5 m from the subject, at −15°, 0°, +15°, and +30° relative to the horizontal. This limited range of angles was used in part because the authors imagined that most situations in which Pratt's effect might be used would involve sound sources near the horizontal plane, and also because the framework supporting the loudspeakers was designed for upper-hemisphere research (making it difficult to support loudspeakers at lower elevations). In the first part of the experiment, the vertical array of loudspeakers was directly ahead of the subject, and so was in the median plane. In the second part of the experiment, the loudspeaker array intersected the interaural axis on the subject's left side, and so was in the lateral plane. The loudspeaker drivers were nominally 100 mm in diameter. They were mounted in closed cylinders, so that the face of the loudspeaker consisted only of the driver.

A-weighted impulse responses were measured from each loudspeaker to a microphone (Bruel & Kjaer 4191) at a position that would be at the intersection of the median plane and interaural axis when a subject was seated. Inverse filters of these impulse responses were generated (8192-sample, 32-bit linear phase FIR filters, for a sampling rate of 48 kHz), using 1/3-octave-band smoothing, for the frequency range of 100 Hz to 16 kHz. The complex tones used in the experiment were convolved with these inverse filters, so that the effective system frequency response for each loudspeaker followed an inverse A-weighted function between 100 Hz and 16 kHz. System gains were adjusted so that the presentation level for every complex tone was 40 dB(A) at the microphone position. When two complex tones were presented simultaneously, their combined sound level was 43 dB(A). This sound level was chosen to avoid audible nonlinear distortion in the loudspeakers. Background noise in the room was below 20 dB(A).

B. Stimulus signals

The complex tone fundamental frequencies used in the experiment were 55.0, 110.0, 220.0, 311.1, 622.2, and 1244.4 Hz. For half of the stimuli, a simultaneous pair of complex tones (a dyad) was presented: these pairs were 55.0 Hz with 1244.4 Hz, 110.0 Hz with 622.2 Hz, and 220.0 Hz with 311.1 Hz. Hence, the geometric mean frequency for each dyad was constant at 261.6 Hz. These frequencies correspond to the musical tuning frequencies of A (the lower 3) and E-flat (the upper 3), with a geometric mean frequency of middle C. With each dyad spanning the musical interval of a tritone (plus zero, two, or four octaves), the

dyads were designed to be easily segregated into two pitches by the subject since they had no harmonics in common (Hutchinson and Knopoff, 1978).

Two timbres were used for the complex tones in the experiment: tones made up of five harmonics, and square waves. The five-harmonic complex tones had a −6-dB decrease in the level of successive harmonics. For all but the 55-Hz fundamental, the harmonics used were $f1-f5$. The 55-Hz complex tone used harmonics $f2-f6$ – i.e., the fundamental was absent, and heard as a virtual pitch. The square wave had a 50% duty cycle, consisting of odd harmonics up to the Nyquist frequency (but with a roll-off above 16 kHz). The 55-Hz square wave included the fundamental, but at a reduced level relative to the remaining harmonics. The two complex tones forming a dyad had the one timbre.

The five-harmonic timbre was chosen so as to have little or no sound in the frequency range of pinna-related vertical localization cues. Roffler and Butler (1968b) found that veridical vertical localization in the median plane requires complex frequency content above 7 kHz, and that this is not affected by the presence or absence of low-frequency content. Related findings have been made by many others, generally showing that spectral cues above about 5 kHz may be influential for vertical localization in the median plane (e.g., Shaw and Teranishi, 1968; Blauert, 1969/70; Hebrank and Wright, 1974; Mehrgardt and Mellert, 1977; Asano *et al.*, 1990). As observed by Shaw and Teranishi (1968) and Asano *et al.* (1990), a spectral notch moves from 6 to 10 kHz for source elevations from −45 to +45 deg, and the height of a spectral peak at 12 kHz also increases. Hence, most of the five-harmonic complex tones did not possess energy in this frequency range, the one exception being the highest component of the highest frequency complex tone (6222 Hz).

The loudspeaker array was conceived of as two overlapping pairs of loudspeakers (i.e., −15° paired with +15°, and 0° with +30°). When dyads were presented, they were emitted from one or other of these loudspeaker pairs. There were 96 stimuli, made up of all combinations of independent variables, namely two timbres, six individual complex tones (monads) presented from each of four loudspeakers, all dyads presented from all four possible configurations within a loudspeaker pair [(i) both complex tones from the low loudspeaker; (ii) both from the high loudspeaker; (iii) the high-frequency complex tone from the high loudspeaker and low-frequency complex tone from the low loudspeaker; and (iv) the high-frequency complex tone from the low loudspeaker and low-frequency complex tone from the high loudspeaker], and two overlapping loudspeaker pairs for the dyads.

Each stimulus was 1.2 s in duration, including a 100-ms fade-in and 100-ms fade-out. Each stimulus was presented five times, with an 800-ms interval between presentations. There was an 8.8-s interval until the next stimulus, so that the time between the start of each new stimulus was 18 s. The inverse filters for the loudspeakers were convolved with the complex tones prior to creating the tone envelopes (using tones of 1.5-s duration). This meant that the filter latency had no effect on the tone envelopes, because the envelopes were applied to the steady-state portions of the filtered complex tones.

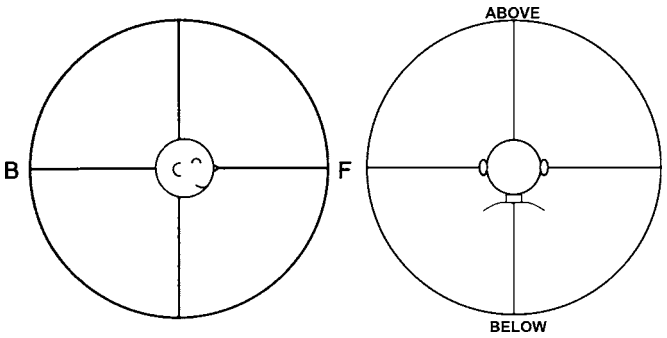


FIG. 1. Response sheets used for the experiment. The median plane response sheet is shown on the left, and the lateral plane sheet on the right.

C. Subjects and their task

Subjects responded to the stimuli with pencil and paper. For each response, they marked the auditory image angle or angles on a printed diagram, showing all possible angles around either the median or lateral plane (Fig. 1). The response marks consisted of a short line intersecting the large circle on the response sheet. When two pitches were heard, they marked both image angles, identifying the high pitched tone angle with an "H" and the low pitched tone angle with an "L." After the experiment, these responses were transcribed with the aid of a protractor to the nearest degree. The experiment was conducted in darkness, apart from a dim spot of illumination on the subject's lap (for the response sheets), and the loudspeaker array was not visible during the experiment. However, the loudspeaker array was visible to the subject as they entered the room and their head was aligned and clamped in position, but as there were about 35 identical loudspeakers in a frame around the subject, those used as sound sources in the experiment were not identifiable by sight. This method of presenting stimuli and recording responses has been used in previous studies of median plane localization (Morimoto, 2001; Morimoto *et al.*, 2003), providing accurate responses for easily localized stimuli such as broadband noise. Subsequent tests using a white noise stimulus indicate that a 95% confidence interval of ±3.6 deg could be expected in the present experiment for a white noise frontal median plane stimulus (assuming one response per subject), which can be compared to the just-noticeable difference of 4 deg determined by Kurosawa *et al.* (1982).

The 96 stimuli were presented in a different random order for each subject. The stimulus set was divided into three sessions of 32 stimuli, and the subject rested and stretched between sessions for between 5 and 15 min. A practice session of ten randomly selected stimuli was given to the subjects prior to the experiment.

One potential problem with this experiment is that some potential subjects might be unable to distinguish two simultaneous complex tones, and hence be unable to report their respective auditory image angles. Therefore, subjects with significant musical training were used, as these were likely to be able to make this distinction, considering that this is a key skill in musical training and that many studies have shown that musically trained listeners perform better than nonmusicians in categorical pitch perception (Burns, 1999). About

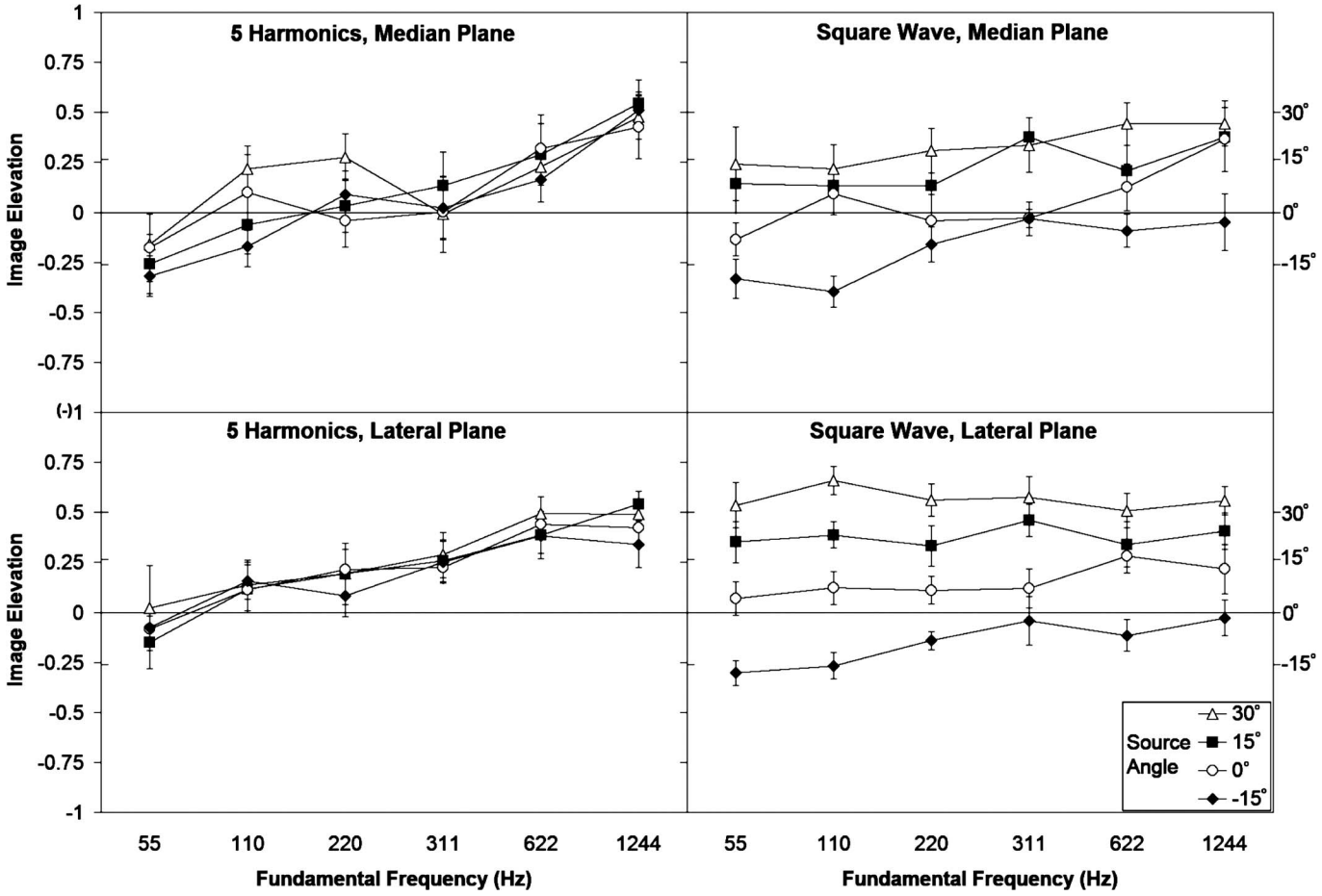


FIG. 2. Mean auditory image elevations for the monads (±1 standard error). Results are shown separately for the five-harmonic timbre and square wave timbre in the median and lateral planes.

half of the subjects were members of *a capella* singing groups.

Eight subjects were tested facing the active loudspeakers (median plane) and nine subjects were tested with the active loudspeakers to their left (lateral plane). Problems with the computer playing the stimuli meant that only partial results were obtained from three of the lateral plane subjects (3, 41, and 52 responses out of 96 were missing for these three subjects, and 25 stimulus responses were missed by two of these subjects). The subjects were aged between 19 and 23, and there were nine male and eight female subjects, approximately evenly distributed between median and lateral planes.

III. RESULTS

The responses for the two planes, for the two timbres, and for the monads and dyads, were analyzed separately, using factorial analysis of variance (ANOVA) to test for the presence and strength of effects, and Scheffe tests to test the significance of mean differences in responses for states of independent variables. For the analysis, responses were expressed as the sine of the image angle from the horizontal, which is a ratio representing the vertical component of that angle—referred to as image elevation. Hence, values have a possible range between 1 (90°, directly above the subject) and −1 (−90°, directly below the subject), with the source positions at −0.26 (−15°), 0 (0°), 0.26 (15°), and 0.5 (30°).

Figures 2, 4, and 5 show the results for the experiment, in terms of the image elevations as functions of stimulus frequency and source configuration. Figures 3 and 6 show the proportion of response variance that can be attributed to the independent variables used in ANOVA (source angle and

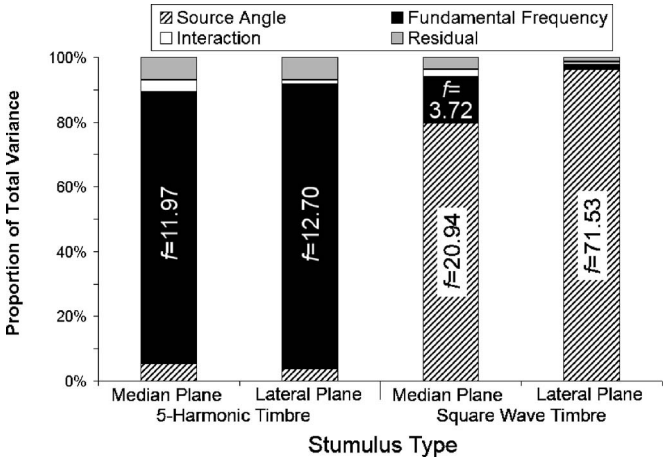


FIG. 3. The proportion of variance in the monad responses related to the source elevation, the complex tone fundamental frequency, and their interaction (based on the mean square deviations from ANOVA). The *f* statistic (which is the ratio of mean square deviation to the residual) is shown for significant effects, representing their absolute strength. Results are shown separately for the five-harmonic timbre and square wave timbre in the median and lateral planes.

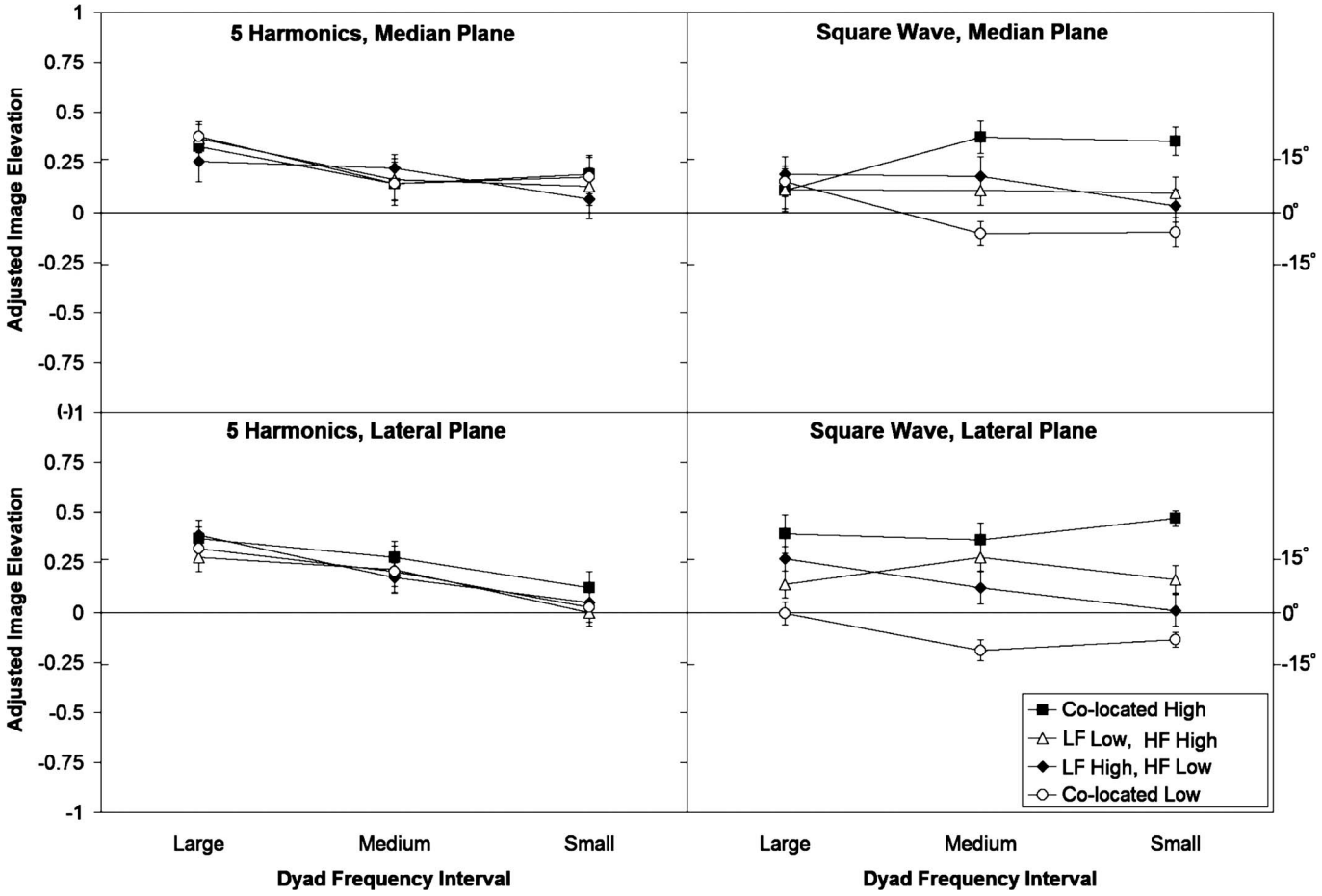


FIG. 4. Mean high-frequency auditory image elevations for the dyads (±1 standard error). Results are shown separately for the five-harmonic timbre and square wave timbre in the median and lateral planes. LF and HF refer to the low-frequency and high-frequency source positions, respectively.

stimulus frequency)—and so show the *relative* strength of effects for each type of stimulus. In *absolute* terms, the strength of effects is represented by the f statistic, significant values for which ($p<0.05$) are stated on these charts.

A. Monads in the median plane

For the five-harmonic timbre, the fundamental frequency significantly affects the auditory image elevation ($f=11.97, p<0.0001$), but the angle of the sound source has no effect ($f=0.77, p=0.51$). Mean responses are shown the upper-left quadrant of Fig. 2, and the general positive slope of the response functions exhibits Pratt's effect. The fact that the four functions representing each source elevation are approximately superimposed indicates that the source positions do not discernibly affect image elevation. While the responses generally exhibit Pratt's effect (with high-frequency images at greater elevations than low-frequency images), there is little or no spatial difference between the midfrequency (110.0–622.2-Hz) stimulus responses. A Scheffe test indicates significant mean differences for six of the 15 frequency pairs—all involving the highest and/or lowest fundamental frequencies. Figure 3 shows the relative influences of independent variables on the elevation responses for these and the other monad stimuli.

The square wave timbre responses in the median plane indicate a significant influence of source positions ($f=20.94, p<0.0001$), and also a weak effect for the fundamental frequency ($f=3.72, p=0.0032$). The mean responses, shown in the upper-right quadrant of Fig. 2, show that greater sound-source elevation corresponds to greater auditory image elevation, since the response functions for each sound source are spatially separated in order. However, localization is not veridical, since Pratt's effect also seems to be evident, with the high-frequency complex tones receiving more elevated responses than the others, and the two lowest frequency complex tones receiving responses with lower elevations. A Scheffe test shows significant differences between mean image elevations for most of the loudspeaker positions, with a nonsignificant difference between 0° and 15° ($p=0.13$), and no apparent distinction between 15° and 30°. However, when considering the square wave tone fundamental frequencies, only the 55- versus 1244-Hz stimuli have significantly different mean image elevations [with a near-significant difference between 110- and 1244-Hz image elevations ($p=0.068$)].

The experiment was not designed for individual subject analysis (as each subject responded to each stimulus just once, notwithstanding problems with the computer playback system). Nevertheless, a visual survey of the subject responses appears to show Pratt's effect for seven of the eight median plane subjects (five-harmonic timbre), and an effect of source position for six subjects (square wave).

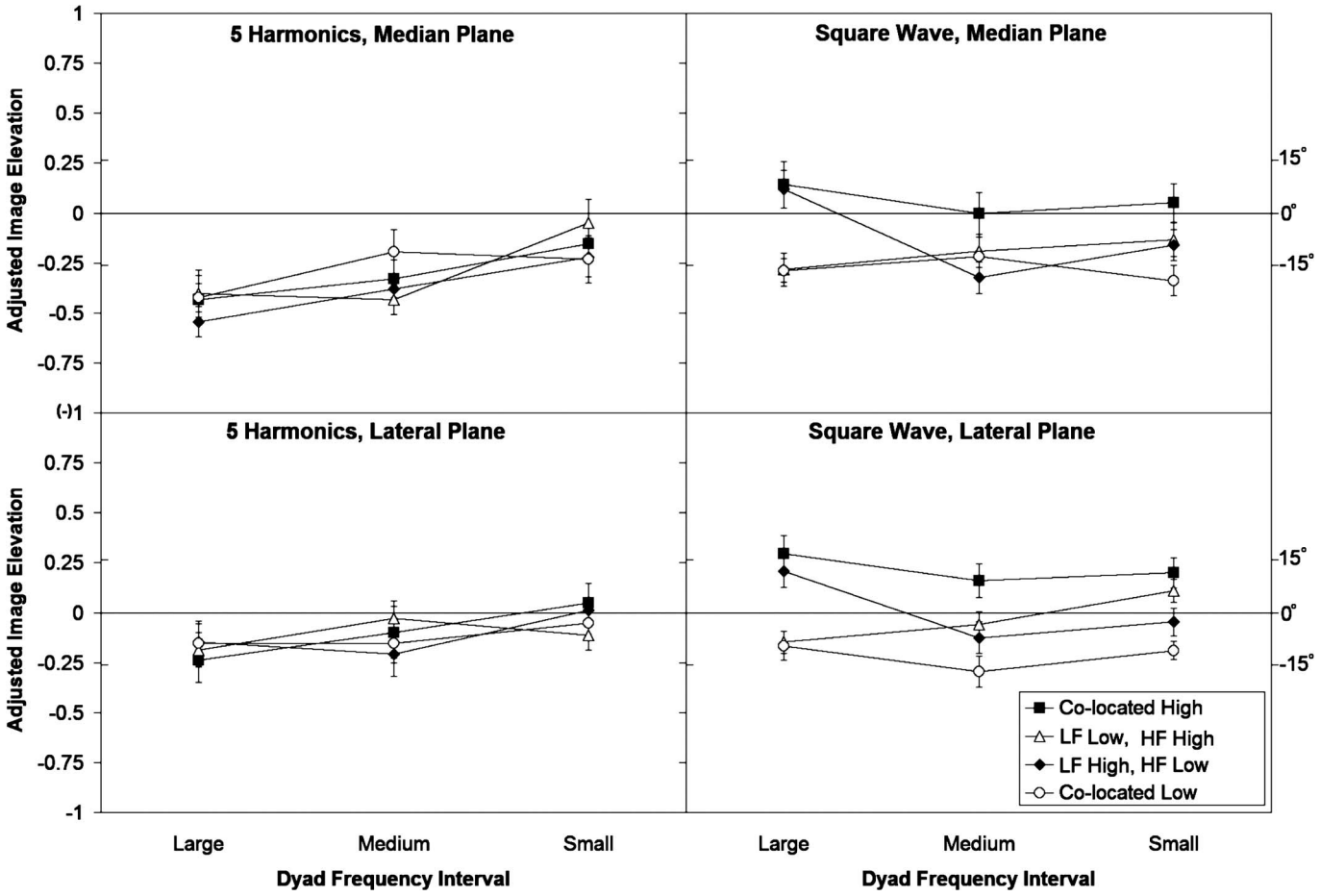


FIG. 5. Mean low-frequency auditory image elevations for the dyads (±1 standard error). Results are shown separately for the five-harmonic timbre and square wave timbre in the median and lateral planes. LF and HF refer to the low-frequency and high-frequency source positions, respectively.

For the five-harmonic timbre, 20.8% of images were to the subject's rear, and so can be classed as front-back reversals. This also occurred for 3.6% of the square wave images. Reversal rates for the various tone fundamental frequencies do not differ significantly. Front-back reversals were included in the analysis (since the elevation of the auditory image was the dependent variable).

B. Monads in the lateral plane

For the five-harmonic timbre, the fundamental frequency significantly affects the auditory image elevation ($f=12.70, p<0.0001$), but the angle of the sound source has no effect ($f=0.56, p=0.65$). As shown in the lower-left quadrant of Fig. 2, the responses exhibit Pratt's effect, with high-frequency images at greater elevations than low-frequency images. Responses are remarkably similar in pattern and scale to the median plane responses. A Scheffe test indicates significant mean differences for six of the 15 frequency pairs—all but one involving the highest and/or lowest fundamental frequencies.

The square wave timbre responses (lower-right quadrant of Fig. 2) show significant image elevation differences between source positions ($f=71.53, p<.0001$), but no effect for the fundamental frequency ($f=1.06, p=0.38$). The source elevations correspond approximately to image elevations. However, the mean auditory image positions appear to be a little higher than the source elevations. It is not known whether this is an artifact of the response interface or a perceptual phenomenon, because the lateral plane response interface has not been used previously by the experimenters. A Scheffe test shows significant differences between the mean image elevations of all loudspeaker pairs.

Results for all seven of the lateral plane subjects with complete or nearly complete results appear to show Pratt's effect for the five-harmonic timbre, as well as some ability to localize the square wave source.

C. Treatment of dyad responses

The dyad stimuli involve overlapping upper $(0°, 30°)$ and lower $(-15°, 15°)$ loudspeaker pairs—with identical and full combinations of the other independent variables presented from each of these pairs. In order to facilitate the analysis, responses for these pairs are combined. Hence, the dyad analysis is concerned with the difference in elevation between the members of each pair, but not the absolute elevation of the images from each of the four loudspeakers. Combining the results for the two pairs is done first by adding 7.5° to the lower pair and subtracting 7.5° from the upper pair responses (front-back reversals were flipped back to the front prior to this). This adjusts the sound sources for both pairs to have effective angles of −7.5° and 22.5°. Then, the sine of the shifted responses is taken, to give the image el-

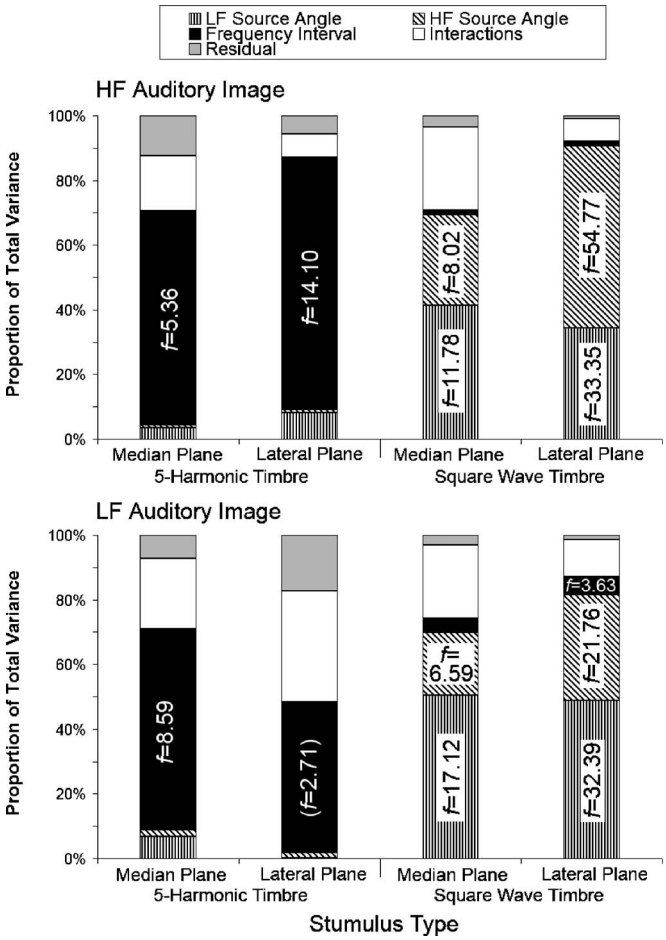


FIG. 6. The proportion of variance in the dyad responses related to the low-frequency source elevation, the high-frequency source elevation, the dyad frequency interval, and their interactions. The f statistic is shown for significant effects, and is parenthesized for one nonsignificant effect ($p=0.065$). The analysis for the high-frequency auditory image is in the upper chart, and the low-frequency auditory image in the lower chart. Results are shown separately for the five-harmonic timbre and square wave timbre in the median and lateral planes.

evation as a ratio (as was done for the monad responses). Finally, 0.126 is subtracted from these ratios to shift the effective sound-source elevation ratios to ±0.257 (corresponding to approximately ±15°, so that the midpoint is 0°). The last step has no effect on the analysis: its purpose is to make the result charts easier to read due to symmetry of the effective array around 0. Performing the shifts and sine transformation in this way and order minimizes errors from the nonlinear sine function to small fractions of a degree, which are negligible. Figure 4 shows the results for the high frequency auditory image, and Fig. 5 shows those for the low-frequency auditory image.

D. High-frequency image for dyads in the median plane

For the five-harmonic dyads, the dyad interval significantly affects the high-frequency auditory image elevation ($f=5.36, p=0.0055$), and the sound-source configuration has no effect ($p=0.89$). Mean responses are shown in the upper-left quadrant of Fig. 4, and the negative slope of the response functions is consistent with Pratt's effect. As the dyad interval decreases (i.e., the high-frequency complex tone's frequency is reduced), the high-frequency auditory image tends to descend, with significant differences between the large and medium dyad intervals ($p=0.037$), the large and small intervals ($p=0.012$), but not the medium and small intervals ($p=0.91$).

For the square wave dyads (upper-right quadrant of Fig. 4), the dyad interval has no primary effect on the high-frequency auditory image elevation ($p=0.67$). However, both the high-frequency source position ($f=8.02, p=0.0052$) and low-frequency source position ($f=11.78, p=0.0007$) significantly affect the high-frequency image elevation, and there is also a small significant interaction effect between dyad interval and high-frequency source elevation ($f=4.36, p=0.014$). The interaction effect between source configuration and dyad interval reflects the fact that the loudspeaker configuration has no apparent effect on the high-frequency image elevation for the large dyad interval, as can be seen in Fig. 4—whereas for the other two dyad intervals, the high-frequency and low-frequency sound sources both appear to influence the high-frequency auditory image. Of course, it is important to bear in mind that a low-frequency square wave is actually a broadband signal with dense spectral peaks, making high-frequency spectral localization cues available. Based on the Scheffe test, the significance levels for the mean difference in image elevation due to the high-frequency sound-source position is $p=0.005$, and for the low-frequency sound source position it is $p=0.0007$.

E. High-frequency image for dyads in the lateral plane

For the five-harmonic dyads, the dyad interval significantly affects the high-frequency auditory image elevation ($f=14.10, p<0.0001$), and the sound-source configuration has no effect ($p=0.50$). The same pattern is found as for the median plane (shown in the lower-left quadrant of Fig. 4), with the high-frequency auditory image less elevated as the dyad frequency interval is reduced, which is consistent with Pratt's effect. The significance of mean differences between the image elevations for the dyads are large-medium, $p=0.09$; large-small, $p<0.0001$; and medium-small, $p=0.01$.

For the square wave, both the high-frequency source position ($f=54.77, p<0.0001$) and the low-frequency source position ($f=33.35, p<0.0001$) strongly affect the high-frequency image elevation. There is no effect of dyad interval ($p=0.29$) and a minor interaction ($f=3.86, p=0.02$) between dyad interval and high-frequency source configuration. The Scheffe test shows a significance of $p<0.0001$ for the mean differences in image elevation due to both the high-frequency source elevation and the low-frequency source elevation. Although the mean responses (lower-right quadrant of Fig. 4) are generally not veridical, they do not reflect Pratt's effect, and instead show a combination of low- and high-frequency source elevation influencing high-frequency image elevation.

F. Low-frequency image for dyads in the median plane

Responses for the low-frequency auditory image elevation of the five-harmonic dyads are consistent with Pratt's effect, as indicated by the positive slope in the response functions in the upper-left quadrant of Fig. 5. In this case, the auditory image descends with an increase in the dyad interval. The dyad interval significantly affects low-frequency image elevation ($f=8.59, p=0.0003$), and the sound source configuration has no effect ($p=0.61$). There is a significant difference between large and small dyad intervals ($p=0.0003$), a marginally significant difference between medium and small ($p=0.054$), and no significant difference between large and medium ($p=0.24$).

The low-frequency auditory image for the square wave (upper-right quadrant of Fig. 5) is not significantly influenced by the dyad interval ($p=0.24$), is significantly affected by the high-frequency source position ($f=6.59, p=0.01$) and the low-frequency source position ($f=17.12, p<0.0001$), and there is a weak interaction effect between dyad interval and low-frequency source position ($f=4.51, p=0.012$). For the large dyad interval, responses indicate that the low-frequency sound source fully determines the low-frequency auditory image elevation—localization is veridical for these stimuli. The medium and small dyad intervals show influences of both sound sources. Based on the Scheffe test, the significance levels for the mean difference in low-frequency image elevation due to the high-frequency sound-source position is $p=0.01$, and for the low-frequency sound-source position it is $p<0.0001$.

G. Low-frequency image for dyads in the lateral plane

The low-frequency image for the five-harmonic dyads yielded the weakest result set in the experiment with respect to the independent variables. For these, there is merely a nonsignificant effect ($f=2.71, p=0.065$) for dyad interval on the low-frequency auditory image elevation. Nevertheless, the nonsignificant effect of dyad interval is consistent with Pratt's effect, evident from the general positive slope of the response functions in the lower-left quadrant of Fig. 5. Sound-source configuration has no effect ($p=0.99$).

The low-frequency auditory image for the square wave is significantly affected by high-frequency source position ($f=21.76, p<0.0001$), low-frequency source position ($f=32.4, p<0.0001$), and to a lesser extent by dyad frequency interval ($f=3.63, p=0.029$), and also shows an interaction effect between low-frequency source position and dyad frequency interval ($f=4.49, p=0.013$). The Scheffe test shows a significance of $p<0.0001$ for the mean differences in image elevation due to both the high-frequency source elevation and the low-frequency source elevation. Mean differences due to dyad interval are nonsignificant ($p=0.057$ for large-medium, $p=0.87$ for large-small, and $p=0.15$ for medium-small). The effect of dyad interval does not appear to be consistent with Pratt's effect in this case. The interaction effect reflects the dominance of the low-frequency source for the large interval dyad, with an increasing influence of the high-frequency source for the medium and small interval dyads, which can be seen in the lower-right quadrant of Fig. 5.

H. Specific interpretation of responses

In some cases, subjects reported two images for a monad. This occurred mainly for the 55-Hz square wave (35% of responses) and 110-Hz square wave (23%). The 220- and 1244-Hz square wave monads, respectively, had 3% and 5% of responses with two images. There were no cases of two images for the five-harmonic timbre, or for the 311- and 622-Hz square wave. The reason for multiple auditory images for low-frequency square waves could be that the high harmonics is more prominent than the low ones, and is close enough in frequency to produce a rough texture which is easily segregated from the low pitch. Also, pitch perception theory (Terhardt *et al.*, 1982) indicates that multiple pitches could be easily heard out from these low-frequency square waves (and indeed that was the experience of the experimenters). However, in order to avoid making assumptions about the subjects' perception, when two images were reported for a monad, the mean image angle was used in the analysis. The mean angle difference between these high-pitch and low-pitch images was 15° (i.e., on average, the high-pitch image was above the low), and included 28 cases where the high-pitch image was above the low-pitch image, seven cases of colocation, and seven cases of low-pitch image above the high-pitch image. The high-pitch image was closer to veridical (on average 4° elevated) than the low-pitch image (on average 11° lower than the source angle).

When one image was reported for a dyad, the auditory images of high- and low-frequency complex tones were treated as colocated in the analysis. For the five-harmonic timbre, this occurred for 1.5%, 0.8%, and 4.0% of the large, medium, and small frequency interval responses respectively. For the square wave timbre, this occurred for 0.8%, 3.1%, and 6.9% of the large, medium, and small frequency interval responses, respectively. Of course, 50% of the dyads were actually emitted from a single sound source.

IV. DISCUSSION

The results of this experiment show a combination of sound-source elevation and complex tone fundamental frequency affecting image elevation, and the relative strengths of these influences depend on factors including the spectral content (or timbre) of the complex tone and, to a much lesser extent, the horizontal angle of the sound source. For complex tones lacking high-frequency content (monads and dyads, in the median and lateral planes), image elevation is mainly affected by the tone fundamental frequency, in a manner consistent with Pratt's effect. Broadband complex tone monads are localized well in the lateral plane, and quite well in the median plane with a small influence of the complex tone fundamental that is consistent with Pratt's effect. For broadband complex tone dyads, a combination of low-frequency and high-frequency source elevation usually affects both of the image elevations, and Pratt's effect is not evident.

The effect of timbre on vertical localization in this study is consistent with the concept that spectral cues above 5 kHz are of prime importance, which is supported by many previous studies (including Roffler and Butler, 1968b). Hence, the square wave stimuli, which carry such spectral cues, have responses that show no or minimal evidence of Pratt's effect. However, a point of interest here is that Pratt's effect is evident for complex tone stimuli lacking high-frequency energy (and for which these pinna-related cues are unavailable). This extends the repertoire of signal types for which Pratt's effect may be found, which previously only included pure tones and bandlimited noise.

The results of this study indicate that Pratt's effect applies beyond median plane localization. The fact that the effect was present for sound sources on and near the interaural axis suggests that it may occur for all horizontal sound-source angles. This finding is consistent with Ferguson and Cabrera's (2005) lateral plane result for simultaneous high- and low-passed noise bands (where the low-passed band's image elevation exhibited Pratt's effect). Another way of thinking of this result is that it demonstrates some importance of spectral cues for accurate localization, even when binaural difference cues are available. Although this is not a common observation, it is also not a new one. For example, Musicant and Butler (1984) show that pinna-related spectral filtering, especially from the near ear, substantially increases localization accuracy in the horizontal plane for wideband noise stimuli that include energy above 4 kHz (this is apart from front-back reversals, which is a cone of confusion effect). Musicant and Butler (1985) find that narrow-band noise stimuli can exhibit the same horizontal localization patterns in monaural and binaural listening, even though binaural difference cues are physically present in the binaural case. In that study, the image angle on the horizontal plane varied very widely, determined by the center frequency of the narrow noise band rather than the sound-source angle—and this effect was related to the spectral localization cue. However, Macpherson and Middlebrooks (2002) find that monaural spectral cues have little influence on lateral localization of broadband noise stimuli, given the availability of binaural difference cues. Slattery and Middlebrooks (1994) and Wightman and Kistler (1997) find that with one ear plugged, subjects with normal hearing are unable to perform lateral localization accurately (although subjects with long-term hearing loss in one ear are able to make such judgments). In reconciling these (and the present) results, binaural difference cues are clearly dominant in auditory localization, and spectral cues may be used to refine the accuracy of localization judgments, both in terms of the image angle around a cone of confusion and for sharpening the cone (which might be better thought of as a disk for source angles near the interaural axis).

Only a small proportion (11%) of the dyad responses indicated colocated or single images (6% of dyad responses had both colocated sources and images). The implication is that in the large majority of cases, subjects could perceptually separate the two signals of the dyad. Nevertheless, the results show that the basis of this separation is not always simple. For the five-harmonic timbre dyads, this spatially expressed separation is due to spectral differences, because the source elevation does not affect the image elevation. For the square wave a combination of high- and low-frequency source locations affect the image elevation, the exception being the 55-Hz dyad component, which is well localized. The 55-Hz square wave tone has the greatest spectral density, and so provides the best stimulus to provide pinna-related spectral cues for vertical localization. Results for the remaining components of the square wave dyads appear to reflect interference between localization cues for each source, which in the median plane consist only of spectral cues.

As mentioned in the Introduction, Pratt's effect is considered to be part of the directional bands phenomenon. The mechanism behind this is explained by the concept of boosted bands, also known as covert peak areas (Butler and Helwig, 1983; Rogers and Butler, 1992), whereby the auditory image of a narrow-band signal tends to be localized to the direction that has the maximum transfer function for that frequency. A related supportive concept is that pinna-related spectral features tend to increase in frequency as source elevation increases, and these features form the boosted bands in the high-frequency range. Hence, Bothe and Elfner (1972) and Gardner (1973) note that the apparent "pitch" of broadband noise increases with source elevation. However, Pratt's effect in the low-frequency range (e.g., below 1 kHz) cannot be explained by directional dependence of pinna-related spectral features, because the pinnae have little effect at low frequencies (although possibly torso reflections may provide some elevation-dependent spectral cues as low as 700 Hz—Algazi *et al.* (2001) give physical evidence for this, and Carlile *et al.* (1999) find modest spectrum-based localization effects for noise low-passed at 2 kHz, but it is not clear that elevation is affected). The fact that low-passed sound stimuli do not produce elevated auditory images (noted earlier) may reflect a default image elevation that occurs in the absence of relevant pinna-related spectral cues, as suggested by Vliegen and Van Opstal (2004). A further possibility is that learning from environmental transfer functions in everyday life develops an association between low frequency and low elevation, as suggested by Cabrera *et al.* (2006), based on a survey of elevation-related transfer functions in 20 everyday environments.

Although this study is concerned with sound localization, there is also a substantial body of literature on metaphoric and linguistic associations between pitch and height, especially with regard to the perception and philosophy of music. Early instances are reviewed by Pratt (1930), and more recent instances are reviewed by Eitan and Granot (2006). In some cases these dismiss anything other than metaphor (e.g., Lippman, 1952; Zuckerkandl, 1956; Scruton, 1983) in linking pitch to height. Some examine the extent to which an association exists, for example in young children (Wagner *et al.*, 1981) or across cultures (Walker, 1987). Most comment on the tendency in many (but not all) languages to use high-low words to describe pitch (and this includes Japanese, which is the native tongue of the subjects in the present study). It is very difficult to examine the extent to which such associations affect auditory vertical localization—but, since they tend to be consistent with Pratt's effect, it is con-

ceivable that they contribute together with the directional bands and related mechanisms in vertical localization. The range of mean image elevations over which Pratt's effect is seen in this experiment matches the range of sound-source elevations (−15° to +30°), at least for the monad responses. This could indicate that responses for the five-harmonic timbres are shaped by the range of image positions experienced for the square wave timbres, considering that Roffler and Butler (1968a) demonstrated that the extent of Pratt's effect is susceptible to influence from expectations (in their case, the visual angle subtended by a vertical reference scale).

This study used stimuli of 40 or 43 dB(A) with a fixed duration, raising the question of whether similar results might be expected for higher sound-pressure levels or different durations. Blauert (1969/70) finds that neither sound-pressure level (30–60 dB) nor duration (100 ms–1 s) affect coarse categorical localization of narrow noise bands in the median plane (whereas frequency does). Using various longer duration stimuli, Davis and Stephens (1974) find that localization accuracy in the median plane increases with sensation level up to 70 dB (compared with 25 dB for horizontal localization), but do not find a systematic shift in image elevation associated with sound level. In an experiment examining Pratt's effect for octave noise bands, Cabrera and Tilley (2003) find that a 20-phon loudness level interval (from 64 to 84 phons) has no effect on mean image elevation (although the image size increases substantially with this level change). However, others have found that sound level can affect the extent to which auditory image elevation deviates from the default horizontal position—that is, the slope of the stimulus-response function for elevation, which is referred to as elevation gain (Macpherson and Middlebrooks, 2000; Vliegen and Van Opstal, 2004). Over a large range of sound-pressure levels (26–73 dB), Vliegen and Van Opstal (2004) find an inverted-U relationship between intensity and elevation gain, with greatest elevation gain in the vicinity of 50 dB. They also find elevation gain to be affected by stimulus duration, at least for the short durations that they tested (3–100 ms), such that greater gain occurs for long durations. For the most part, elevation gains were less than or equal to 1, meaning that the elevation of auditory images tended to be veridical or compressed relative to the source elevations. These studies indicate that, while level and duration may have some effect on vertical localization, they do not provide a significant confound with Pratt's effect.

The combined findings of this and previous studies indicate that, while Pratt's effect may not result in large localization shifts, it can be a robust effect for stimuli that lack dense spectral content above 5 kHz (in the region of pinna-related elevation cues). Pratt's effect has been found for signals including pure tones, narrow noise bands, octave noise bands, low-passed noise bands, and complex tones. It is present in both anechoic and normally echoic rooms (at least some of Roffler and Butler's experiments were conducted in nonanechoic conditions). There is some reason to speculate that the effect could be stronger in nonanechoic conditions because tones are difficult to localize accurately in such conditions, and Hartmann and Rakerd (1989) show that at least some localization illusions are present in nonanechoic rooms because of this, but absent in anechoic rooms. Roffler and Butler (1968a) made several findings including: Pratt's effect is evident in the responses of young children; the effect is affected by visual stimuli (the visual scale used to form responses); the effect is not dependent on visual associations (since it is found in responses of congenitally blind subjects); and the elevation shifts observed tend to be related to the subject's body, rather than the direction of gravity (for subjects listening when lying on their backs or sides). A key finding of the present study is that the effect is equally present for some stimulus signals in median plane and lateral plane presentation, and so is likely to be independent of azimuth for such signals. While the present study used precise head positioning and restraint, other studies have not necessarily used any head restraint, and have still found the effect. As shown by this study, the effect may be observed in the simultaneous localization of two stimuli. This extends to high- and low-passed noise bands, and even to the high- and low-passed components of a music signal produced through a two-way loudspeaker (Ferguson and Cabrera, 2005). As such, the effect could have practical relevance to the arrangement of drivers on the face of a multiway loudspeaker (a woofer above a tweeter yields a more compact overall auditory image than the more conventional inversion of this). Auditory display is another area in which the effect could practically exploited, most obviously in the representation of elevation through sound, as has been done by Susnik *et al.* (2005).

V. CONCLUSIONS

The results of this study indicate that both the source elevation and the fundamental frequency of complex tones can systematically affect the vertical elevation of auditory images, and the strengths of these influences depend on the spatial and spectral attributes of the stimulus. As might be expected from previous findings on pinna-related spectral cues, the presence of high-frequency harmonic content does make sound sources easier to localize, thereby reducing or eliminating Pratt's effect. A difference between the median and lateral plane response is only clearly evident for the full-bandwidth complex tone—for complex tones with no high-frequency components, there is little difference apparent in vertical localization. Regarding Pratt's effect for dyads—the effect is clearly evident for the complex tones possessing no high-frequency components, both through the localization of the individual auditory images of the two complex tones, and through the accord between spectral interval and spatial interval. Pratt's effect is not evident at all for full-bandwidth complex tone dyads—instead, image elevation is mainly determined by a combination of the low- and high-frequency source elevations. The relative and absolute strengths of these effects are summarized in Figs. 3 and 6.

ACKNOWLEDGMENTS

The authors thank Kimihiro Sakagami, Hayato Sato, and Taku Nishimura for their contributions to the experimental work of this project. They also thank the experiment volunteers, and the masters students in the Environmental Acous-

tics Laboratory at Kobe University who assisted with the experiment. They thank Simon Carlile, Simon Hayman, Jens Brosbøl, and two anonymous reviewers for advice in preparing the manuscript.

Algazi, V. R., Avendano, C., and Duda, R. O. (**2001**). "Elevation localization and head-related transfer function analysis at low-frequencies," J. Acoust. Soc. Am. **109**, 1110–1122.

Asano, F., Suzuki, Y., and Sone, T. (**1990**). "Role of spectral cues in median plane localization," J. Acoust. Soc. Am. **88**, 159–168.

Best, V., Schaik, A., and Carlile, S. (**2003**). "Separation of concurrent broadband sound sources by human listeners," J. Acoust. Soc. Am. **115**, 324–336.

Blauert, J. (**1969/70**). "Sound localization in the median plane," Acustica, **22**, 205–213.

Blauert, J. (**1997**). *Spatial Hearing: The Psychophysics of Human Sound Localization*, rev. ed. (MIT Press, Cambridge, MA).

Bothe, S. J., and Elfner, L. F. (**1972**). "Monaural vs binaural auditory localization for noise bursts in the median vertical plane," J. Aud Res. **12**, 291–296.

Burns, E. M. (**1999**). "Intervals, scales and tuning," in *The Psychology of Music*, 2nd ed., edited by D. Deutsch (Academic, London, UK), pp. 215–264.

Butler, R. A., and Helwig, C. C. (**1983**). "The spatial attributes of frequency in the median sagittal plane and their role in sound localization," Am. J. Otolaryngol. **4**, 165–173.

Cabrera, D., Morimoto, M., Sakagami, K., and Sato, H. (**2006**). "Potential low frequency environmental cues for sound source elevation," in Proceedings of the 9th Western Pacific Acoustics Conference (Seoul).

Cabrera, D., and Tilley, S. (**2003**). "Parameters for the auditory display of height and size," in Proceedings of the International Conference on Auditory Display (Boston, MA), pp. 29–32.

Carlile, S., Delaney, S., and Corderoy, A. (**1999**). "The localization of spectrally restricted sounds by human listeners," Hear. Res. **128**, 175–189.

Davis, R. J., and Stephens, S. D. G. (**1974**). "The effect of intensity on the localization of different acoustical stimuli in the vertical plane," J. Sound Vib. **35**, 223–229.

Eitan, Z., and Granot, R. Y. (**2006**). "How music moves: Musical parameters and listeners' images of motion," Music Percept. **23**, 221–247.

Ferguson, S., and Cabrera, D. (**2005**). "Vertical localization of sound from multi-way loudspeakers," J. Audio Eng. Soc. **53**, 163–173.

Gardner, M. B. (**1973**). "Some monaural and binaural facets of median plane localization," J. Acoust. Soc. Am. **54**, 1489–1495.

Hartmann, W. M., and Rakerd, B. (**1989**). "Localization of sound in rooms IV. The Franssen effect," J. Acoust. Soc. Am. **84**, 1366–1373.

Hebrank, J., and Wright, D. (**1974**). "Spectral cues used in the localization of sound sources on the median plane," J. Acoust. Soc. Am. **56**, 1829–1834.

Hutchinson, W., and Knopoff, L. (**1978**). "The acoustical component of western consonance," Interface (USA) **7**, 1–29.

Itoh, M., Iida, K., and Morimoto, M. (**2006**). "Individual differences in directional bands in median plane localization," Appl. Acoust. **68**, 909–915.

Kurosawa, A., Takagi, T., and Yamaguchi, Z. (**1982**). "On the transfer function of the human ear and auditory localization," J. Acoust. Soc. Jpn. **38**, 145–151. (in Japanese)

Lipmann "Music and Space: A Study in the Philosophy of Music." (**1952**). Ph.D. thesis, Columbia University.

Macpherson, E. A., and Middlebrooks, J. C. (**2000**). "Localization of brief sounds: Effects of level and background noise," J. Acoust. Soc. Am. **111**, 2219–2236.

Macpherson, E. A., and Middlebrooks, J. C. (**2002**). "Listener weighting of cues for lateral angle: The duplex theory of sound localization revisited," J. Acoust. Soc. Am. **108**, 1834–1849.

Mehrgardt, S., and Mellert, V. (**1977**). "Transformation characteristics of the external human ear," J. Acoust. Soc. Am. **61**, 1567–1576.

Middlebrooks, J. C. (**1992**). "Narrow-band sound localization related to external ear acoustics," J. Acoust. Soc. Am. **92**, 2607–2624.

Morimoto, M., and Aokata, H. (**1984**). "Localization cues of sound sources in the upper hemisphere," J. Acoust. Soc. Jpn. (E) **5**, 165–173.

Morimoto, M. (**2001**). "The contribution of two ears to the perception of vertical angle in saggital planes," J. Acoust. Soc. Am. **109**, 1596–1602.

Morimoto, M., Yairi, M., Iida, K., and Itoh, M. (**2003**). "The role of low frequency components in median plane localization," Acoust. Sci. & Tech. **24**, 76–82.

Musicant, A. D., and Butler, R. A. (**1984**). "The influence of pinnae-based spectral cues on sound localization," J. Acoust. Soc. Am. **75**, 1195–1200.

Musicant, A. D., and Butler, R. A. (**1985**). "Influence of monaural spectral cues on binaural localization," J. Acoust. Soc. Am. **77**, 202–208.

Noble, W. (**1987**). "Auditory localization in the vertical plane: Accuracy and constraint on bodily movement," J. Acoust. Soc. Am. **82**, 1631–1636.

Pratt, C. C. (**1930**). "The spatial character of high and low tones," J. Exp. Psychol. **13**, 278–285.

Roffler, S. K., and Butler, R. A. (**1968a**). "Localization of tonal stimuli in the vertical plane," J. Acoust. Soc. Am. **43**, 1260–1266.

Roffler, S. K., and Butler, R. A. (**1968b**). "Factors that influence the localization of sound in the vertical plane," J. Acoust. Soc. Am. **43**, 1255–1259.

Rogers, M. E., and Butler, R. A. (**1992**). "The linkage between stimulus frequency and covert peak areas as it relates to monaural localization," Percept. Psychophys. **52**, 536–546.

Scruton, R. (**1983**). "Understanding music," Ratio **25**, 97–120.

Shaw, E. A. G., and Teranishi, R. (**1968**). "Sound pressure generated in an external-ear replica and real human ears by a nearby point source," J. Acoust. Soc. Am. **44**, 240–249.

Slattery, W. H. I., and Middlebrooks, J. C. (**1994**). "Monaural sound localization: Acute versus chronic impairment," Hear. Res. **75**, 38–46.

Susnik, R., Sodnik, J., and Tomazic, S. (**2005**). "Coding of elevation in acoustic image of space," in *Proceedings of the Australian Acoust. Society Conference* (Busselton, Australia), pp. 145–150.

Terhardt, E., Stoll, G., and Seewann, M. (**1982**). "Algorithm for extraction of pitch and pitch salience from complex tonal signals," J. Acoust. Soc. Am. **71**, 679–688.

Trimble, O. C. (**1934**). "Localization of sound in the anterior-posterior and vertical dimensions of 'auditory' space," Br. J. Psychol. **24**, 320–334.

Urbantschitch, V. (**1889**). "On the science of sound sensations" ("Zur Lehre von den Schallempfindungen"), Pfluegers Arch. **24**, 574–595.

Vliegen, J., and Van Opstal, J. (**2004**). "The influence of duration and level on human sound localization," J. Acoust. Soc. Am. **115**, 1705–1713.

Wagner, Y. S., Winner, E., Chicchetti, D., and Gardner, H. (**1981**). "'Metaphorical' mapping in human infants," Child Dev. **52**, 728–731.

Walker, R. (**1987**). "The effects of culture, environment, age, and musical training on choices of visual metaphors for sound," Percept. Psychophys. **42**, 491–502.

Wightman, F. L., and Kistler, D. J. (**1997**). "Monaural localization revisited," J. Acoust. Soc. Am. **101**, 1050–1063.

Zuckerkandl, V. (**1956**). *Sound and Symbol* (Pantheon, New York).

Front-back confusion resolution in three-dimensional sound localization using databases built with a dummy head

Alexander Ovcharenko, Sang Jin Cho, and Ui-Pil Chong[a)]
School of Computer Engineering and Information Technology, University of Ulsan, 7-318-1, San 29, Moogeodong, Namku, Ulsan, South Korea, 680-749

(Received 19 May 2006; revised 19 April 2007; accepted 21 April 2007)

Sound localization plays an important role in everyday life. It helps us to separate sounds coming from different sources and thus to acquire acoustic information. This paper describes an algorithm for localizing the position of a sound source, as recorded by dummy head microphones. The recorded signals are considered to be basic, random signals within an imaginary round room. The goal of this research is to localize random signals produced from different positions using information about basic signals. The method used is based on the identification of similarities between basic and random signals. It includes an interaural time difference comparison at the beginning, and continues with further analysis of the differences in signal spectrums. One of the main issues arising in sound localization is the problem of front-back confusion, and this paper shows how it was resolved by the use of reference signals. © *2007 Acoustical Society of America.*
[DOI: 10.1121/1.2739415]

PACS number(s): 43.66.Qp [EJS] Pages: 489–495

I. INTRODUCTION

The spatial separation of our ears allows us to localize sounds from three physical factors: time, intensity and spectrum. Naturally, people are more sensitive to sounds coming from the horizontal plane, as it provides the maximum difference for sound events occurring around the listener, rather than from below or above.[1] In this case we use the interaural time difference (ITD) and interaural intensity difference (IID) attributes of sound for localization of the sound source. An ITD is defined as the difference in the time reaching the left and right ear drums by a sound. IID is defined as the amplitude difference (usually in dB) in the wave front of a sound between left and right ear.[2] However, ITD and IID cues are limited, and cannot answer all the questions about sound's spatial location. Moreover, let us consider left or right hemisphere of the head in the interaural horizontal plane. When you begin to move a sound source symmetrically, relatively 90° to each side of the head, an ambiguity of interaural differences is evident because they are the same from both the positions.[3] Thus, there is a confusion in distinguishing sound sources in the front, back, and overhead. We came to the conclusion that we should keep one of the interaural cues and resolve the front-back confusion problem with further spectral analysis of the signal we wanted to localize.

It is not always suitable to use real human ears for binaural recording. First, every human head has its own specifications. Second, as every signal is additionally modified by the shape of ears, we would have to mount microphones in the ear canals and calibrate them to perceive the equal wave fronts, which by itself is not an easy procedure. Third, it would be very difficult for one to fix the head, and prevent it from moving and making noise.[4] Thus, the technique of dummy head recording has become more and more popular.

A Dummy head is a mannequin with a head and a torso for sound quality testing and recording. Two microphones, positioned at the entrances to the manikin's ear canals, simulate the human ear separation and ensure a signal that includes the interference patterns caused by the head and upper body. This gives an extremely accurate three-dimensional, binaural recording.[5]

In the work presented here we used a dummy head to record signals from different spatial locations of an imaginary round-shaped room and tried to analyze information about the obtained signals with the sound localization cues mentioned above. We recorded signals that were considered to be basic signals and placed them in the database. Then, we also recorded random signals from the same distance to the dummy head and tried to localize them using the previously gathered information about basic signals.[6–11]

The problem of front-back confusion was one of the main issues during sound localization. To our knowledge, all algorithms use the rotation of the dummy head to catch the slight difference in ITD of the sound source. In this work there was an attempt to use static signals (recorded and put into the database) and resolve front-back confusion using reference signals, which were also taken from the database. For the simulation of our algorithm we used MIT head related transfer function (HRTF) data and our own data sets recorded in an anechoic room as well as in a common lecture room.

The paper is organized as follows. Section II presents previous experiments and results. The algorithm of sound localization with front-back confusion resolution is proposed in Sec. III. Simulation results are given in Sec. IV. Conclusions and future works are presented in Sec. V.

[a)]Author to whom correspondence should be addressed; Electronic mail: upchong@mail.ulsan.ac.kr

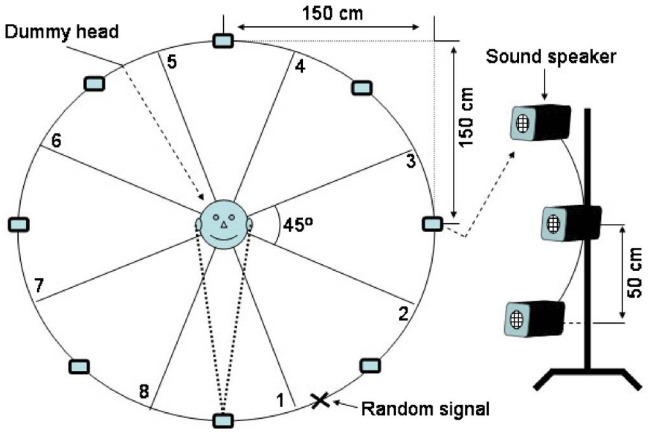


FIG. 1. (Color online) Scheme of the sound recording.

II. PREVIOUS EXPERIMENTS AND RESULTS

It is important to discuss previous experiments and the preliminary results we obtained while designing our algorithm. At first, we used only interaural cues (ITD and IID) for sound localization. It is well known from duplex theory that ITD and IID are not sufficient to localize a sound source's position because of the limited potentials and because of front-back confusion.

However, sound localization cues from duplex theory are worth considering as they can be basic beginning cues for the localization of the sound. For example, with the help of ITD information it is easy to identify a sound source's position accurately within a confusion block. It is shown below how ITD and IID can be analyzed and used to localize an upcoming wave front.

A. Experiment in an anechoic room

The goal of the experiment was to create a database of several basic sound sources with the subsequent recognition of where each random sound signal comes from. A scheme of the sound recording is shown in Fig. 1. As the picture shows, we divided an anechoic room into 24 equal domains as follows: eight domains covered 360° of azimuth changing, and three levels of such domains were taken to provide elevation difference. Then, we put a sound speaker device into each domain, producing a mono sound through the sound card to be localized by the dummy head. Every speaker was mounted at a 1.5 m distance from the head and was supposed to be a center in the board line of a domain.

We recorded basic signals from specific blocks (one sound speaker was responsible for one domain) with digital tape and enumerated them according to increases in the azimuth. After that, we obtained random signals from the same distance and elevation we used for producing basic signals, but with a different azimuth location. Finally, we considered random signals to be sound signals with an unknown spatial location and tried to use different sound localization cues to identify the number of the domain with the random wave front produced.

Output signals were obtained using two different input signals for further comparison. In both cases we used signals which covered all components in the frequency domain.

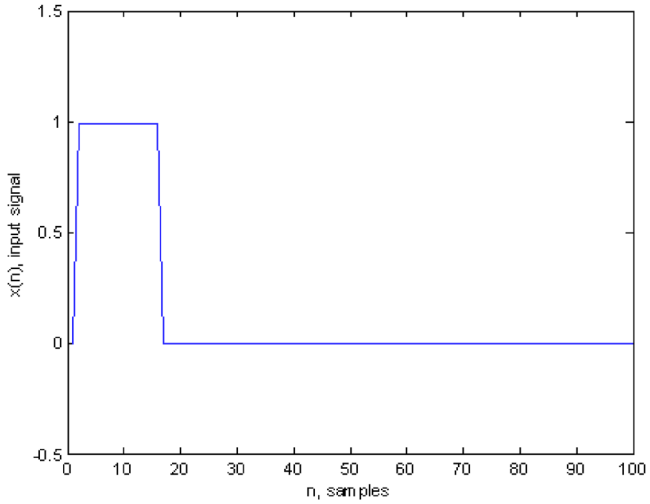


FIG. 2. (Color online) First 100 samples of the impulse signal.

Each sequence length contained 4096 samples and was produced with 44.1 kHz sampling frequency. The first signal (which is called an Impulse signal) is a combination of two unit step functions as depicted in Fig. 2. The second signal (also known as a Clap signal) was formed by a hand clap sound and recorded by a microphone. Figure 3 illustrates the Clap input signal.

B. Spatial coordinate system

A spatial coordinate system has been chosen as it frequently appears in most three-dimensional sound papers and books. The azimuth is described as increasing counterclockwise from 0 to 360° in the horizontal plane. Elevation increases from 0 to +90° directly above the listener and decreases from 0 to −90° directly below the listener's head.[1]

C. Duplex theory application

Overall, we made 24 sound recordings for each of the basic original signals and 12 records for each of the random input sound sources.

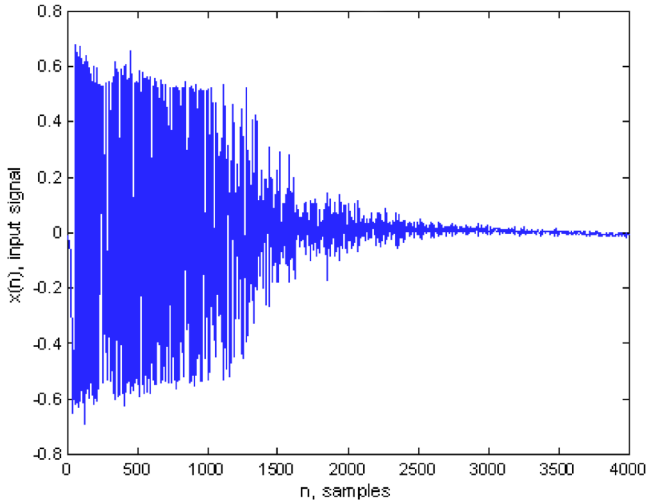


FIG. 3. (Color online) Clap signal.

TABLE I. Numbers of blocks found with IID and ITD comparison of basic and random signals. Elevation is −18°.

		Random signal azimuth			
Type of signal	Method of comparing	30°	120°	210°	300°
Impulse	IID	2	3	6	6
	ITD	4	2	8	6
Clap	IID	1	2	6	8
	ITD	2	2	8	6
	Correct	2	4	6	8

We took random signals one by one, calculated their IIDs and ITDs, and compared them with the interaural attributes of all the basic signals. As we enumerated database signals with increasing azimuth, the index of minimum IID or ITD difference pointed us to a block number (two neighboring basic signals define one block) which was assumed to contain a random signal source. For example, the minimum ITD difference index for each jth random signal can be computed as

$$k = index\{\min|ITDb_i - ITDr_j|\}, \quad (1)$$

where $ITDb_i$ and $ITDr_j$ are basic ith and random jth ITDs respectively. As an example, the results of calculations with −18° elevation are given in Table I.

It can be seen from the presented results that ITD provides us with more precise information about the location of random signal than IID. The main problem to point at the right azimuth location of the random sound source was front-back confusion. Figure 1 shows three pairs of blocks which have almost identical ITD, they are: [1,5], [2,4], [8,6]. We try to keep the ITD data and eliminate front-back confusion with further spectral analysis of the signals.

III. ALGORITHM OF SOUND LOCALIZATION WITH THE RESOLUTION OF FRONT-BACK CONFUSION

It is said that complete characterization of the auditory cues associated with a single spatial location involves the measurement of three quantities: left- and right-ear magnitude responses and the ITD.[2] We use this property to build our algorithm for finding the location of a random signal source relying on the database of the information about basic signals.

First, we get ITD information of the considered cth random signal to identify in which part of the head we are. Suppose that all the tables with ITD and fast Fourier transform (FFT) data of basic signals are arranged by the azimuth growth. For each degree interval $[n,n+1]$ we compare the ITD difference (*ITDdiff*) between two basic FFTs of nth and $n+1$th sound sources ($ITDdiff_{n,n+1}$) and the difference between each of them and the cth one ($ITDdiff_{n,c}$, $ITDdiff_{n+1,c}$) in absolute values. If $|ITDdiff_{n,c}| \le |ITDdiff_{n,n+1}|$ and $|ITDdiff_{n+1,c}| \le |ITDdiff_{n,n+1}|$, we can say that our arbitrary signal is possibly located between nth and $n+1$th signals (let us call the $[n,n+1]$ interval as *suitable interval*).

Second, we use two database signals of 0 and 180° of azimuth location and try to identify for which of them the rest of the sound sources in the database are closer. As the signals produced from 0 and 180° cause confusion by themselves, it was interesting to investigate their relation to the other signals in the frequency domain. We calculate the differences between each pair of reference signals and database signals which are in the suitable interval. If we use the mean value method to compare signals, for the nth database signal it can be written mathematically as

$$diff_n = \frac{1}{n}\sum_{i=1}^{N} |X_n(i) - X_0(i)| - \frac{1}{N}\sum_{i=1}^{N} |X_n(i) - X_{180}(i)|, \quad (2)$$

where X_n is the FFT of the nth sound, X_0 and X_{180} are FFTs of the sounds located in 0 and 180° of azimuth, respectively, N is a number of frequency values. Figure 4 depicts how $diff_n$ of MIT database changes with increasing angle for the left and right ear.

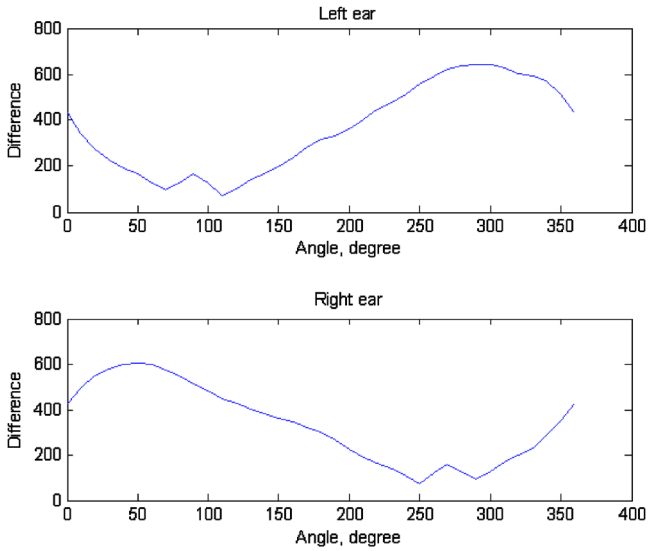


FIG. 4. (Color online) The arithmetical mean of differences between two reference and each database FFTs.

Then, we take the current $diff_n$ and the next $diff_{n+1}$ values to determine whether the value of the considered signal $diff_c$ lies between them. If it is located in the interval $[diff_n, diff_{n+1}]$, we can conclude that our random signal has some similarity with the nth and $n+1$th.

Generally, we make our calculations for both the left and right ears and then compare the numbers of signals from our database for each ear. The co-inside indexes of the left and right ears can point us at the nearest position of the random signal.

If we consider that the difference intervals $[diff_n, diff_{n+1}]$ have a linear structure, we can build a line between the closest points of the indexes we have found. After that, we can point to the approximate angle yielding the information we know about an arbitrary signal.

The algorithm of sound localization for the mean method of spectral comparison is described below:

1. For each degree interval $[n,n+1]$ check the $|ITDdiff_{n,c}|$, $|ITDdiff_{n+1,c}|$ and $|ITDdiff_{n,n+1}|$.
2. For the found values, if $|ITDdiff_{n,c}| \le |ITDdiff_{n,n+1}|$ and

$|ITDdiff_{n+1,c}| \leq |ITDdiff_{n,n+1}|$, remember the nth signal number and move to step 3. If not, we move to the next interval $[n+1, n+2]$ and return to step 1.

3. For the nth signal, calculate the difference $diff_n = \frac{1}{N}\Sigma_{i=1}^{N}|X_n(i) - X_0(i)|\,\frac{1}{N}\Sigma_{i=1}^{N}|X_n(i) - X_{180}(i)|$. Calculate in the same way $diff_{n+1}$ and $diff_c$.
4. If $diff_c \notin [diff_n, diff_{n+1}]$, leave the n value in the array with ITD information only (*ITD_only*) for further consideration (if it is needed), and move to the next interval $[n+1, n+2]$. If $diff_c \in [diff_n, diff_{n+1}]$, save the n value in array with signal indexes (L or R, in accordance with left or right ear, respectively) which satisfies both ITD and difference value requirements.
5. Complete steps 1–4 for each the left and right ears. Finally, there are two arrays $(L=\{l_1, l_2, \ldots\}, R=\{r_1, r_2, \ldots\})$ for each ear with the signal indexes which were found after ITD and difference matching. Find similar index numbers in arrays: $index = find(L=R)$.
6. If there is only one co-inside value $(\ell_i = r_i)$, the right number of the signal from our database has been found, the closest one to the cth.
7. If $L=\varnothing$ or $R=\varnothing$, head on the nonsingular array of signal indexes, either R or L.
8. If there are no matching indexes, compare $|ITDdiff_{i,c}|$ and $|ITDdiff_{j,c}|$, where $i \in L$, $j \in R$ and take the index from array with minimum ITD difference, i.e., $index = \min_{i,j}\{|ITDdiff_{i,c}|, |ITDdiff_{j,c}|\}$.
9. If both R and L arrays are empty and *ITD_only* consists of nothing, consider that ITD of the cth signal is either greater or smaller than other ITD values in the database. Find a suitable interval for this case and follow steps 3 and 4 to identify the location of a sound source.

IV. SIMULATION RESULTS

Three methods of spectral comparison have been studied. In addition to the mean method which is presented in formula (2), we also used the norm method and the coherence method. For the norm method, the difference between the two reference signals and the nth database signal can be written mathematically as

$$diff_n = \|X_n(i) - X_0(i)\| - \|X_n(i) - X_{180}(i)\|, \tag{3}$$

where $\|*\|$ denotes the Euclidian norm of the value. In the coherence method we use a function which is identical to the cross-correlation function in frequency domain. We calculate the maximum values between the obtained cross-correlation vectors. It can be written as:

$$corr_{n,0} = \arg\max_k \sum_i X_n(i)X_0(i+k); \tag{4}$$

$$corr_{n,180} = \arg\max_k \sum_i X_n(i)X_{180}(i+k); \tag{5}$$

$$diff_n = corr_{n,0} - corr_{n,180}. \tag{6}$$

While executing the algorithm, besides our own databases, we also examined MIT HRTF data. All the database signals were from 0° to 360° with different degree increments. Random signals lied in some location between the basic signals. For the reference signals first we used two database signals of 0 and 180° of azimuth location. Later we considered another set of reference signals. Generally, calculations were made in the azimuth plane. But there was also an attempt to estimate the effectiveness of the algorithm with changing elevation.

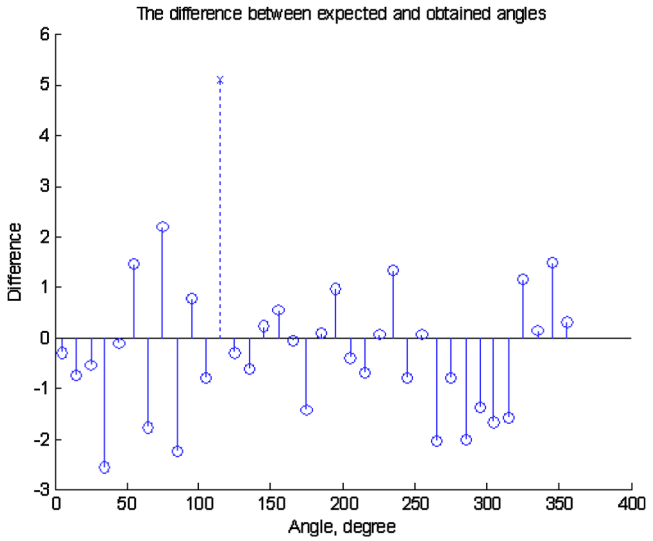


FIG. 5. (Color online) The information about found angular positions of random signals using coherence method. MIT HRTF data, 10° azimuth increment.

A. MIT HRTF database

It was of great interest to experiment with the efficiency of our algorithm when the ITD difference between neighboring blocks and random signal in the middle of them became small. At first we considered database signals started from 0° in 10° increment and random signals we took from 5° in 10° increment. The sound source's distance from the dummy head was 140 cm. Figure 5 depicts the results of degree differences between ideal and computed locations of random signals for coherence method. Lines with circled symbols correctly depict found angles. Dotted lines with x-marked symbols show found signals with wrong angles or signals which were not found at all. Relative errors for the whole set of random signals and the number of angles found correctly obtained with the MIT database are given in Table II. X there denotes a set of degree location of random signal HRTFs, X^*

TABLE II. The information about found angular positions of random signals using different methods. MIT HRTF database and impulse signal data in anechoic room with 10° azimuth increment with 10° azimuth increment.

Type of signal	Method	$\frac{\|X-X^*\|}{\|X\|}$	Correctly found angles
MIT HRTF	Mean	1.0703e-002	35/36
	Norm	1.4042e-002	34/36
	Coherence	7.0122e-003	35/36
Impulse signal in anechoic room	Mean	7.1918e-002	33/36
	Norm	1.2202e-002	34/36
	Coherence	4.9962e-003	36/36

TABLE III. The information about found angular positions of random signals using different methods. Random signals are not in the center of basic signals. MIT data with 20° azimuth increment.

Method	$\frac{\|X-X^*\|}{\|X\|}$	Correctly found angles
Mean	1.7811e-002	17/18
Norm	1.3947e-002	15/18
Coherence	9.6233e-003	18/18

- set of degrees obtained during the algorithm execution. We assume that the angle of a random signal has been found correctly if the angular difference between the found and the ideal value of a random signal degrees is less than the difference between the angular position of random and each of the neighboring basic signals, i.e., for MIT data that difference cannot exceed 5°. Here we can see that the coherence method performs with better results than two other methods.

Next, calculations have been made for the random signals not in the middle of neighboring basic signals. We tested all methods with MIT data. Database signals were taken as to begin from 0° in 20° increment. Random signals started from 5° in 20° increment also. Relative errors for the whole set of random signals and the number of angles found correctly are given in Table III. As we can find, all the random signals have been found correctly. The interesting thing here was that the 90° signal with highest ITD value was not in the database. However, the random signal which is between 80° and 100° was found correctly.

Let us now make the angular difference bigger and experiment with database signals started from 0° in 30° increments and random signals we took from 15° in 30° increments. It can be seen from the relative error results in Table IV that, once again, the coherence method behaved better and as a more stable method. However, this test shows us that the effectiveness of the algorithm decreased. The explanation can be taken out from the fact that the greater the azimuth difference between basic signals the more ambiguous intervals of spectral differences exist. As it is used in our algorithm, after finding suitable intervals we search for the proper interval in the spectral range. When those intervals for comparison become bigger, the confusion in finding the appropriate interval becomes less. We can conclude here that we need more basic signals in the database for more accurate sound localization.

TABLE IV. The information about found angular positions of random signals using different methods. MIT HRTF database, impulse, and clap signal data in anechoic room, and impulse signal data in lecture room with 30° azimuth increment.

Type of signal	Method	$\frac{\|X-X^*\|}{\|X\|}$	Correctly found angles
MIT HRTF	Mean	7.3528e-002	10/12
	Norm	3.3895e-000	9/12
	Coherence	3.3438e-003	11/12
Impulse signal in anechoic room	Mean	7.0156e-002	9/12
	Norm	6.1842e-002	10/12
	Coherence	5.6551e-002	10/12
Clap signal in anechoic room	Mean	9.3841e-002	9/12
	Norm	1.2840e-001	8/12
	Coherence	7.8143e-002	10/12
Impulse signal in lecture room	Mean	2.6267e-001	8/12
	Norm	3.4586e-001	7/12
	Coherence	8.9134e-002	10/12

TABLE V. The information about found angular positions of random signals using different methods and different elevation of the database signals. Elevation of random signals is −20°. MIT data with 10° azimuth increment.

Random signals elevation, degrees	Method	$\frac{\|X-X^*\|}{\|X\|}$	Correctly found angles
−20	Mean	7.5557e-002	32/36
	Norm	3.8548e-002	34/36
	Coherence	6.5829e-003	36/36
0	Mean	3.0907e-001	12/36
	Norm	3.8366e-001	7/36
	Coherence	4.4220e-001	8/36
20	Mean	8.9904e-001	15/36
	Norm	9.6928e-001	8/36
	Coherence	1.3426e-000	5/36

The last thing we would like to consider about the MIT HRTF database is the application of our algorithm for elevation changing. Database signals were taken as to begin from 0° in 10° increments. Random signals also started from 5° in 10° increments. We changed the elevation of database signals from −20° to 20° in 20° increments and used random signals with −20° elevation. The resulting table for this test is given in Table V.

It can be seen from Table V that the coherence method shows less stability to the elevation changing when basic and random signals are not in the same elevation plane. Quite to the contrary, the mean method presents the most stable results of the algorithm application. These results follow the fact that elevation changing is conducive to great modification in spectral range and an attempt to compute correct angle positions of random signals with spectral cues becomes unimportant. Thus, random signals can be localized correctly if and only if they are in the same elevation with basic signals.

B. Anechoic room experiment database

For our experimental data, we used sounds from the horizontal plane recorded in the same way we described in Sec. II. In this paragraph we consider data taken from the anechoic room. The distance between the sound source and the dummy head is 150 cm.

First, we took impulse database signals (Fig. 2) beginning from 0° in 10° increments and we used random signals starting from 5° in 10° increments. In that case each random signal was in the middle of two neighboring basic signals. Since 0° and 180° signals were in the database, we used them as reference signals we described before. Relative errors for the whole set of random signals and the number of angles found correctly for the experimental data are given in Table II.

Next, we took impulse database signals beginning from 0° in 30° increments and we used random signals starting

from 15° in 30° increments with other conditions the same as those used in the previous test. We also calculated random signals with the clap database (Fig. 3) with the same parameters we used for the impulse database. Relative errors for the experimental data with impulse and clap signals are given in Table IV.

If we compare three of the presented methods for all the listed cases, we see that the coherence method could find all the random signals. The difference in relative errors also shows that the coherence method works better.

C. Lecture room experiment database

In this paragraph we present results obtained from data taken from the original lecture room. We used the same impulse signals for recording as we used in the anechoic room. The significant thing here is that all the database and random signals were taken with ordinary noise and reverberations existing in the lecture room. It was of big importance to learn about the effectiveness of our algorithm on the recognition of signals in a real situation. The sound source's distance from the dummy head is 180 cm.

We recorded signals from the same positions we did in anechoic room. Database signals started from 0° in 30° increments and random signals began from 15° in 30° increments. In this test we also used reference signals in 0° and 180° of azimuth location. Relative errors for the whole set of random signals and the number of angles found correctly for the experimental data are given in Table IV.

Comparing three of the presented methods we can see that mean and norm methods show robust results whereas the coherence method provides us with more precise information about random signals' locations. The quantity of correctly found signals by the coherence method emphasizes its stability against other methods.

D. Further simulations

There has also been an attempt to apply our algorithm for different signals' distances from the dummy head. Except for the signals located 180 cm from the dummy head, we also recorded signals from 150 and 120 cm. With the same algorithm we have found that the angular difference was slightly changed. However, we could localize all the random sound sources we localized with 1.8 meter distance from the dummy head. The results for 120, 150, and 180 cm distance are depicted in Fig. 6.

In addition, we would like to mention that different results of sound localization can be obtained if we use different reference signals. To compute our results in previous tests we used reference signals in 0° and 180° of azimuth location. But what happens when we change reference signals? If we take reference signals in 0° and 150° of azimuth location, we will find one of the signals we could not find with previous reference signals and will not find one of the signals we could find with the previous ones. Again, if we take reference signals in 0° and 120° of azimuth location, we can see that all the signals which were not localized with 0° and 180° reference signals were found correctly.

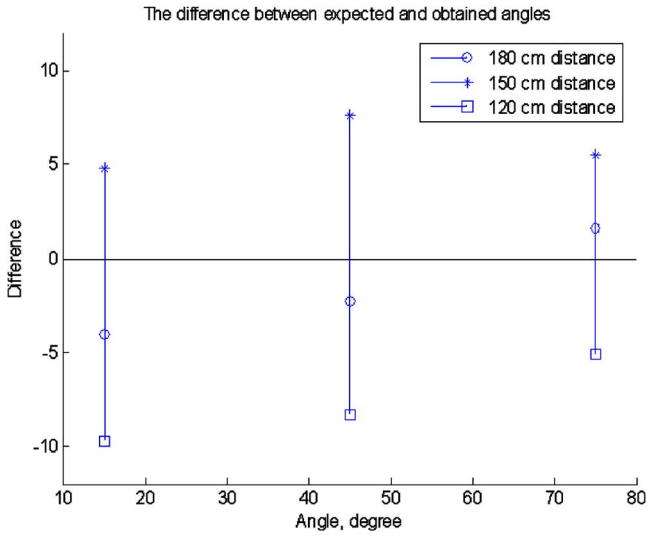


FIG. 6. (Color online) Lecture room experimented data, mean method with 30° azimuth increment. Random signals are in 15°, 45°, and 75° of azimuth location. Distance from the dummy head is varied from 120 to 180 cm in 30 cm interval.

In brief, using combinations of different reference signals we can localize all random signals. We can also use three or more reference signals in our algorithm for sound localization. However, this statement requires additional consideration and calculations.

V. CONCLUSIONS AND FUTURE DIRECTIONS

In this paper we proposed an algorithm for the localization of random signals using information about basic signals with sound localization cues such as ITD and spectrum analysis. We found that interaural cues are not sufficient for solving the front-back confusion problem. Thus, we used spectral information about the signals while creating our algorithm. ITD allows us to quickly get information about the approximate random signal location with front-back confusion error. One more advantage of ITD information is that it has a small data size and is relatively easy and quick to calculate. Further spectral analysis improves results obtained by ITD, and in combination they provide a good method for sound localization.

However, in the algorithm described above we need at least two reference signals from the database, preferably in opposite directions under the assumption that they give us more information about direction differences.

In this work we analyzed several methods of signal spectrum difference. It was found by the calculations that the coherence method in most cases gave the best results in comparison with all the other methods studied. However, with elevation changing, the mean method works better than norm and coherence methods.

We have also provided experimental results of our algorithm on those random signals which lie between neighboring basic signals but at different distances from the dummy head. As it is easy to see, even though the angular difference was slightly changed, we could localize all the random sound sources we localized with other distances and same azimuth degrees.

Changing elevation decreased the efficiency of the algorithm. These results follow the fact that elevation changing is conducive to great modification in spectral range. Thus, computing correct angle positions of the random signals with spectral cues becomes unimportant. However, the algorithm is stable if random signals have the same elevation as the basic ones.

The results of the experiment for sound localization of signals with noise are also provided. In spite of noise and reverberation additions to the signals, almost all of them could be localized correctly. We are confident that if we recorded more database signals in the lecture room it could be possible to localize all the random signals. So far, using the data from the anechoic room we can predict sound location with an error of one neighboring block.

While simulating experimental results we could conclude that we need more basic signals in the database for more accurate sound localization. The explanation can be taken from the fact that the more azimuth difference we have between basic signals the more ambiguous intervals of spectral differences we have. As it is used in our algorithm, after finding suitable intervals we search for the proper interval in the spectral range. When those intervals for comparing become bigger, the confusion in finding the appropriate interval becomes less.

Moreover, the assumption about linear behavior of the signal intervals is not entirely correct. The difference between the two reference signals and each database signal consists of a number of complex arithmetic and signal performances. Thus, the use of splines could make our calculated angles become closer to exact values.

Finally, the creation of a software application for the algorithm testing would be useful to simplify calculations for nonprogrammers. We can combine all the databases and algorithms into one program which will provide support for sound localization speed up. It can be underscored here that automation of the localization process reduces time the user has to spend executing the algorithm and finding appropriate results.

Future work in making the algorithm more stable is required. We need to provide additional analysis and calculations to find how combinations of different reference signals could be effective in order to localize all random signals. It would also be of great importance to create a system to recognize sounds with multiple locations of random sound sources using information about basic sound signals.

[1]D. R. Begault, *3-D Sound for Virtual Reality and Multimedia* (Academic, Professional, New York, 1994), pp. 1–82.

[2]C. I. Cheng and G. H. Wakefield, "Introduction to head-related transfer functions (HRTFs): Representations of HRTFs in time, frequency, and space," J. Audio Eng. Soc. **49**, 231–249 (2001).

[3]W. M. Hartmann, "How we localize sound," *Phys. Today* (**1999**). http://www.aip.org/pt/nov99/locsound.html. Last viewed: 04.18.2007.

[4]T. Lee, D. Jang, K. Kang, J. Kim, D.-G. Jeong, and H. Hamada, "A spatial audio system using multiple microphones on a rigid sphere," ETRI J. **27**, 153–165 (2005).

[5]Bruel & Kjaer, *Sound Quality Head and Torso Simulator Type 4100 & 4100D User Manual* (Bruel & Kjaer, 2001), pp. 2–8.

[6]R. O. Duda, "Modeling head related transfer functions," in *Proceedings of the 27th Annual Asilomar Conference on Signals, Systems and Computers*, Asilomar, CA, Vol. **2**, pp. 996–1000 (1993).

[7]M. Marolt, "Spatial audio synthesis," in *Proceedings of Multimedia and Hypermedia Systems Conference MIPRO'96*, Opatija, Croatia, pp. 119–122 (1996).

[8]J. C. Middlebrooks, J. C. Makous, and D. M. Green, "Directional sensitivity of sound-pressure level in the human ear canal," J. Acoust. Soc. Am. **86**, 89–108 (1989).

[9]F. Rumsey, *Spatial Audio* (Elsevier, Amsterdam, 200l), pp. 21–118.

[10]P. A. Hill, P. A. Nelson, O. Kirkeby, and H. Hamada, "Resolution of front-back confusion in virtual acoustic imaging system," J. Acoust. Soc. Am. **108**, 2901–2910 (2000).

[11]S. S. Jarng, Y. J. Kwon, and J. H. Lee, "Binoural directivity pattern simulation of the KEMAR head model with two twin hearing aid microphones by boundary element method," J. Acoust. Soc. Korea, **24**, 115–122 (2005).

Using a signal cancellation technique to assess adaptive directivity of hearing aids[a)]

Yu-Hsiang Wu[b)] and Ruth A. Bentler
Department of Speech Pathology and Audiology, University of Iowa, Iowa City, Iowa 52242

(Received 1 June 2006; revised 13 March 2007; accepted 5 April 2007)

The directivity of an adaptive directional microphone hearing aid (DMHA) cannot be assessed by the method that calls for presenting a "probe" signal from a single loudspeaker to the DMHA that moves to different angles. This method is invalid because the probe signal itself changes the polar pattern. This paper proposes a method for assessing the adaptive DMHA using a "jammer" signal, presented from a second loudspeaker rotating with the DMHA, that simulates a noise source and freezes the polar pattern. Measurement at each angle is obtained by two sequential recordings from the DMHA, one using an input of a probe and a jammer, and the other with an input of the same probe and a phase-inverted jammer. After canceling out the jammer, the remaining response to the probe signal can be used to assess the directivity. In this paper, the new method is evaluated by comparing responses from five adaptive DMHAs to different jammer intensities and locations. This method was shown to be an accurate and reliable way to assess the directivity of the adaptive DMHA in a high-intensity-jammer condition. © *2007 Acoustical Society of America.*
[DOI: 10.1121/1.2735804]

PACS number(s): 43.66.Ts, 43.66.Yw [DOS] Pages: 496–511

I. INTRODUCTION

The directional microphone is one of few hearing aid features that can improve speech intelligibility in noisy backgrounds (e.g., Bentler *et al.*, 2004a; Ricketts and Dhar, 1999; Walden *et al.*, 2000). It is designed to retain good sensitivity to signals arriving from the frontal direction (on-axis), while suppressing signals coming from other directions (off-axis). Many systems on current digital directional microphone hearing aids (DMHA) are first-order dual-microphone systems; that is, the spatial sensitivity is achieved by combining two matched omnidirectional microphones with an appropriate electronic delay and signal subtraction. After a certain electronic delay is programed, the null (the least sensitive angle) is fixed at a specific azimuth (Dillon, 2001). In contrast to the conventional fixed directional microphone, the adaptive null-steering directional system can move the null to different angles to achieve maximal noise attenuation. More recently, some DMHAs have been equipped with multichannel adaptive directional systems that allow several adaptive systems to operate independently in different frequency channels.

A. Quantifying microphone directivity

One typical way to describe the directivity of a directional system is the polar plot, which depicts the spatial characteristic of the system in a single plane (e.g., *x-y* Cartesian coordinates). The polar plot records the attenuation pattern of a directional system in different azimuth angles at a given frequency. Four basic theoretical attenuation patterns (i.e., polar patterns) of the first-order directional system are commonly mentioned in the literature: dipole (with null at 90° and 270°), hypercardioid (with null at 110° and 250°), supercardioid (with null at 125° and 235°), and cardioid (with null at 180°) (Dillon, 2001; Valente *et al.*, 2000). Different internal electronic delays result in different polar patterns. It is important to note that these theoretical patterns can only occur with a DMHA in the free field, and do not occur when a DMHA is fitted on the head. Three-dimensional (3D) spatial directivity patterns can be reported by polar plots at several elevation angles or 3D polar balloon plots.

Another way to describe directivity is by using a single number, such as the directivity index (DI) (for review, see Ricketts, 2000; Ricketts and Mueller, 1999). The definition of the DI is the ratio (in decibels) of the sound pressure of the microphone on-axis response to a sound source in a free sound field to the sound pressure of the microphone off-axis response in a diffuse sound field (Beranek, 1949). A free sound field is "unbounded, free of external interference and objects that would introduce diffraction and reflections." A diffuse sound field is one in which "the sound pressure level is uniform and the angle of incidence of a sound wave is randomly distributed within the field at any instant in time." (Ghent, 2005) Thus, the on-axis response of an omnidirectional microphone to a single noise source in a free field would be equal to the off-axis response of the same microphone in a diffuse sound field resulting in a DI of 0 dB (Ricketts, 2000). A directional system may have different DI values at different frequencies. Directivity in such a system can be described by plotting DI as a function of frequency.

The directivity of a DMHA is traditionally obtained in an anechoic chamber by presenting a signal (i.e., the probe signal) from a probe loudspeaker at different elevation angles relative to the DMHA, which rotates by a turntable to differ-

[a)]Portions of this work were presented at the annual convention of the American Academy of Audiology, Minneapolis, MN, April 2006.
[b)]Electronic mail: yu-hsiang-wu@uiowa.edu

 0001-4966/2007/122(1)/496/16/$23.00

ent azimuth angles discretely (ANSI, 2004; Ricketts and Mueller, 1999). Although multiple loudspeakers may be employed, only a single loudspeaker is active at any given point in time. Therefore, this method is referred to as the single-loudspeaker method in this article. The response of the DMHA to the probe signal (i.e., the probe response) at each angle is recorded, and then subtracted from the probe response at 0° (i.e., the on-axis response) to derive the directional response (ANSI, 2004). The frequency-specific directional responses at different azimuth and elevation angles can be used to calculate the 3D DI. Compared with the 3D DI, the planar DI is a simplified measurement in which only directional responses on the horizontal plane at 0° elevation are taken into account. The frequency-specific directional responses on the horizontal plane can also be used to construct the polar plot. The planar DI is calculated by the equation

$$\mathrm{DI} = 10 * \mathrm{Log}_{10}\left[\frac{1}{\pi/(2n)\sum_{j=1}^{n} 10^{R(0,\theta_j)/10}|\sin\,\theta_j|}\right], \qquad (1)$$

where n is the number of measurements in the 360° azimuth, θ is the azimuth angle, and R is the directional response at the indicated azimuth. The DI can be derived from measurements of the DMHA suspended, in isolation, in a free sound field (i.e., the free-field DI), or from measurements of the DMHA on a manikin in a free sound field (i.e., the simulated real-ear aided DI, SREADI; ANSI, 2004). Theoretically, the maximal free-field DI for a first-order directional system is around 6 dB in a hypercardioid polar pattern (Valente *et al.*, 2000).

DI can be used to compare different directional microphones. Mean SREADI (3D or planar) across frequencies correlates to directional benefit for the average hearing aid user in diffuse noisy listening environments (Dittberner and Bentler, 2007; Ricketts *et al.*, 2005). In order to predict the outcome of a subjective test of directional performance, applying frequency-specific importance weightings to the DI values before averaging across frequencies has been suggested (e.g., Killion *et al.*, 1998). The weightings most often used are based on the band importance function, which has been used in calculating speech intelligibility index (SII) for years (ANSI, 1997). This weighted average DI is called the speech-intelligibility weighted directivity index (SII-DI) in ANSI S3.35-2004. Although it is logical to apply the weighting function, Ricketts and his colleagues (2005) found that the differences between unweighted linearly averaged DI and SII-DI were very small. They also demonstrated that the choice of weighting functions (flat or band importance function based) did not improve the accuracy of prediction. Therefore, a simple average DI without weighting is suggested (ANSI, 2004).

B. Assessing adaptive directional microphone systems

Because the probe responses at different angles are collected sequentially in the single-loudspeaker method, an important underlying assumption is that directivity of the DMHA is time invariant during DMHA rotation. This assumption is generally true for the conventional fixed directional microphone because its polar pattern does not change regardless of the angle from which the sound is presented. However, this assumption does not hold for the adaptive DMHA.

Adaptive refers to the ability of the directional system to change the polar pattern in different noise configurations. Recall that, in a dual-microphone system, different polar patterns can be achieved by changing the internal electronic delay between the outputs of two microphones. The algorithm of the adaptive DMHA continually updates the internal delay to provide the greatest reduction of noise, placing the nulls at appropriate azimuth angles (usually between 90° and 270°, the back hemisphere). Therefore, the polar pattern of an adaptive system depends on the location of the signal. It also depends on the intensity and characteristics of the signal, how the system detects and analyzes the signals, and the decision-making rules used to determine the appropriate polar pattern. For example, adaptive systems may determine the polar pattern by adopting the internal electronic delay that gives the lowest power output from the directional microphones. Alternatively, the hearing aid may estimate the signal-to-noise ratio (SNR), and choose the electronic delay that generates the highest SNR (for review, see Chung, 2004). This adaptivity is achieved independently in different frequency channels in a multichannel adaptive DMHA. Moreover, many adaptive DMHAs can switch not only between different polar patterns in the directional mode, but also between omnidirectional and directional modes (i.e., automatic adaptive directional system). As a result, an adaptive DMHA may generate different polar patterns in different noise configurations. Even with the same noise configuration, different adaptive DMHAs may not necessarily generate the same polar pattern. Due to this adaptivity, it is difficult to assess the directivity of adaptive directional systems. Specifically, in the single-loudspeaker method, because the probe signal is presented at different azimuth angles one by one, the probe signal itself changes the polar pattern when the probe is presented between 90° and 270°. For instance, when the probe signal moves from 90° to 180° relative to the DMHA in a free field, the polar pattern may change from a dipole to a hypercardioid, and then to a cardioid pattern (Chung, 2004; Powers and Hamacher, 2004). Therefore, the single-loudspeaker method is invalid for the adaptive DMHA. The free-field polar pattern derived from this method is not consistent with any theoretical pattern, and the DI is always overestimated. That is why in ANSI S3.35-2004, the DMHA needs to be in a fixed directional mode.

The rationale of the adaptive directional system is to improve speech intelligibility by moving the null to appropriate angles. Although the effectiveness of these instruments has not been observed in field trials (Palmer *et al.*, 2006), the efficacy of the adaptive DMHA has been proven in some laboratory environments (Bentler *et al.*, 2004b; Ricketts and Henry, 2002). Therefore, many manufacturers currently implement adaptive directional technology on newly designed hearing aids.

Despite the fact that much research has been conducted studying the adaptive DMHA and many manufacturers adopt

adaptive directional technology, there is no standard way of assessing, verifying, and comparing adaptive directional systems. In personal communications, Ricketts (2006) suggested using "average angular attenuation," which is the average amount of attenuation to a discrete noise source from all azimuth angles, to describe the average performance of an adaptive DMHA. Dittberner and Bondy (2006) proposed another method to assess the DI of the adaptive DMHA on a Knowles electronics manikin for acoustic research (KEMAR) in a simulated diffuse field, which is created by eight off-axis loudspeakers. The DI is obtained by comparing the DMHA's response to one on-axis loudspeaker to the DMHA's output to the eight off-axis loudspeakers. Although these two methods probably can estimate the average performance, neither of them can describe the directivity that an adaptive system can produce across different noise configurations. For example, what is the polar pattern if a single loud noise is presented at 150° relative to the DMHA? What is the polar pattern if a second less intense noise source is added at 270°? An adaptive directional system can generate almost infinite directivity patterns to adapt to different noise configurations. How can we verify this adaptivity? How can a researcher explain the difference between directional benefits provided by different adaptive DMHAs if he/she does not know the exact polar patterns that the DMHAs produce in the test environment? Because, as mentioned, different adaptive systems may not generate the same polar pattern in the same environment, it is crucial to understand how an adaptive directional system reacts to a certain noise configuration. Such information is important in verifying the adaptivity of the DMHA. It is also important for researchers when designing studies on the adaptive DMHA. At present, this information is only available to manufacturers who have the ability to read from the integrated circuit (IC) chips in DMHAs, and the results may not be open to the public. Therefore, a system that can acoustically assess the directivity of the adaptive DMHA under a given noise configuration is necessary. This method should be (1) accurate; (2) reliable; (3) able to assess most adaptive DMHAs; and (4) able to assess directivity in any noise configuration.

In this article, as a first step, we propose a new method that can measure the planar polar pattern of the adaptive DMHA in high-intensity-noise conditions.

C. Theory

To assess the directivity of the adaptive system, the noise configuration must first be defined. In this given condition, the polar pattern will be driven to a certain configuration. If the noise intensity is high enough, the polar pattern may be "frozen;" in other words, the pattern may not change if a second low-intensity sound is presented simultaneously. Therefore, a low-intensity probe signal that is presented at different angles may not change the polar pattern if a high-intensity noise is presented to the DMHA from a certain angle. In the proposed method, at least two loudspeakers are needed: a probe loudspeaker to present the low-intensity probe signal from different angles, and a "jammer loudspeaker" to present a high-intensity jammer signal at a fixed angle relative to the DMHA. If the probe response at each azimuth can be recorded, the polar pattern can be obtained. However, because the probe and jammer signals are presented simultaneously, the probe response and the jammer response are mixed together in the output of the DMHA. In fact, the probe response will be embedded in the jammer response because the probe intensity is much less than the jammer intensity.

In order to extract the probe response from the output of the DMHA, a special signal cancellation technique is utilized. This technique was successfully used to investigate the noise reduction algorithms of hearing aids (Hagerman, 2004). Specifically, two signals are presented to the hearing aid sequentially. They are

$$A_{\text{in}} = u(t) + v(t)$$

and

$$B_{\text{in}} = u(t) - v(t),$$

where $u(t)$ is the speech signal, $v(t)$ is the noise signal, and t is time. The speech signals in A_{in} and B_{in} are the same, whereas the noise signals are phase-inverted in these two input signals. The responses of the hearing aid will be

$$A_{\text{out}} = u'(t) + v'(t) + e_1(t)$$

and

$$B_{\text{out}} = u'(t) - v'(t) + e_2(t),$$

where $u'(t)$ is the output speech signal, $v'(t)$ is the output noise signal, and $e_1(t)$ and $e_2(t)$ are error signals. The output speech signal can be extracted as following:

$$(A_{\text{out}} + B_{\text{out}})/2 = u'(t) + [e_1(t) + e_2(t)]/2. \quad (2)$$

The output noise signal can be extracted in a similar way. With this technique, Hagerman (2004) calculated the SNR of input and output of hearing aids, and determined the performance of the noise reduction algorithm.

This technique can be used in our new method to extract the probe response. Instead of speech and noise signals, we use different signals: probe and jammer signals (their characteristics are discussed below). Specifically, $u(t)$ is the probe signal and $v(t)$ is the jammer signal. Different from Hagerman's method, which presents $u(t)$ and $v(t)$ from the same loudspeaker, the probe signal $u(t)$ is presented from the probe loudspeaker, and the jammer signal $v(t)$ is presented from the jammer loudspeaker simultaneously. With the same procedure, the response of the jammer signal $[v'(t)]$ can be canceled out. Then, the embedded probe response $[u'(t)]$ can be extracted at each angle, and used to construct the polar plot. Because of this signal cancellation technique, the new method is referred to as the SC method in this article.

II. METHOD FOR EXPERIMENT 1

Recall that the new method should be (1) accurate; (2) reliable; (3) able to assess most adaptive DMHAs; and (4) able to assess directivity in any noise configuration. The procedures were designed to examine how the SC method meets the first three criteria by testing five adaptive DMHAs avail-

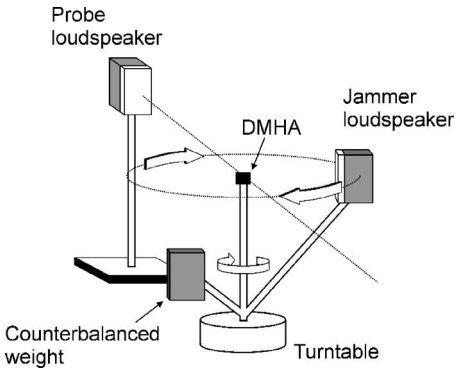


FIG. 1. Equipment setup (DMHA: directional microphone hearing aid).

able on the market. The SC method, by definition, fails the fourth criterion as it is designed to test the adaptive DMHA with high-intensity noise.

A. Equipment setup

Although ANSI S3.35-2004 requires elevated probe loudspeakers for 3D DI measurements, in order to simplify equipment setup and data interpretation, only planar DI was measured in this study. The DMHA was mounted on an Outline ET1.1 turntable without KEMAR and was positioned at the center of an anechoic chamber with 0° elevation. The rationale of this free-field measurement will be discussed later. The turntable rotation was controlled via an external controller and a personal computer. The azimuth resolution of the turntable was set to 10°. The probe loudspeaker was fixed in position, and was on the same horizontal plane as the DMHA. The distance between the DMHA and the probe loudspeaker was 1.1 m. A second loudspeaker to present the jammer signal was installed on the turntable so as to rotate synchronously with the DMHA. An illustration of this setup is shown in Fig. 1. With this arrangement, the probe signal can be presented from the probe loudspeaker at different azimuth angles, and the jammer signal can be presented from the jammer loudspeaker at a fixed angle during the measurement. Although it is desirable to place the jammer loudspeaker on the same horizontal plane with the DMHA and the probe loudspeaker, it is not possible to do so because one loudspeaker will block the other when both loudspeakers are at the same azimuth angle. Therefore, the jammer loudspeaker was placed approximately 7° below the horizontal plane. The distance between the DMHA and the jammer loudspeaker was 0.9 m. Two Tannoy i5 AW point-source loudspeakers served as the probe and jammer loudspeakers.

The probe and jammer signals were generated by ADOBE AUDITION 1.0, and output from a RME DIGI96/8 PAD sound card with a 44.1-kHz sample rate and 16-bit resolution. From the sound card, the probe and jammer signals were routed to the probe and jammer loudspeakers, respectively, via an ART 355 equalizer and an ADCOM GFA5002 amplifier. A G.R.A.S. IEC 126 2-cc coupler and a Larson-Davis 2560 1/2-in. random incidence microphone were used to collect the DMHA output. The recorded signals were routed through a Larson-Davis preamplifier and a LISTEN SoundConnect microphone power supplier to the RME sound card. The sound card digitized the signals for further analysis.

B. Stimuli and signal analysis

A white noise was used as the jammer signal because this wideband signal can drive the polar pattern at all frequencies simultaneously. In order to obtain data across frequencies in one measurement, a wideband noise was also chosen as the probe signal. The selection of the intensity and spectrum of the probe signal is critical. For a given jammer intensity, the probe intensity should be as low as possible so as not to change the polar pattern. On the other hand, in order to assess the null of the polar pattern, the probe intensity should be high enough to provide an appropriate probe-to-noise floor ratio across frequencies. That is, the intensity of the extracted probe response at 0° must be higher than the noise floor by a certain amount; otherwise, the noise floor will obscure the null. Generally, a 20- to 25-dB probe-to-noise floor ratio is required to detect the null. Therefore, the probe intensity should be within a certain range. The upper limit of this range is the highest intensity that does not change the polar pattern driven by the jammer signal, and the lower limit is the lowest intensity that can achieve the 20- to 25-dB probe-to-noise floor ratio. Since the upper limit is difficult to determine, the probe signal is always set to the lower limit in the SC method. By choosing this lowest probe intensity, the possibility that the probe changes the polar pattern is minimized. The noise floor was obtained by the same signal cancellation technique, with the jammer signal presented from the angle at which the jammer loudspeaker would be in the measurement. After canceling out the jammer response, the remainder (i.e., the noise floor) was compared to the probe response at 0° to compute the probe-to-noise floor ratio. Note that the noise floor is the summation of the residual jammer signal after accounting for cancellation, background noise, and the internal noise of the DMHA. Therefore, the spectrum and intensity of the probe signal depend on the spectrum, intensity, and location of the jammer signal. The probe signal is DMHA-dependent also. In this study, each probe signal was determined after the intensity and azimuth of the jammer were given. A white noise was spectrally shaped to achieve the 20- to 25-dB ratio across frequencies to serve as the probe signal.

During the measurement, the probe signal was presented about the 360° horizontal plane in 10° increments. At each azimuth angle, the first signal set (A_{in}) was presented: probe signal $[u(t)]$ from the probe loudspeaker, and the jammer signal $[v(t)]$ from the jammer loudspeaker. Before the probe signal presentation, the jammer signal was presented for 20 s to ensure the directivity pattern of the DMHA. From a pilot study,[1] it was determined that a 20-s interval was long enough for the null placement of most DMHAs to stabilize. After 20 s, the probe signal was presented with the jammer simultaneously for 12 s. After a 2-s silent interval, the second signal set with the inverted jammer phase (B_{in}) was presented with the same order. The outputs of the DMHA (A_{out} and B_{out}) were recorded and the probe responses $[u'(t)]$ were extracted by Eq. (2). The middle 10-s probe response was converted from the time to the frequency domain using the

MATLAB fast Fourier transform (FFT) function. Then, the one-third-octave power across the bandwidth of 0.4 to 6.3 kHz was computed at each azimuth and used to create the planar polar plot.

C. Evaluation procedures

1. Accuracy

Ideally, the easiest way to assess the accuracy of the SC method is to freeze the polar pattern of an adaptive system in a given jammer condition by proprietary software from the manufacturer, and compare this polar pattern to the result generated by the SC method. However, because the SC method will be used when the proprietary software is not available, methods to assess accuracy without proprietary software should be developed. Instead of using software to freeze the polar pattern, we decided to pursue an alternative way to assess the accuracy of the SC method.

Although ANSI S3.35-2004 indicates that the DMHA is to be tested on a KEMAR, in order to investigate the accuracy of the SC method, the DMHAs were tested free field on a 2-cc coupler. From these measures, the shape of the polar pattern can be used to evaluate accuracy, at least for the first-order adaptive directional system (the most common system on the hearing aid market). If a polar pattern is not consistent with any known theoretical pattern, such as a hypercardioid pattern, it can be inferred that the result is not accurate: the probe has changed the polar pattern during the measurement. Another way to evaluate accuracy is by calculating the planar DI, which can be obtained by Eq. (1). If the measured DI is larger than the theoretical upper limit of a directional system in a free field, we can conclude that the result is inaccurate. It should be noted that the DI is defined as the ratio of the on-axis response in a free field to the off-axis response in a diffuse field. For an adaptive DMHA, the DI represents the DMHA's directivity that adapts to a diffuse sound field. However, the polar pattern from the SC method is the DMHA's response to one or more discrete high-intensity jammer and a single low-intensity probe signal. Although the same calculation can be done, it is not appropriate to call the result a DI. We will refer to this measurement as "polar index" (PI) in this article. The polar index cannot be used to quantify the performance of a directional microphone system in a diffuse field, and it will not correlate to the average directional benefit of hearing aid users in a diffuse noisy environment. Yet the polar index can be used to evaluate the accuracy of the SC method. Specifically, if the jammer loudspeaker is placed at 110° to make the DMHA generate a hypercardioid polar pattern, which has the highest DI value among the polar pattern family of the first-order directional microphone, and if the probe signal does not alter the polar pattern, then the polar index should be lower than its theoretical value: 6 dB. If a polar index larger than 6 dB is observed, then we can conclude that the polar plot is not accurate. For this reason, all the DMHAs were tested with a single jammer signal presented at 110° to evaluate accuracy. A 75-dB-SPL white noise was chosen as the jammer signal because the intensity is high enough to drive most adaptive directional systems (75-dB, 110°-jammer condition).

The accuracy of the SC method can also be evaluated by the extracted probe response, $u'(t)$. As mentioned, the probe signal is presented for 12 s at each azimuth. If the amplitude of the probe response is the same from the beginning to the end, it is reasonable to assume that the probe signal does not change the polar pattern. On the other hand, if the system changes the polar pattern adaptively after the presentation of the probe signal, the amplitude of the beginning of the probe response will be larger than the amplitude of the end of the probe response. In order to compare the amplitude of the probe response, the waveform between 5 and 10 ms from the start of the probe signal and the waveform during the last 5 ms of the probe signal were designed to be identical. After extracting the probe response, the root-mean-square (rms) amplitude of these two intervals (rms_{5-10} and rms_{L5}) were calculated and compared. The difference between them (rms_{5-10} minus rms_{L5}, in decibels) is called the error value in this article. The 5-ms duration was chosen because most DMHAs require more than 10 ms to integrate information and adapt their polar patterns (Chung, 2004). If a longer duration is used, the error value may not reflect the change of the polar pattern. In the design stage of the SC method, we found that the output waveform of the first 5 ms of the probe response was not stable for some DMHAs. Thus, the interval from 5 to 10 ms was used instead. The rms_{5-10} provides a rough estimation of the directivity that is not affected by the probe signal, and the rms_{L5} represents the directivity after probe presentation. If the probe signal does not alter the polar pattern at all, the error value will be zero. The output waveforms recorded between 5 to 10 ms and the last 5 ms of the extracted probe response were filtered by an octave filter bank centered at 0.5, 1, 2, and 4 kHz. Then, the rms_{5-10} and rms_{L5} were calculated for each frequency band at each azimuth to obtain the frequency-specific error value. Next, the error values were added back to the measured probe response at the corresponding frequency to obtain the "corrected probe response." For example, if the measured probe response at 1 kHz is 30 dB, the rms_{5-10} is 35 dB, and the rms_{L5} is 25 dB, then the error value will be $35-25=10$ dB, and the corrected probe response at 1 kHz will be $30+10=40$ dB. The "corrected polar index" calculated by the corrected probe response was then used to evaluate the accuracy of the SC method. If the measured and corrected PIs are very close, then it is reasonable to assume that the SC method is accurate. For convenience, different error values at different azimuth angles are summarized by a single number. In this article, it is referred to as the "error index" (EI). The EI is the root-mean-square (rms) level of the error values across all azimuth angles in a certain frequency band. The EI can be expressed by the equation

$$\text{EI} = \sqrt{\frac{\sum_{j=1}^{n} E_j^2}{n}}, \tag{3}$$

where n is the number of measurements in the 360° azimuth, and E is the error value at each azimuth angle. The EI can be used to quantify result validity, i.e., the lower the EI, the more accurate the result.

TABLE I. The characteristics of the test hearing aids.

	Microphone order	Channel number of the adaptive directional system	Others
HA1	First order below 1 kHz Second order above 1 kHz	Four channels	Right-ear hearing aid
HA2	First order	Channel-less	
HA3	First order	Twenty channels	
HA4	First order	Single channel	
HA5	First order	Four channels	

In brief, metrics (i.e., the accuracy metrics) were used to examine the accuracy of the SC method: the shape of the polar pattern, the PI value, the difference between the measured and corrected PI, and the EI value.

2. Reliability

To examine the reliability of the SC method, five consecutive measures of each DMHA were obtained. The DMHA was removed from the coupler, replaced, and repositioned for each of the repeated measures. The standard deviation of PI values and the standard deviation of directional responses at each azimuth were calculated for each DMHA. The repeated measurements were done in the 75-dB, 110°-jammer condition. If the standard deviations were small (e.g., smaller than 0.1 dB), the SC method was considered reliable.

3. Generalizability

Five hearing aids equipped with adaptive directional systems were tested, here denoted as HA1–HA5. All hearing aids were behind the ear style. The characteristics of the adaptive directional systems are listed in Table I. HA1 consists of two different systems: a first-order system at frequencies below 1 kHz, and a second-order system above 1 kHz. Unlike the other DMHAs, the microphone ports of HA1 were isolated to the right side of the hearing aid case (i.e., right-ear hearing aid). All the DMHAs were programed to fit a flat 50-dB hearing loss by manufacturers' fitting software, and then modified into a linear gain mode. The adaptive directional mode was enabled. Features other than the directional microphone, such as noise reduction and feedback cancellation algorithms, were disengaged. The generalizability of the SC method was considered good if accurate and reliable results could be obtained from all DMHAs.

Each DMHA was mounted on a 2-cc coupler in free field and tested following the SC method in the 75-dB, 110°-jammer condition, five separate times. The DMHAs were tested in three additional conditions: (1) with a 60-dB-SPL white noise jammer at 110° (60-dB, 110°-jammer condition); (2) without any jammer signal (no-jammer condition); and (3) with a 75-dB-SPL white noise jammer located at 180° (75-dB, 180°-jammer condition). The purpose of the first two conditions was to examine whether the SC method could be used with less intense jammer signals. The no-jammer condition is equivalent to the traditional single-loudspeaker method. The purpose of the third condition was to verify the adaptivity of these DMHAs.

The spectra and the intensities of the probes were determined by the jammer intensity and location according to the principle mentioned before. The probe intensities for the different conditions are listed in Table II. A test was conducted to ensure the DMHA worked linearly before measurement. Equipment was calibrated daily. All tests were conducted in a 30-cubic-ft. anechoic chamber that conforms to ANSI S3.35-2004. An algorithm written in MATLAB controlled the signal presentation, turntable rotation, and probe response extraction. The extracted probe responses were then used to construct the polar plot.

Although the polar plot is a good way to represent spatial directivity, a single polar plot can only demonstrate directivity at a limited number of frequencies. In other words, it cannot display spatial directivity in continuous frequency. Therefore, we also present our data by the "polargram," which provides more directivity information across frequencies. The polargram is essentially a contour plot representing the attenuation of the DMHA as a function of azimuth angle and frequency. The polargram can display directivity in a continuous fashion for both azimuth and frequency.

III. RESULTS OF EXPERIMENT 1

A. 75-dB, 110°-jammer condition

Each DMHA was tested five times in this condition. The results are shown in Fig. 2. Polar plots that display the spatial attenuation pattern at frequencies of 0.5, 1, 2, and 4 kHz are shown on the left side. The center of the circle corresponds to −25 dB, and the radius of the outermost circle corresponds to +5 dB. The arrow at 110° of the outermost circle indicates the azimuth of the jammer signal. The polargram is shown on

TABLE II. The probe intensity (dB SPL) of different hearing aids and jammer conditions.

	Jammer condition			
	75 dB, 110°	60 dB, 110°	No jammer	75 dB, 180°
HA1	61.3	59.4	59.4	61.3
HA2	57.5	55.7	55.1	56.2
HA3	63.2	57.4	56.0	63.2
HA4	60.4	58.6	52.1	60.4
HA5	59.0	57.7	55.0	59.0

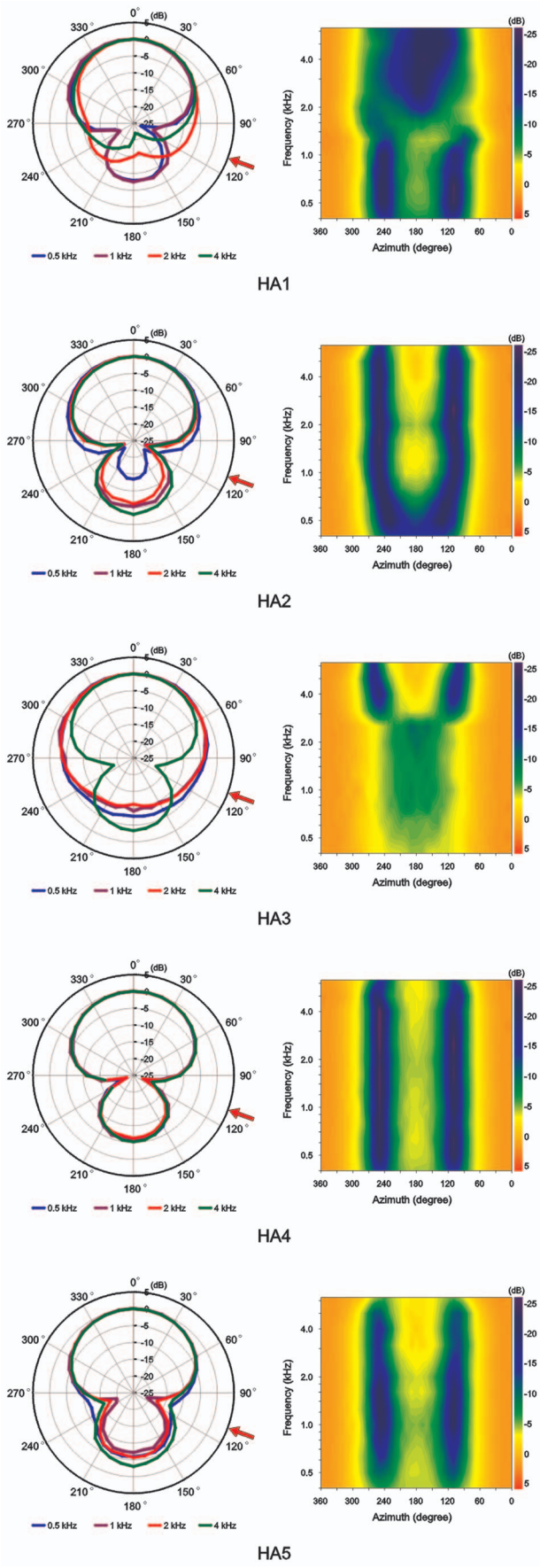


FIG. 2. The directivity of five adaptive DMHAs in the 75-dB, 110°-jammer condition. The polar plots are shown on the left side and the polargrams are shown on the right side of each panel. The arrow indicates the azimuth of the jammer signal.

the right side of each panel. The polargram displays the directional response as a function of azimuth and frequency. The *x* axis is the azimuth, and the *y* axis is the frequency (0.4 to 6.3 kHz). Note that the *x* axis is reversed (360° to 0°, from left to right) in order to be more consistent with the direction of azimuth angles of the polar plot (clockwise). The color represents the directional response, from −25 dB (dark purple) to +5 dB (orange), as indicated by the bar over the right side of the polargram. The polar plot and polargram present the same data in different ways. For example, the polar pattern of HA4 on the polar plot is a hypercardioid pattern (Fig. 2), which corresponds to two dark purple columns across frequencies at 110° and 250° in the polargram.

From Fig. 2, one can see that the responses to the 110° jammer are not the same for all DMHAs. Typical hypercardioid polar patterns, which were consistent with the jammer from 110°, were found in HA4 and HA5 at all frequencies. These adaptive systems steered the null accurately to 110° and 250°. This pattern held true in HA2 except at 500 Hz.

Different patterns were noted in HA1 and HA3. For HA1, the polar patterns differed from the hypercardioid pattern in three ways. First, the polar patterns at frequencies below 1 kHz and above 2 kHz were different. This can be seen clearly from the polargram. This is consistent with a first-order system at frequencies below 1 kHz, and a second-order system at frequencies above 1 kHz. Second, although a hypercardioid-like pattern was found at low frequencies, it was not symmetric. The whole curve seems to rotate 10° counterclockwise, and the null at the left side points to 240°, instead of 250°. The polar patterns at high frequencies were also asymmetric. It is possible that the directivity of this right-ear DMHA has been optimized to achieve the best directivity on the human head by placing the most sensitive azimuth angle toward the left side. Third, the shapes of the polar patterns at high frequencies were uncommon, and the nulls did not point to the jammer signal. Instead, the nulls were between 150° and 180°.

The polar patterns of HA3 were also unusual. At frequencies above 3 kHz, they were the typical hypercardioid pattern. However, there was no apparent null pointing to the azimuth of the jammer at frequencies below 3 kHz, resembling instead a mixture of a cardioid and an omnidirectional pattern. The above measurements were not adequate to determine the true directivity of HA1 and HA3.

The mean PIs at 0.5, 1, 2, and 4 kHz of five consecutive measures of each DMHA are shown in Fig. 3. Except for the PI at 4 kHz of HA1, which has the second-order microphone system at high frequencies, all the PIs were below 6 dB. The mean error values of the five measures at each azimuth angle are shown in Fig. 4. A positive error value means that the amplitude of the beginning of the probe response is larger than the end of the probe response, and it suggests that the system steers the null to reduce the sensitivity to the probe signal. As expected, the error values around 110° and 250° were more variable. There are two reasons for this outcome. First, when the probe signal is presented around the jammer signal, the adaptive system tends to move the null to the probe signal because a small shift of the null can cause a big change in the probe response. Second, when the probe ap-

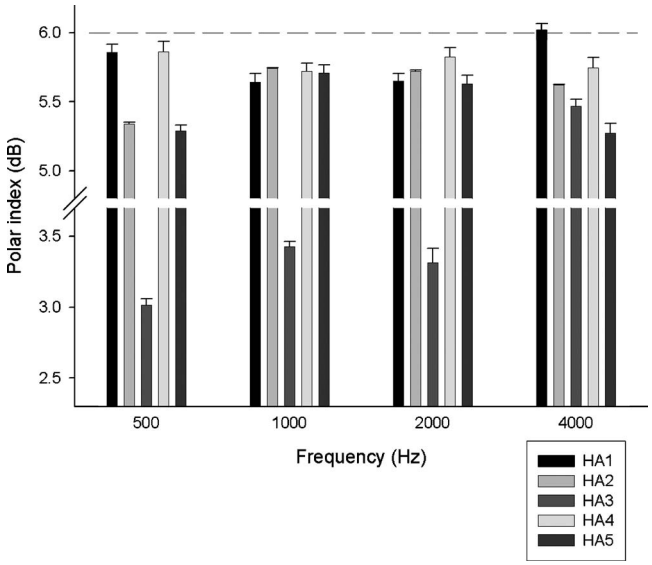


FIG. 3. The mean polar indexes as a function of frequency. The error bars represent 1 standard deviation. The dashed line represents the theoretical upper limit of the free-field polar index of the first-order directional microphone.

proaches the null, the probe response is much attenuated and close to the noise floor. Thus, the probe response (and the error value) is easily contaminated by the noise floor, especially at low frequencies. Similarly, the error values at

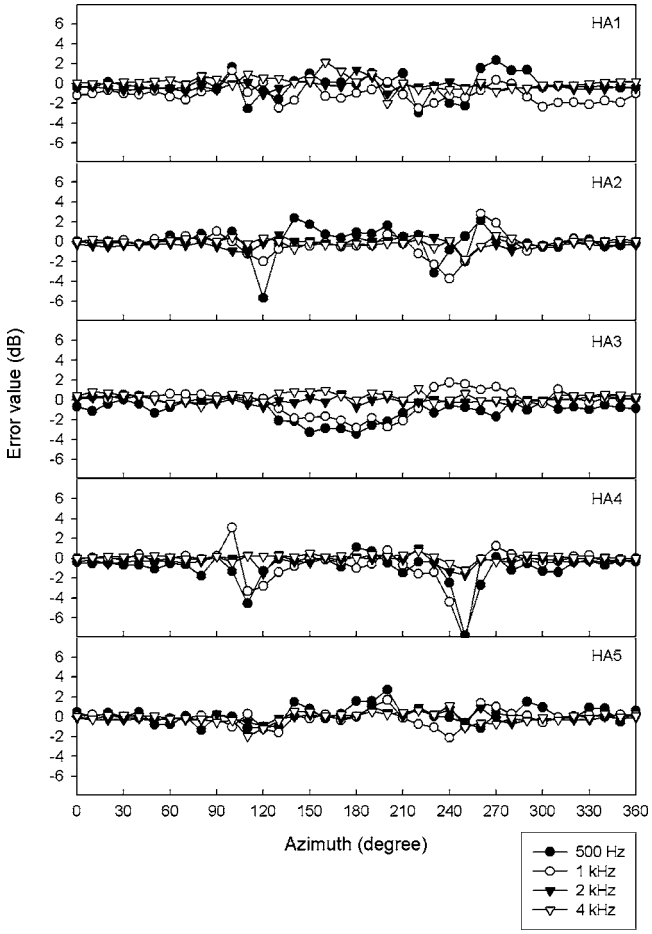


FIG. 4. The mean error values as a function of azimuth angle.

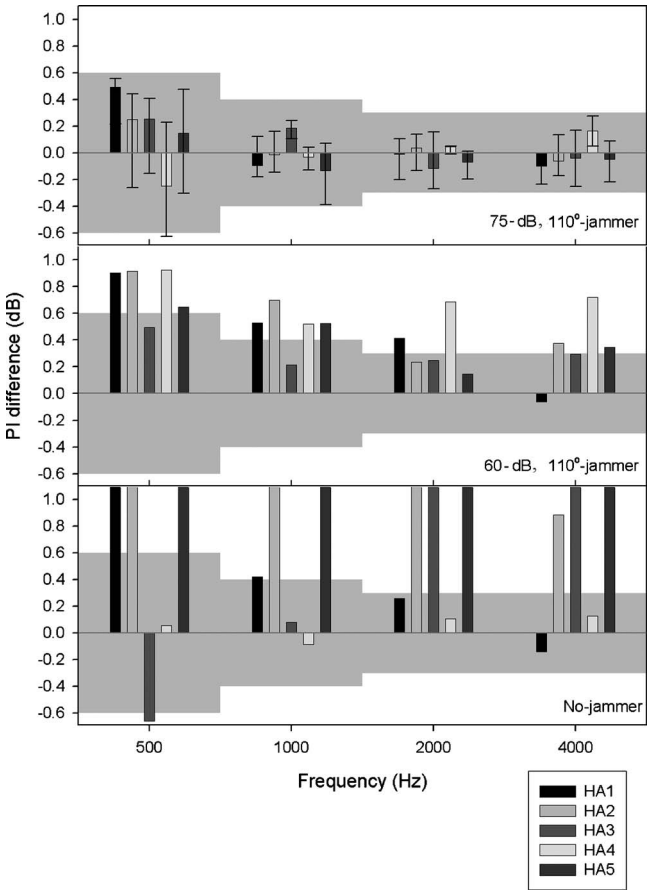


FIG. 5. The difference between measured PI and corrected PI as a function of frequency in the 75-dB, 110°-jammer, 60-dB, 110°-jammer, and no-jammer conditions. The error bars in the upper panel represent the data range from five measures. The shaded areas represent the norm. The data beyond the scale of y axis are truncated. (PI: polar index).

500 Hz (solid circle) were more easily contaminated by low-frequency background noise, and generally fluctuated more than at other frequencies. Large error values at azimuth angles between 120° and 240° were noted in HA3 at 0.5 and 1 kHz. This implied that the probe signal altered the polar pattern systemically when the probe was presented from the rear azimuth of HA3. However, the negative values indicated that the sensitivity to the probe increased after the probe signal was presented. All of this might relate to the uncommon polar pattern shown in Fig. 2. Except for HA3, no other systemic positive error values were observed (including HA1 at high frequencies), indicating that the 75-dB jammer could effectively freeze the polar patterns.

The differences between the measured PI and corrected PI at 0.5, 1, 2, and 4 kHz of each DMHA are shown in the upper panel of Fig. 5. A positive PI difference means the measured PI is larger than the corrected PI (i.e., overestimation). The error bars represent the maximum and minimum of the five measures. Apparently, the PI difference is larger and more variable at 500 Hz. Possible explanations include an alteration of the polar pattern by the probe signal at low frequencies in all DMHAs, or that the error values are more easily contaminated by background noise (as previously described), or both. Except for 500 Hz, the PI differences were generally small, indicating that the SC method is accurate.

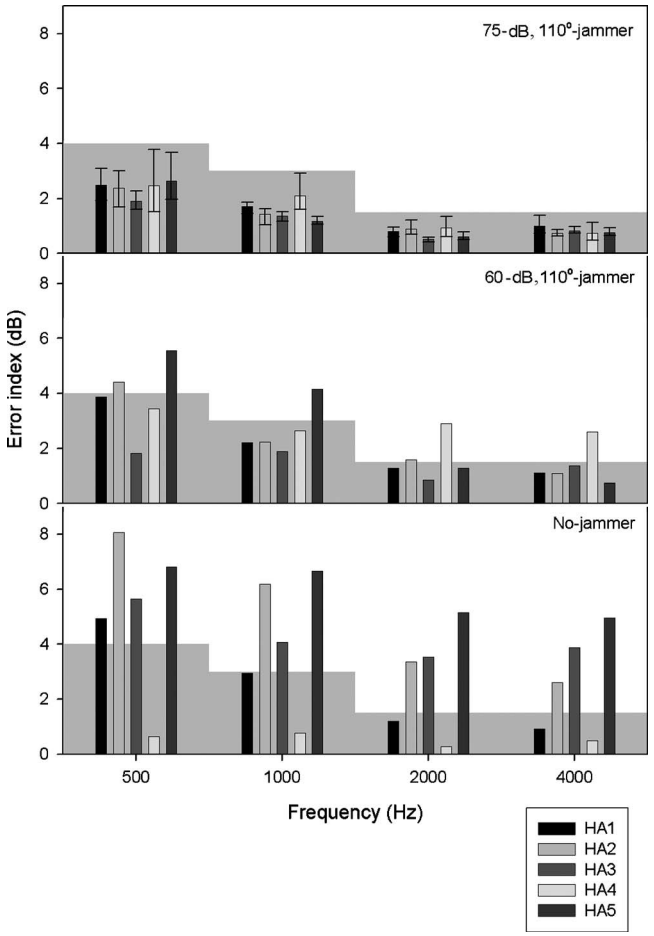


FIG. 6. The error index as a function of frequency in the 75-dB, 110°-jammer, 60-dB, 110°-jammer, and no-jammer conditions. The error bars in the upper panel represent the data range from five measures. The shaded areas represent the norm.

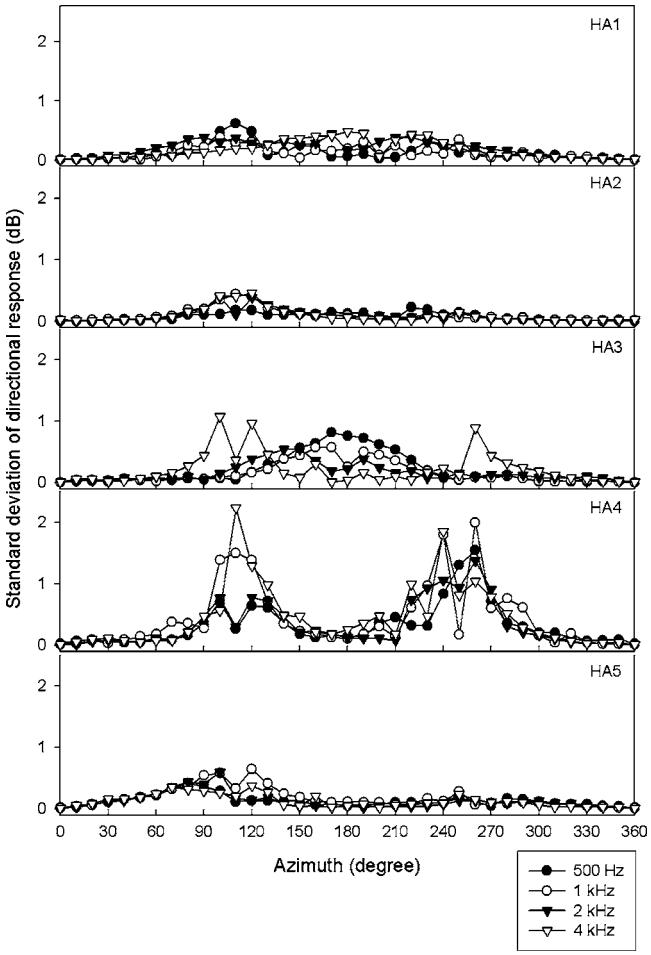


FIG. 7. The standard deviation of the directional responses as a function of azimuth angle.

Almost all the PI differences were within ±0.6 dB (0.5 kHz), ±0.4 dB (1 kHz), ±0.3 dB (2 kHz), and ±0.3 dB (4 kHz). We can use these ranges to establish the norm. If a PI difference from one measurement is within the range of the norm, we can expect the result to be accurate. More precisely, the result is as accurate as in the 75-dB, 110°-jammer condition we tested here. With this norm, we can examine the accuracy of the results obtained from other test conditions, such as conditions with different jammer intensity and location. The norm is shown in Fig. 5 by the shaded area.

The error index of each DMHA at 0.5, 1, 2, and 4 kHz is shown in the upper panel of Fig. 6. The error bars represent the range of the EI values of five consecutive measures. Recall that the EI is the rms of the error values across azimuth angles. It consists of any change of the polar pattern (both positive and negative), and the effect of background noise. Therefore, it is not surprising to note that the index tends to be larger in the low frequencies. All EI values were smaller than 4 dB (0.5 kHz), 3 dB (1 kHz), 1.5 dB (2 kHz), and 1.5 dB (4 kHz). Again, the results suggest that the SC method is more accurate in the mid- to high-frequency range. We can treat the range of the EI values obtained from the 75-dB, 110°-jammer condition as a norm if we assume that the results in this jammer condition are accurate. The norm can be used to examine the accuracy of the SC method in other jammer conditions. It is shown in Fig. 6 by the shaded area.

To evaluate the reliability of the SC method, the standard deviation of the PI and directional response at each azimuth angle were computed. In Fig. 3, the error bars represent 1 standard deviation of the measured PIs. One can see that the PI variation was very small for all DMHAs at all frequencies. Most PI standard deviations were smaller than 0.1 dB. The standard deviations of the directional response as a function of azimuth angle at 0.5, 1, 2, and 4 kHz are shown in Fig. 7. Consistent with the error values shown in Fig. 4, there were more variations of the directional response when the probe signal was presented around the azimuth of the jammer signal. Except for this fluctuation, the only device that had a large systemic variation was HA3 at around 180° azimuth, indicating that, with the jammer at 110°, the adaptive system was not stable when the probe signal was from the rear azimuth. The standard deviations were generally small in other hearing aids studied here, indicating the SC method is reliable for most DMHAs.

B. 60-dB, 110°-jammer and no-jammer conditions

The polar plots and polargrams from the SC method with and without the 60-dB-SPL jammer are shown in Figs.

8 and 9, respectively. The PI values are shown in Fig. 10. With the less intense 60-dB-SPL jammer signal, some DMHAs (HA1, HA2, HA5) changed the polar patterns at low frequencies. In the no-jammer condition, most polar patterns could not be classified under any known theoretical pattern. Many PI values shown in Fig. 10 are close to or larger than 6 dB, indicating that most polar patterns were changed by the polar signals in these two conditions. From the middle and lower panels of Fig. 5 and 6, one can see that most PI differences and EI values are larger than the norm, especially for the no-jammer condition. All these parameters suggest that the SC method works only with a high-intensity jammer. The only exception is HA4. In the no-jammer condition, the polar pattern of HA4 was an omnidirectional pattern (Fig. 9). The PI differences and EI values were very close to zero, indicating that the measured pattern was the real polar pattern. This is because the HA4 has an automatic adaptive directional system. It can switch not only between different polar patterns, but also between the directional and omnidirectional mode. Because the probe intensity was lower than the threshold for switching from the omnidirectional mode to the directional mode, the DMHA maintained its omnidirectional polar pattern.

C. 75-dB, 180°-jammer condition

The polar plots and polargrams for the 75-dB, 180°-jammer condition are shown in Fig. 11. Except for the high frequencies of HA1, all the DMHAs steered the polar patterns into a cardioid pattern appropriately to respond to the jammer from 180°. Surprisingly, HA3 also steered the null to the same azimuth. From Fig. 2, it seems that HA3 is not adaptive below 3 kHz because the null does not point to the jammer at 110° for frequencies below 3 kHz.

D. Summary

For HA2, HA4, and HA5, the results in experiment 1 reveal that the SC method meets our criteria. With the 75-dB-SPL jammer, the SC method is accurate because (1) the polar pattern shape is consistent with the theoretical pattern; (2) the PI values are smaller than 6 dB; (3) the error values are small except for at the azimuth angles around the jammer; (4) the differences between measured PI and corrected PI are small; and (5) the EIs are generally small. The methodology has greater accuracy for mid to high frequencies than for low frequencies. The SC method is reliable because (1) the variations of the PI values are small in five consecutive measures; and (2) the variations of the directional responses are small except for at the azimuth angles around the jammer. Compared with the 60-dB-jammer and no-jammer conditions, the SC works best with the high-intensity jammer. Before we generalize the conclusion to HA1 and HA3, we need to explain the observed uncommon polar patterns of these two hearing aids.

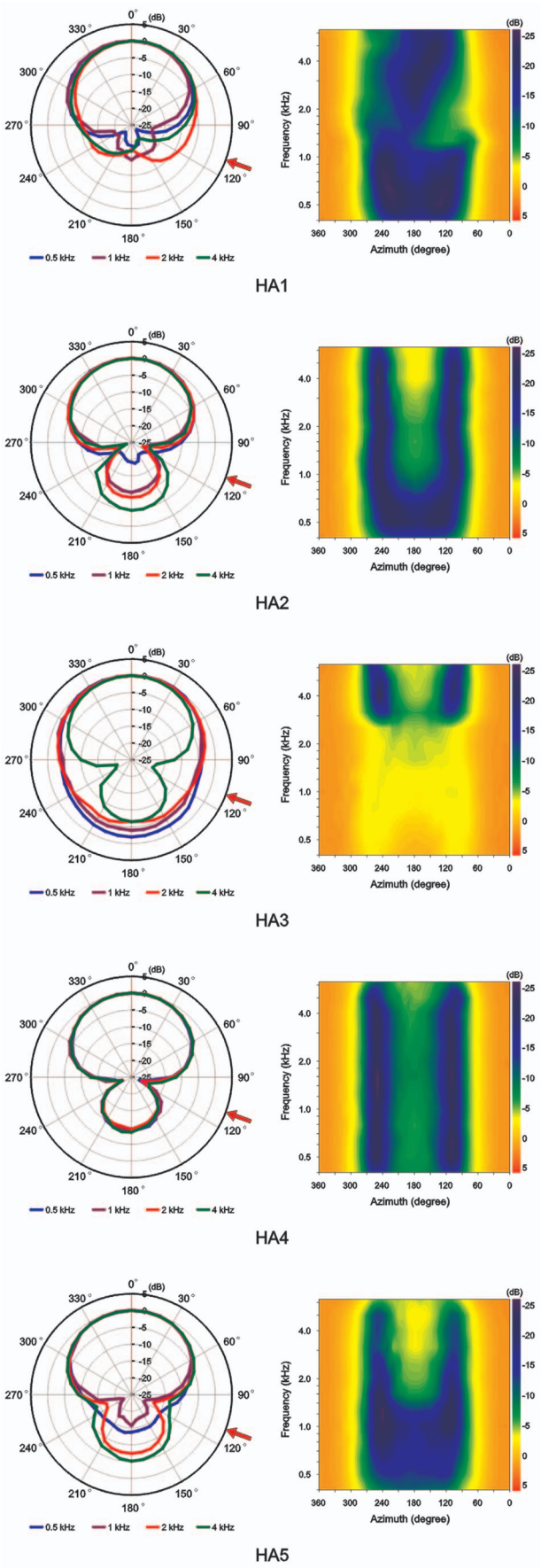


FIG. 8. The directivity of five adaptive DMHAs in the 60-dB, 110°-jammer condition. The polar plots are shown on the left side and the polargrams are shown on the right side of each panel. The arrow indicates the azimuth of the jammer signal.

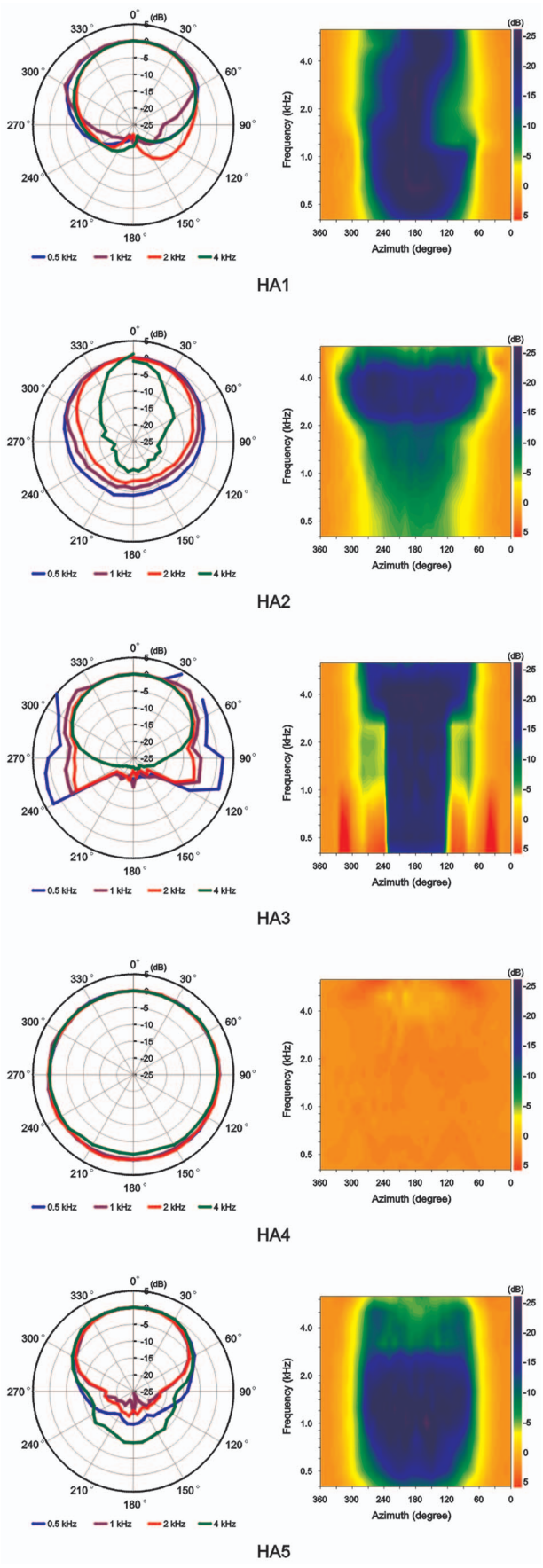


FIG. 9. The directivity of five adaptive DMHAs in the no-jammer condition. The polar plots are shown on the left side and the polargrams are shown on the right side of each panel.

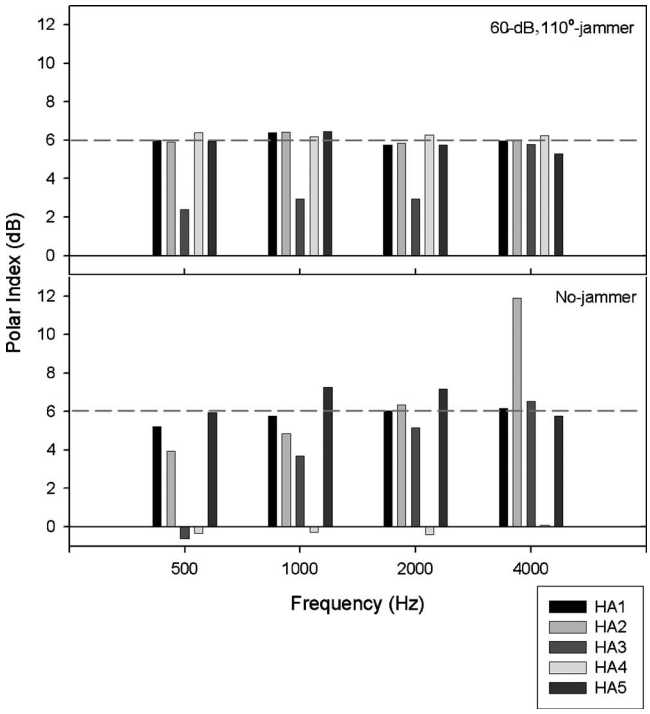


FIG. 10. The polar index as a function of frequency in the 60-dB, 110°-jammer and no-jammer conditions. The dashed line represents the theoretical upper limit of the free-field polar index of the first-order directional microphone.

IV. METHOD FOR EXPERIMENT 2

A. Rationale

Although the PI differences were small, the polar patterns of HA1 (at high frequencies) and HA3 (at mid to low frequencies) were uncommon. It was necessary to investigate the directivity of these two DMHAs to answer the question: Are these unusual patterns due to an error of the SC method, or to the intrinsic nature of these DMHAs?

HA1 has a first-order directional microphone system at frequencies below 1 kHz. Above 1 kHz, it is a second-order system (i.e., three omnidirectional microphones used). In the 75-dB, 110°-jammer and 75-dB, 180°-jammer conditions (Figs. 2 and 11), the system steered the nulls appropriately at the low frequencies. In contrast, at frequencies above 2 kHz, the polar pattern was nearly identical in all test conditions in experiment 1 (Figs. 2, 8, 9, and 11). Between 1 and 2 kHz, the polar pattern looked like a transition between low- and high-frequency patterns. The measured PIs at high frequencies were very similar in different conditions. For example, the PIs at 4 kHz in the 75-dB, 110°-jammer, 60-dB, 110°-jammer, and no-jammer conditions were 6.02, 5.98, and 6.17 dB, respectively. The EI values at 4 kHz for these three conditions were also close and small: 1.00, 1.10, and 0.91 dB. This suggested that for high frequencies, the polar pattern did not change. In other words, although HA1 was programed into the adaptive mode, the microphone at high frequencies is still in a fixed directional mode. However, before we make this conclusion, we have to consider that a single jammer may not be able to freeze the polar pattern of this second-order directional system, because, theoretically, an adaptive array comprising multiple microphones (M) can

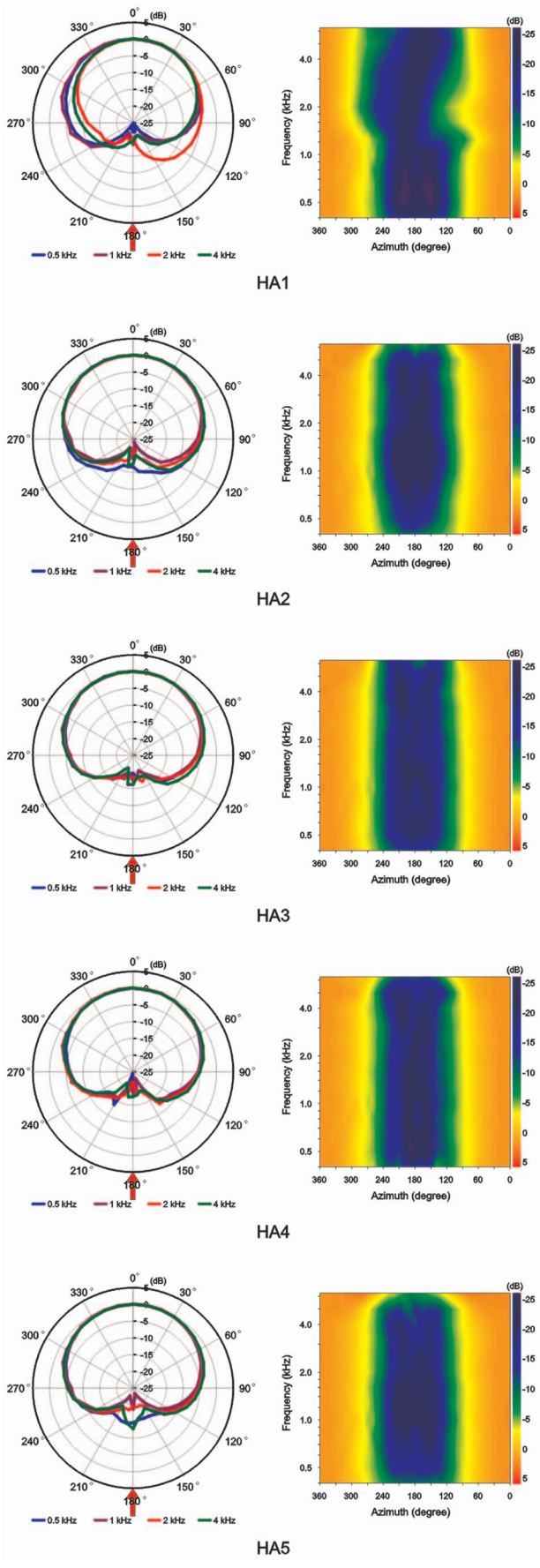


FIG. 11. The directivity of five adaptive DMHAs in the 75-dB, 180°-jammer condition. The polar plots are on the left side and the polargrams are on the right side of each panel. The arrow indicates the azimuth of the jammer signal.

steer M-1 nulls. It is possible that the unusual polar pattern was not the real pattern because HA1 placed one null to the jammer signal at 110°, and steered the other null to the moving probe signal. Thus, two jammer loudspeakers are needed to freeze the polar pattern of the three-microphone array in HA1. To test this hypothesis, HA1 was measured by the SC method with a second 75-dB-SPL jammer signal located at 180° (75-dB, 110°–180°-jammer condition). With two jammer signals, the polar pattern of the second-order system should be frozen. In addition, HA1 was reprogramed into the fixed directional mode and tested in the 75-dB, 110°-jammer condition. If the polar patterns at high frequencies are similar to those in experiment 1, we can conclude that the directional system at those frequencies is in the fixed mode, even though the DMHA is programed into the adaptive mode.

The behavior of HA3 below 3 kHz is of interest. The system could steer the nulls adaptively to 180° if the jammer loudspeaker was placed at that azimuth (Fig. 11). But, there was no apparent null when the 75-dB and 60-dB-SPL jammers were presented from 110° (Figs. 2 and 8). The negative and highly variable error values (Figs. 4 and 7) between 120° and 240° suggested that, when the probe moved to the rear azimuth of HA3 in the 110°-jammer conditions, the system did not behave appropriately and was not stable. After carefully reviewing all the data, we found that the polar plot in the no-jammer condition provided the clue. From the middle panel of Fig. 9, one can see the directional response dropped abruptly between 130° and 250° at the frequencies below 3 kHz. This suggests that the system only steers nulls appropriately within this range. To test this hypothesis, HA3 was tested with a 75-dB-SPL jammer presented from 60°, 90°, 120°, 130°, and 150°. We hypothesized that no null would be seen in the 120°-jammer condition, and the null would point to the jammer at the 130°- and 150°-jammer conditions.

B. Equipment setup

The equipment setup, signal presentation, and analysis were the same as in experiment 1. The spectra and intensities of the probe signals were determined by the same principle. To test HA1, a second jammer loudspeaker was installed on the turntable. The second jammer loudspeaker was placed 7° below the horizontal plane, and the distance between HA1 and the jammer loudspeaker was 0.9 m. From the output of the ART 355 equalizer, the jammer signal was split and routed to different channels of the ADCOM GFA5002 amplifier, and then to the two jammer loudspeakers, respectively.

V. RESULTS OF EXPERIMENT 2

The result of HA1 tested with two jammer signals is shown in Fig. 12(a). At low frequencies, the polar pattern was close to a supercardioid pattern. That is consistent with two jammer signals from 110° and 180°. At high frequencies, especially 4 kHz, the polar patterns were very similar to the pattern in experiment 1. The measured PI and the EI at 4 kHz were 5.94 dB and 1.00 dB, respectively, which were very close to the values mentioned previously. The polar pattern of the fixed directional mode is shown in Fig. 12(b).

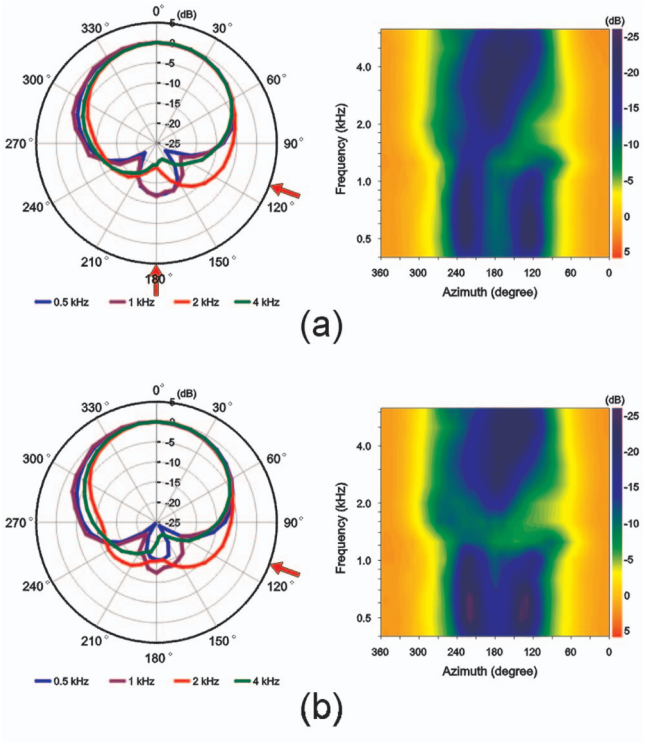


FIG. 12. The directivity of HA1 in the (a) 75-dB, 110°–180°-jammer condition (programed to the adaptive mode) and (b) 75-dB, 110°-jammer condition (programed to the fixed mode). The polar plots are shown on the left side and the polargrams are shown on the right side of each panel. The arrow indicates the azimuth of the jammer signal.

Again, the polar patterns at high frequencies are almost identical to the patterns we have shown before. These results confirmed that the directional system of HA1 at high frequencies is not adaptive, regardless of which mode the DMHA is programed to. With these results, we can conclude that the SC method can assess HA1 accurately.

The results of HA3 tested with different jammer azimuth angles are shown in Fig. 13. The PI differences and EI values are shown in Fig. 14. From these two figures, three different directivity types are observed. First, in the 60°-jammer and 90°-jammer conditions, the polar patterns were the dipole pattern (Fig. 13). Because the PI differences and EI values at 0.5 and 1 kHz were larger than the norm established in experiment 1 (i.e., the bars are higher than the shaded area in Fig. 14), the dipole patterns were altered by the probe signal. Second, similar to the 110°-jammer condition in experiment 1, when the jammer was presented from 120°, no null could be found below 3 kHz. This unusual pattern might not deviate from the real pattern too much because the PI differences and EI values were smaller than the norm. Third, when the jammer was located at 130° and 150°, HA3 steered the nulls adaptively at all frequencies. Because almost all PI differences and EI values were smaller than the norm, the polar patterns in these two azimuth angles were accurate. Given the results from experiments 1 and 2, it seems that the "back" of HA3 is defined as the azimuth angles between 130° and 250°, and the "front" is defined as within ±90°. If a loud sound is presented from the back of HA3, the adaptive system steers the null appropriately. If the sound is coming

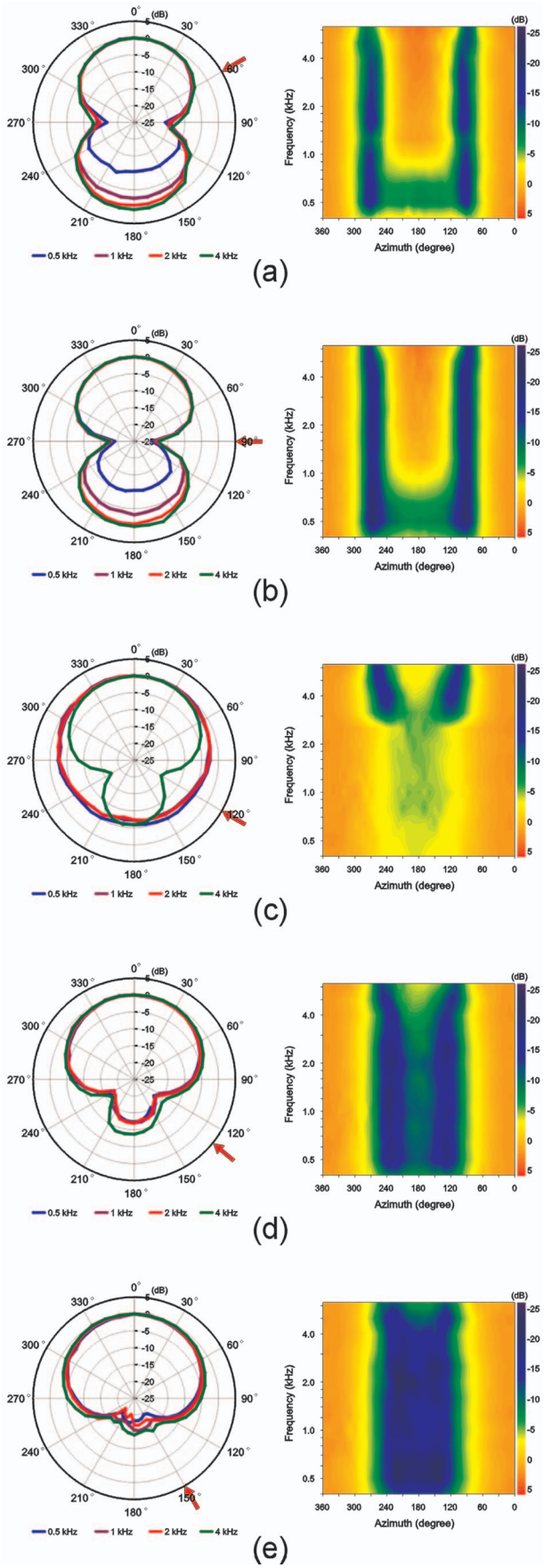


FIG. 13. The directivity of HA3 with a 75-dB-SPL jammer from (a) 60°; (b) 90°; (c) 120°; (d) 130°; and (e) 150°. The polar plots are shown on the left side and the polargrams are shown on the right side of each panel. The arrow indicates the azimuth of the jammer signal.

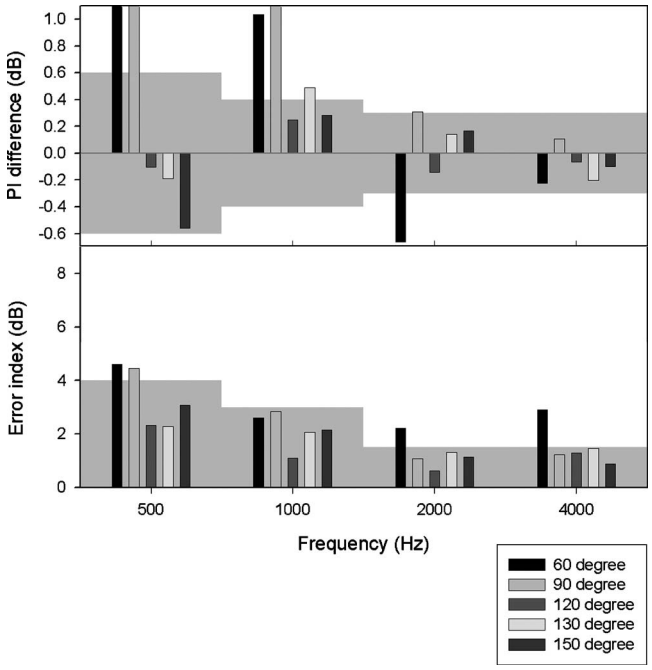


FIG. 14. The PI difference and error index of HA3 with a 75-dB jammer from different azimuth angles as a function of frequency. The shaded areas represent the norm. The data beyond the scale of *y* axis are truncated.

from HA3's front, the pattern is set to a dipole pattern, which is the polar pattern that can reduce the sensitivity most to sound sources from the frontal hemisphere. In this configuration, a probe signal from the rear azimuth of the DMHA can alter the polar pattern easily because the null can be steered to the probe signal and attenuate the microphone output more. This explains the inaccurate polar patterns at low frequency in the 60°- and 90°-jammer conditions. It seems that the azimuth angles from 100° to 120° (and 240° to 260°) form the transition zone, or the boundary of the DMHA's front and back. If a sound is presented from this transition zone, the algorithm becomes unstable and cannot behave appropriately, resulting in the uncommon polar pattern.

From experiment 2, we believe that the unusual polar patterns at high frequencies for HA1 are because the directional system is not adaptive in those frequencies. The unusual polar patterns at mid to low frequencies for HA3 are because the jammer signal is located in the transition zone. The SC method worked well with these two DMHAs and gave the correct results. Therefore, we can conclude that, with a high-intensity jammer, the SC method is accurate, reliable, and can be used to assess the directivity of most DMHAs.

VI. DISCUSSION

A. General discussion

The SC method is designed to assess the directivity of the adaptive DMHA in the presence of a high-intensity jammer. The accuracy, reliability, and generalizability were examined on five DMHAs by the accuracy metrics (the shape of the polar pattern, the measured PI, the difference between measured PI and corrected PI, and the EI values), variations of PI values, and variations of directional responses.

In the SC method, we used one or more high-intensity jammers to drive and freeze the polar pattern, and then used a low-intensity probe signal to assess the polar pattern. The key point of the SC method is the signal cancellation technique. If the jammer signal cannot be canceled out completely, the probe response cannot be extracted correctly, and a probe with higher intensity has to be used to obtain the 20- to 25-dB probe-to-noise floor ratio. Hagerman (2004) has shown that this technique can evaluate noise reduction algorithms in hearing aids. In the current study, we used this technique to evaluate the adaptive directional system with the noise reduction algorithm disabled. In the 75-dB-jammer condition, the probe intensities were around 60 dB SPL for all DMHAs (Table I); this level could achieve the 20- to 25-dB probe-to-noise floor ratio. Therefore, the signal cancellation technique we used here can cancel the jammer response by 35–40 dB, which is compatible with Hagerman's result. Also from Table I, one can see the probe intensities in the 75-dB-jammer condition are higher than the probe intensities in the no-jammer condition by only several decibels. These two facts indicate that, with the algorithm and equipment setup in this study, most jammer response can be removed from the output of the DMHAs.

In the accuracy metrics, the rationale to use the polar pattern shape and PI value is that, if the probe signal changes the polar pattern, the polar pattern will not be a hypercardioid pattern (in the 110°-jammer condition), and the PI may be larger than 6 dB. It seems that these two are good parameters because most polar patterns were the hypercardioid pattern in the 75-dB, 110°-jammer condition, and many polar patterns deviated from the hypercardioid pattern in 60-dB, 110°-jammer and no-jammer conditions. The same trend can be found in the PI values. However, there are some drawbacks to using these two parameters (polar shape and PI value). First, it is not always easy to determine if a polar pattern is different from the theoretical pattern. For example, the polar pattern shapes of HA4 were similar in the 75-dB- and 60-dB-jammer conditions (Figs. 2 and 8), but the PI values in the 60-dB-jammer condition were larger than 6 dB at all frequencies (Fig. 10), indicating that the polar patterns were inaccurate. It is more difficult to compare the polar pattern shapes of higher-order directional systems. Second, the absolute PI value is not sensitive enough to determine directional accuracy. For example, in the no-jammer condition, except for the polar patterns of HA1 at frequencies above 2 kHz (fixed second-order system) and HA4 at all frequencies (omnidirectional mode), the remaining 15 polar patterns are unusual (Fig. 9), indicating that these curves are not the real patterns. However, only five of the 15 PIs are larger than 6 dB (the lower panel of Fig. 10). Third, if the DMHA is tested on a KEMAR, it is impossible to compare the shape with any theoretical pattern, or use the 6 dB as a standard, because the theoretical pattern has been changed by acoustic reflections from KEMAR's pinna and head.

In contrast, the PI difference and EI values are better parameters to evaluate accuracy. Both are based on the error value: the difference between the beginning and the end of the probe response amplitude. The norms of the PI difference and EI values are established based on the assumption that

the results in the 75-dB, 110°-jammer condition are accurate. This assumption is supported by other parameters, including small error values across azimuth angles, theory-compatible polar patterns and PI values, and results from experiment 2. The PI difference and EI value are very sensitive parameters. As mentioned previously, in the no-jammer condition, there are 15 unusual polar patterns, and only five PI values are larger than 6 dB (Fig. 10). On the other hand, in the 15 inaccurate results, 13 PI differences and 13 EI values were beyond the range of the norms (Figs. 5 and 6). The PI difference and EI value are also very consistent with each other. For example, for the 13 PI differences beyond the norm in the no-jammer condition, 12 of them have EI values larger than the norm. The PI difference and EI can be very useful parameters not only to evaluate the results with different jammer intensities and azimuth angles, but also to examine the accuracy when the DMHA is measured on a KEMAR. If both the PI difference and EI values are within the norm, we can conclude that the accuracy of the measurement is as good as the results in this study.

B. Limitations of the SC method

Although the SC method is accurate and reliable, it has some limitations. As mentioned, the probe intensity should be high enough to achieve the 20- to 25-dB probe-to-noise floor ratio. Recall that the noise floor of this ratio is the summation of the residual jammer signal after accounting for cancellation, the background noise, and the internal noise of the DMHA. After signal processing, most of the jammer response can be canceled out, but the ambient noise and the internal noise still remain due to their random phases. Because the energy of these two noises is concentrated in the low-frequency range, the intensities of the probe at low frequencies usually have to be increased to obtain an enough probe-to-noise floor ratio. As a result, the accuracy of the SC method at low frequencies is reduced because the higher intensity probe may change the low-frequency polar pattern. It is one of the reasons that the range of the norm is larger in the PI difference and EI values at low frequencies.

The second limitation also concerns the accuracy of the SC method. It is not always possible for all adaptive DMHAs on the market (now and in the future) to find a probe intensity that is low enough not to change the directional pattern under a given jammer, while high enough to achieve the 20- to 25-dB probe-to-noise floor ratio. For these directional systems, the SC method cannot provide an accurate measurement. Fortunately, by the accuracy metrics, we will know if the result is accurate or not.

The third limitation involves the jammer signal. In order to freeze the directional patterns, a high-intensity jammer is used. In other words, the measured data only reflect the directivity under a high-intensity jammer signal. If one is interested in the directivity of a DMHA in the presence of a low-intensity noise (for example, the directivity of a DMHA with a fan noise behind it), the SC method is less useful because, with the low-intensity jammer, the probe may change the polar pattern. Besides, the SC method is not appropriate to measure the directivity of a microphone array that has multiple microphones because multiple jammer signals are needed to freeze the directivity pattern. Although we used two jammer loudspeakers to freeze the second-order microphone of HA1, it is almost impossible to install more loudspeakers and counterbalanced weights on a turntable. Therefore, an "inaudible" probe that never changes the polar pattern and a system that can hold multiple loudspeakers and rotate with the DMHA are necessary. The first author is currently working on these issues.

C. Application

The SC method can be implemented in most hearing aid laboratories with anechoic chambers because it can be converted easily from the equipment setup required for ANSI S3.35-2004. With different jammer signals and azimuth angles, the SC method can be used to assess the directivity in many conditions. For example, two jammers can be placed in different azimuth angles to verify the adaptivity of a system. Speech can be used as the jammer signal to simulate a talker presented from 0°, and babble can be presented from a second jammer loudspeaker at different azimuth angles. With these jammer settings, the polar patterns can be measured by the probe and used to predict the performance of the adaptive DMHA. By knowing the characteristics of adaptive systems, the researchers can design the laboratory setup to maximize the directional benefit. For example, if research is designed to investigate the adaptive system of HA1 by moving noises (e.g., Bentler *et al.*, 2004b), and if the moving noise is high frequency in spectrum, one may not see any difference between the adaptive and fixed directional modes. If a researcher used HA3 in a study, and placed the noise source at 110° (or the corresponding azimuth angle if HA3 is mounted on a man's head), he/she might not see much directional benefit from this adaptive DMHA.

The other application is using the signal cancellation technique to derive the SNR improvement provided by the directional microphone. Specifically, speech and noises can be presented from different azimuth angles to simulate different listening environments. With the appropriate signal design, the speech and noises can be extracted separately from the output of the DMHA, and the SNR can be calculated. With the SNRs of adaptive, fixed, and omnidirectional modes, the directional benefit in a specific environment can be predicted. Because the speech and noises are presented simultaneously, the DMHA does not need to be in a linear gain mode. Other features, such as the noise reduction algorithm, can also be enabled. Therefore, the performance of the directional microphone can be examined together with other features. For example, the authors used this signal cancellation technique to investigate the interaction between directional technology and compression in open-canal hearing instruments (Bentler *et al.*, 2006). In this application, no turntable is required and various speech and noise levels can be utilized.

The SC method can be used to assess the DMHA that switches between omnidirectional and fixed directional mode automatically (automatic directional system). If the jammer intensity and signal presenting sequence can be set to meet

the switching rule of the DMHA (e.g., Olson *et al.*, 2004), the system will be driven to the directional mode. Because the polar pattern of the directional mode is fixed, it can be assessed by the probe signal accurately. The SC method might also be able to assess the omnidirectional mode of an automatic directional system if a low-intensity probe signal will not move the system into the directional mode (e.g., HA4 in Fig. 9).

D. Polargram

The conventional polar plot can provide detailed information about the spatial characteristics of a DMHA's directivity. Since each frequency is represented as a single plot, the available information is limited. In contrast, the polargram can display continuous directivity in terms of frequency. If the polar patterns are different across frequencies, it can be read easily from the polargram. Therefore, we propose that the best way to express the directivity of a DMHA is the combination of the polar plot and the polargram; the directivity of different azimuth angles is read from the polar plot, and the directivity at different frequencies is read from the polargram.

VII. CONCLUSIONS

The SC method is an accurate and reliable method to evaluate the directivity of the adaptive DMHA in the presence of high-intensity jammer signals. It can be used to evaluate most adaptive DMHAs currently marketed. With the accuracy metrics, such as the PI difference and EI values, the accuracy of the SC method can be quantified. For research purposes, the SC method provides a powerful way to assess and verify the adaptive directional microphone system.

ACKNOWLEDGMENTS

The authors would like to acknowledge Steve Armstrong and an anonymous reviewer for providing a variety of insightful comments and suggestions regarding this work.

[1]In the pilot study, time for changing the polar pattern was measured in two ways. First, two loudspeakers were placed at 90° and 180° relative to the DMHA. A 75-dB-SPL pure tone at 1 kHz was presented from 90° for 1 min, and then was shifted to the loudspeaker at 180° instantaneously. The amplitude of the DMHA output was large at the moment of the pure-tone "jumping," which indicated that the null was not pointed to 180°, and then decreased gradually, which indicated that the system steered the null from 90° to 180°. The "engage time" of an adaptive DMHA is defined as the time taken for the output level to decrease to within 3 dB of its final value. The second way was, after 1-min silence, presenting a 75-dB-SPL pure tone at 1 kHz from 90°. The engage time (with the same definition) represents the time for an adaptive DMHA to switch the default pattern in quiet, which can be a hypercardioid, cardioid, or omnidirectional pattern, to a dipole pattern. The same DMHAs used in this study were tested. The DMHAs were programed into the linear gain mode, and the features such as the noise reduction algorithm were disengaged. The longest engage time of all results determined the 20-s interval of the jammer signal.

ANSI (**1997**). ANSI S3.5-1997, "American National Standard: Methods for the Calculation of the Speech Intelligibility Index" (American National Standards Institute, New York).

ANSI (**2004**). ANSI S3.35-2004, "American National Standard: Method of Measurement of Performance Characteristics of Hearing Aids Under Simulated Real-Ear Working Conditions" (American National Standards Institute, New York).

Bentler, R. A., Egge, J. L., Tubbs, J. L., Dittberner, A. B., and Flamme, G. A. (**2004a**). "Quantification of directional benefit across different polar response patterns," J. Am. Acad. Audiol **15**, 649–659.

Bentler, R. A., Palmer, C., and Dittberner, A. B. (**2004b**). "Hearing-in-Noise: comparison of listeners with normal and (aided) impaired hearing," J. Am. Acad. Audiol **15**, 216–225.

Bentler, R. A., Wu, Y., and Jeon, J. (**2006**). "Effectiveness of directional technology in open-fitting hearing instruments," Hear. J. **59**(11), 40–47.

Beranek, L. L. (**1949**). *Acoustic Measurements* (Wiley and Sons, London).

Chung, K. (**2004**). "Challenges and recent developments in hearing aids. I. Speech understanding in noise, microphone technologies and noise reduction algorithms," Trends. Amplif. **8**, 83–124.

Dillon, H. (**2001**). *Hearing Aids* (Thieme, New York).

Dittberner, A. B., and Bentler, R. A. (**2007**). "Predictive measures of directional benefit. II. Verification of different approaches to estimating directional benefit," Ear Hear. **28**, 46–61.

Dittberner, A. B., and Bondy, J. (**2006**). "A method is proposed to evaluate directional-microphone strategies," Hear. J. **59**(7), 28–33.

Ghent, R. M. (**2005**). "A tutorial on complex sound fields for audiometric testing," J. Am. Acad. Audiol **16**, 18–26.

Hagerman, B. (**2004**). "A method to measure the effect of noise reduction algorithms using simultaneous speech and noise," Acta Acust. (Beijing) **90**, 356–361.

Killion, M. C., Schulien, R., Christensen, L., Fabry, D., Revit, L., and Niquette, P. (**1998**). "Real-world performance of an ITE directional microphone," Hear. J. **51**, 24–38.

Olson, L., Ioannou, M., and Trine, T. (**2004**). "Appraising an automatically switching directional system in the real world," Hear. J. **57**, 32–38.

Palmer, C., Bentler, R. A., and Meuller, H. G. (**2006**). "Evaluation of a second-order directional microphone hearing aid. II. Self-report outcomes," J. Am. Acad. Audiol **17**, 190–201.

Powers, T., and Hamacher, V. (**2004**). "Proving adaptive directional technology works: A review of studies," Hear. Rev. **46**, 48–49, 69.

Ricketts, T. (**2000**). "Directivity quantification in hearing aids: Fitting and measurement effects," Ear Hear. **21**, 45–58.

Ricketts, T. (**2006**). "Directional hearing aids: New designs and predictions applied to old concepts," Hear. J. **59**(6), 10–18.

Ricketts, T., and Dhar, S. (**1999**). "Comparison of performance across three directional hearing aids," J. Am. Acad. Audiol **10**, 180–189.

Ricketts, T., and Henry, P. (**2002**). "Evaluation of an adaptive, directional-microphone hearing aid," International J. Audiol. **41**, 100–112.

Ricketts, T., Henry, P. P., and Hornsby, B. W. (**2005**). "Application of frequency importance functions to directivity for prediction of benefit in uniform fields," Ear Hear. **26**, 473–486.

Ricketts, T., and Mueller, H. G. (**1999**). "Making sense of directional microphone hearing aids," Am. J. Audiol. **8**, 117–127.

Valente, M., Hosford-Dunn, H., and Roeser, R. J. (**2000**). *Audiology Treatment* (Thieme, New York).

Walden, B. E., Surr, R. K., Cord, M. T., Edwards, B., and Olson, L. (**2000**). "Comparison of benefits provided by different hearing aid technologies," J. Am. Acad. Audiol **11**, 540–560.

Tongue movement kinematics in long and short Japanese consonants

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

(Received 22 November 2006; revised 30 March 2007; accepted 30 March 2007)

This paper examines tongue movements in stop and fricative consonants where the duration of the oral closure/constriction for the consonant is varied for linguistic purposes. Native speakers of Japanese served as subjects. The linguistic material consisted of Japanese word pairs that only differed in the duration of the lingual consonant, which was either long or short. Recordings were made of tongue movements using a magnetometer system. Results show a robust difference in closure duration between the long and short consonants. Overall, the path of the tongue movement during the consonant was longer for the long than for the short consonant. All speakers decreased the speed of the tongue movement during the long consonant. These adjustments in tongue movements were most likely made to maintain the contact between the tongue and the palate for the closure and constriction. © *2007 Acoustical Society of America.* [DOI: 10.1121/1.2735102]

PACS number(s): 43.70.Aj, 43.70.Bk [BHS] Pages: 512–518

I. INTRODUCTION

This paper examines tongue movements in stop and fricative productions where the duration of the oral closure/constriction for the consonant is varied for linguistic purposes, using speakers of Japanese. In Japanese, the ratio of closure duration for long and short consonants is about 2:1 (Beckman, 1982; Han, 1994; Hirata and Whiton, 2005). There is an extensive body of acoustic studies of the Japanese sound system with particular emphasis on the role of the mora for speech timing. A mora is traditionally regarded as a unit of timing in Japanese, but its existence and acoustic manifestations are debated, and the reader is referred to Warner and Arai (2001) for a review. The long consonants in Japanese are sometimes referred to as "geminates," and such a consonant contributes one mora; it is also traditionally assumed that a mora boundary occurs in the long consonant (Vance, 1987). However, the primary focus of this paper is not on mora timing but on speech motor control, capitalizing on the length distinction in Japanese sound structure to study how tongue movements are controlled when the duration of a consonant is changed.

If the duration of the oral closure for the consonant is increased, a speaker is still constrained to maintain the contact between the tongue and the palate to make the closure or constriction for the consonant. To do this, the speaker can in principle use two strategies for controlling the tongue movement, both of which involve modulating the speed of the tongue movement. One strategy would be to momentarily stop the tongue from moving. Alternatively, the speaker could slow down the tongue movement for the long consonant. Of these two possibilities, the second one is the most likely, since a large body of research on tongue movements in speech suggests that the tongue hardly ever stops moving. For example, the material presented by Mooshammer *et al.* (1995) showed that the tongue body is moving during the closure for a velar stop (see also Houde, 1968; Perkell, 1969). The amount of tongue movement during the closure varied with vowel context. Thus, the tongue body movement during the closure was about 1 mm when the preceding vowel was /i/, and between 4 and 10 mm when the preceding vowel was /u/ or /a/.

Studies of tongue movements during velar stop consonant production have also shown that the tongue tends to move in curved paths (Houde, 1968; Perkell, 1969; Kent and Moll, 1972; Schönle, 1988; Munhall *et al.* 1991; Mooshammer *et al.*, 1995; Löfqvist and Gracco, 1994; 2002). In the production of velar stops, the tongue movement trajectory into the stop closure is usually going forward, and the location of the point of contact between the tongue body and the palate at stop closure is influenced by the phonetic context of the stop (cf., Dembowski *et al.* 1998). Less information on movement kinematics is available for stops produced with the tongue tip and tongue blade, /t, d/, although the data published by Kent and Moll (1972) show that the location of the point of contact between the tongue tip and the palate, or alveolar ridge, is influenced by phonetic context. In addition, the pattern of contact between the tongue and the alveolar ridge appears to vary with the shape of the palate (Hiki and Itoh, 1986). Löfqvist and Gracco (2002) recorded tongue movements for stop consonants produced with the tongue tip and the tongue body in different vowel contexts. Their results indicated that both the tongue body and the tongue tip moved during the closure for the consonant. The tongue tip movement was similar to that of the tongue body. In both cases, the movement was upward and forward at consonant onset, and downward and backward at the release. However, both the magnitude and direction of the movement were heavily influenced by the vowel context.

Electropalatographic studies of Italian stops have shown that the amount of tongue palate contact is larger for geminate than for single stops, and also that there is a general

[a)]Electronic-mail: lofquist@haskins.yale.edu

 0001-4966/2007/122(1)/512/7/$23.00

TABLE I. The linguistic material, text used for presentation, and glosses in English.

/hata/	kanji	"flag"
/hatta/	kanji&hiragana	"(someone) pasted (something)"
/muda/	kanji	"useless"
/budda/	katakana	"Buddha" katakana
/hosa/	kanji	"assistant"
/hossa/	kanji	"attack (medical condition)"
/haʃa/	kanji	"winner"
/haʃʃa/	kanji	"departure (of a train or a vehicle)"
/saka/	kanji	"slope"
/sakka/	kanji	"author, writer"
/tagu/	katakana	"price tag"
/tagau/	katakana	"tug"

increase in the extent of tongue-palate contact with increasing closure duration (Farnetani, 1990); similar results for American English have been presented by Byrd (1995). An x-ray study of French consonants by Vaxelaire (1995) suggested that the area of tongue palate contact was larger for the long (abutted) stops than for the short ones.

The purpose of the present study was to extend the findings of Löfqvist and Gracco (2002) to the control of tongue movements in consonants with different durations. As noted above, a speaker is constrained in producing lingual consonants to maintain the contact between the tongue and the palate. Thus, it was hypothesized that the speed of the tongue movement would be slower for a long than for a short consonant. A second hypothesis to be evaluated is that the magnitude of the tongue movement will be larger during a long than during a short consonant. To examine these hypotheses, tongue movements were recorded in native Japanese speakers.

II. METHOD

A. Subjects

Five native speakers of Japanese, two male and three 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 an alveolar or velar consonant. These words formed minimal pairs, where the only difference between the pairs was the duration of the consonant. The words are listed in Table I together with English glosses. The linguistic material was organized into randomized lists and presented to the subjects in Japanese writing, with the words occurring in the frame sentence "karewa_to itta" ("He said _"). Fifty repetitions of each word were recorded.

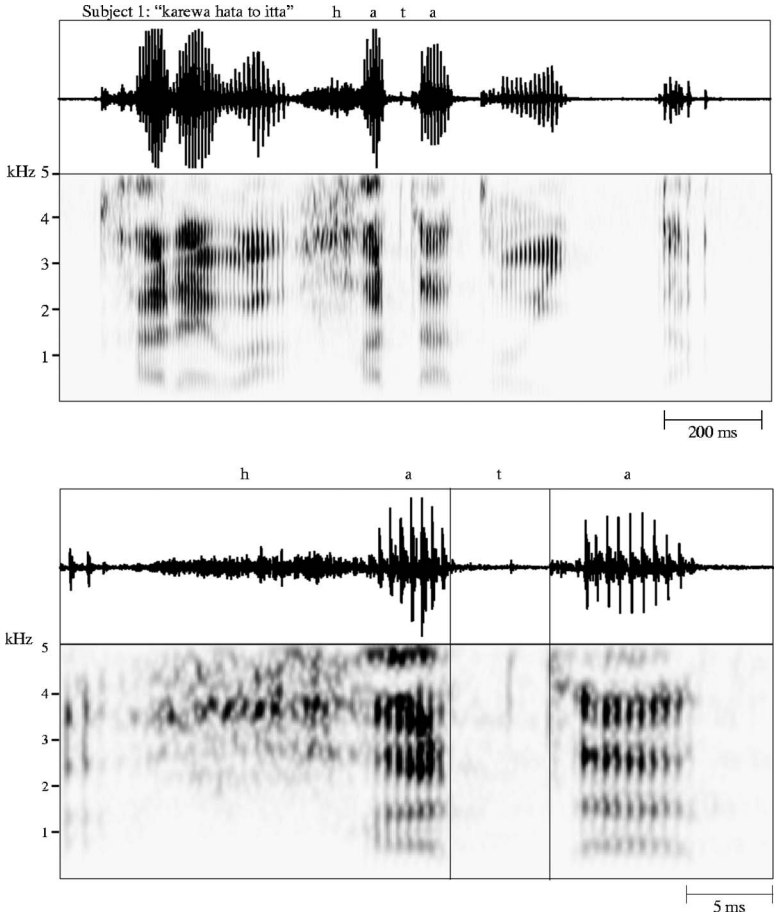


FIG. 1. Waveforms and spectrogram of the utterance "karewa hata to itta" produced by Speaker 1. The top panel shows the whole utterance, and the bottom panel only the word "hata" with the vertical lines showing the onset and release of the voiceless consonant /t/ in "hata."

C. Movement recording

The movements of the lips, the tongue, and jaw were recorded using a three-transmitter magnetometer system (Perkell *et al.*, 1992); when proper care is taken during the calibration, the spatial resolution of the system is on the order of 0.5 mm. Receivers were placed on the vermilion border of the upper and lower lip, on three points 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. The lip and jaw receivers were attached using Isodent, a dental adhesive, while the tongue receivers were attached using Ketac-Bond, another dental adhesive. Care was taken during each receiver placement to ensure that it was positioned at the midline with its long axis perpendicular to the sagittal plane. 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). For one speaker, S4, the signal from the tongue body receiver was degraded during the recording session, so no tongue body data are reported for this subject; the same thing happened for subject S3, so only 30 tokens could be used for the words with a velar consonant.

The articulatory movement signals (induced voltages from the receiver coils) 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 instantaneous velocity of the tongue receivers, the first derivative of the position signals was calculated using a 3-point central difference algorithm. For each tongue receiver, its speed $[v = \sqrt{(\dot{x}^2 + \dot{y}2)}]$ was also calculated. The velocity signals were smoothed using the same triangular window. All the signal processing was made using the Haskins Analysis Display and Experiment System (Rubin and Löfqvist, 1996). The acoustic signal was pre-emphasized, low-pass filtered at 4.5 kHz, and sampled at 10 kHz.

The onset and release of the oral closure/constriction for the consonant were identified in waveform and spectrogram displays of the acoustic signal. They were both identified by a change in the amplitude and the spectral properties, cf. Fig. 1.

The magnitude of the tongue movement trajectory during the consonant was obtained by summing the Euclidean distances between successive samples of the tongue tip and tongue body vertical and horizontal receiver positions from the acoustically defined consonant onset to offset. The average speed of the tongue tip and tongue body during the consonant was obtained by adding the speed of all the individual samples between consonant onset and offset and then dividing by the number of samples in the interval. The kinematic signals are expressed in a maxilla-based coordinate system.

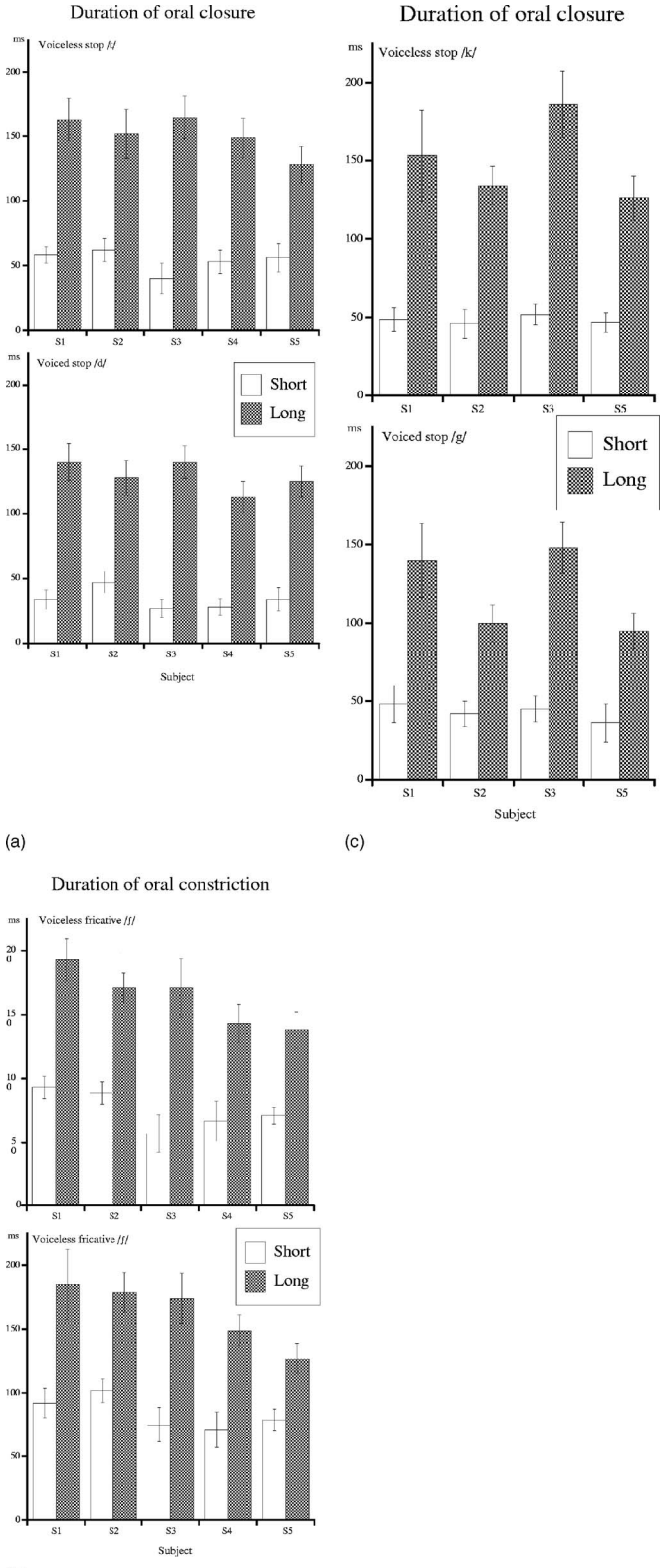


FIG. 2. Closure duration of the consonants (mean and standard deviation). (a) alveolar stops; (b) alveolar fricatives; and (c) velar stops.

Thus, the tongue movement includes the contribution of the jaw, which 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 num-

TABLE II. Results of the *t*-tests for the duration of the oral closure/constriction.

	S1	S2	S3	S4	S5
/t, tt/	41.30	29.8	42.92	37.54	28.92
/d, dd/	40.56	36.20	55.93	43.65	42.69
/s, ss/	37.55	39.80	29.51	24.20	30.39
/ʃ, ʃʃ/	22.11	29.85	29.18	29.28	23.94
/k, kk/	35.65	40.87	43.06		36.21
/g, gg/	38.91	28.67	38.90		25.48

ber of comparisons, a conservative α-level of 0.001 was adopted. Since the variances usually differed between the long and short consonants, as shown by Levene's test, the statistical tests assumed unequal variances, so the degrees of freedom were adjusted (e.g., Winer *et al.*, 1991, p. 67).

III. RESULTS

A. Consonant duration

The duration of the oral closure/constriction for the long and short consonants is summarized in Fig. 2. For all speakers, there is a clear and robust difference between the long and short consonants, with the long ones having twice, or more, the duration of the short ones. The results of the *t*-tests are summarized in Table II. Comparing the durations of the different consonant categories, the voiceless alveolar stop /t/ was always longer than its voiced cognate /d/. Also, the stops were usually shorter than the fricatives.

B. Tongue movements

Overall, the tongue movement patterns were similar to those reported previously (e.g., Houde, 1968; Mooshammer *et al.*, 1995; Löfqvist and Gracco, 2002). That is, both the tongue tip and the tongue body were moving up and forward towards the consonant constriction and down and backward after the consonant. There was some variability between and within speakers, but it is beyond the scope of the present paper to present a detailed analysis of these movement patterns.

The first analysis focused on the magnitude of the tongue movement during the consonant, while the second one examined the average speed of the tongue movement during the consonant.

1. Magnitude of tongue movement path

Figure 3 shows the magnitude of the tongue movement during the consonant. The path was generally longer for the long than for the short consonants, irrespective of place and

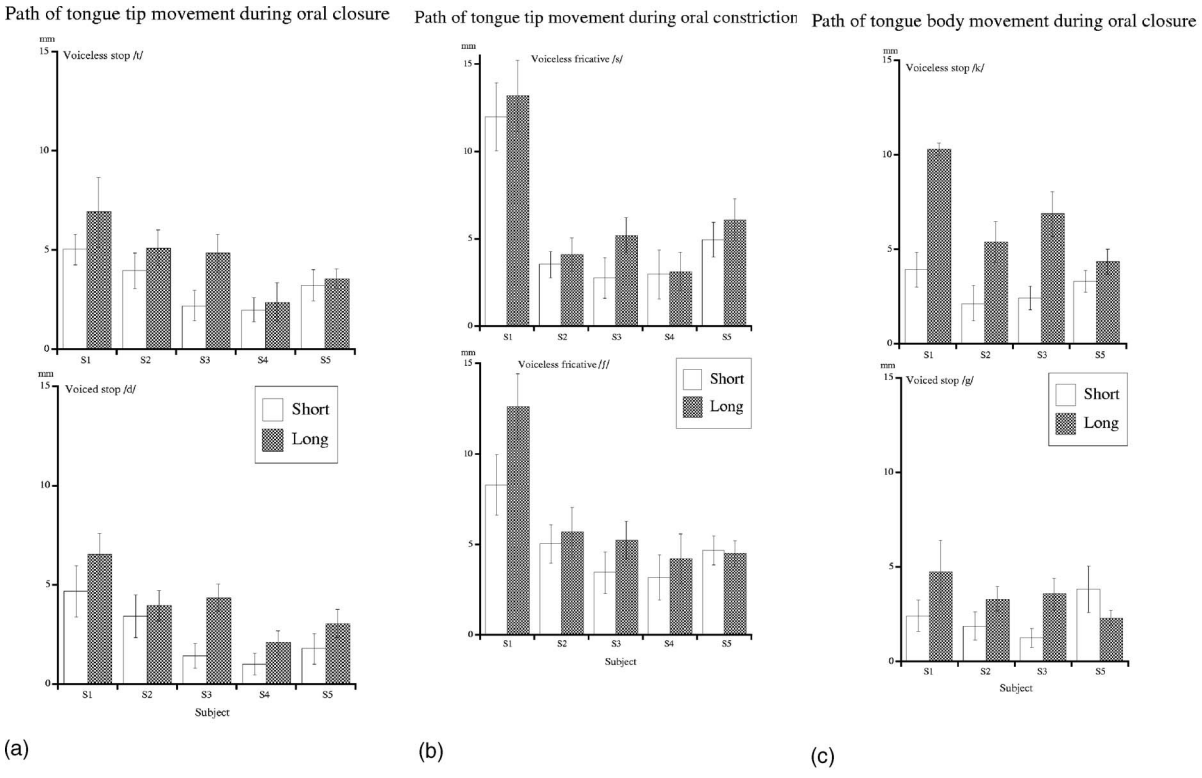
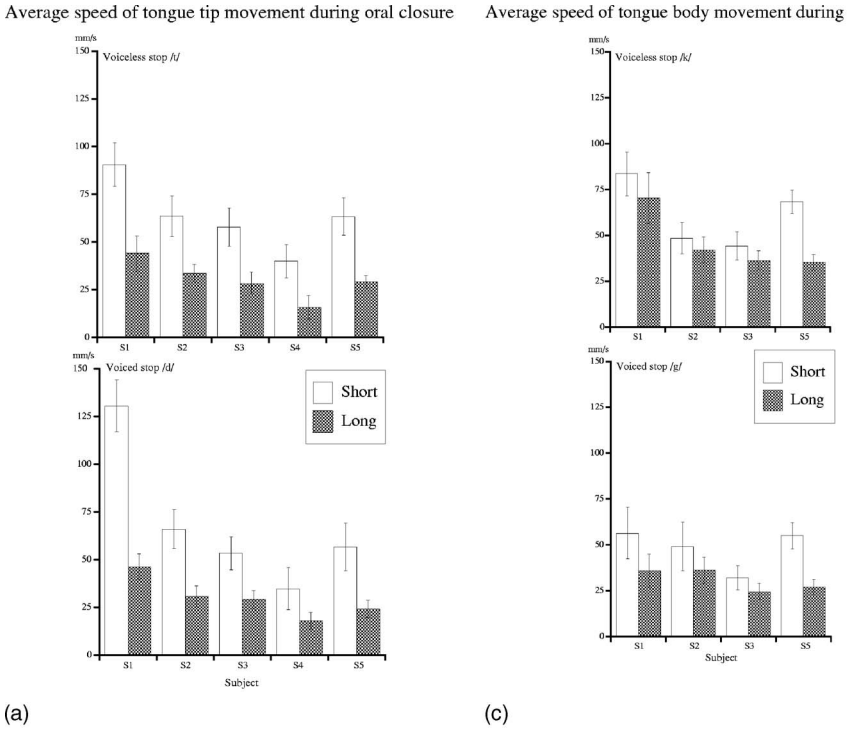


FIG. 3. Path of the tongue movement during the consonant (mean and standard deviation). (a) alveolar stops; (b) alveolar fricatives; and (c) velar stops. For the velar stops, the tongue body movement was measured, while for the other consonants, the tongue tip was used.

TABLE III. Results of the *t*-tests fo the path of the tongue during the oral closurre/constriction.

	S1	S2	S3	S4	S5
/t, tt/	6.85	6.35	15.16	2.19	2.42
/d, dd/	6.75	2.54	22.07	9.00	8.46
/s, ss/	3.03	3.33	10.90	0.57	5.15
/ʃ, ʃʃ/	12.44	2.72	8.14	3.95	1.09
/k, kk/	15.0	15.57	17.9		8.25
/g, gg/	15.71	9.92	12.64		3.53



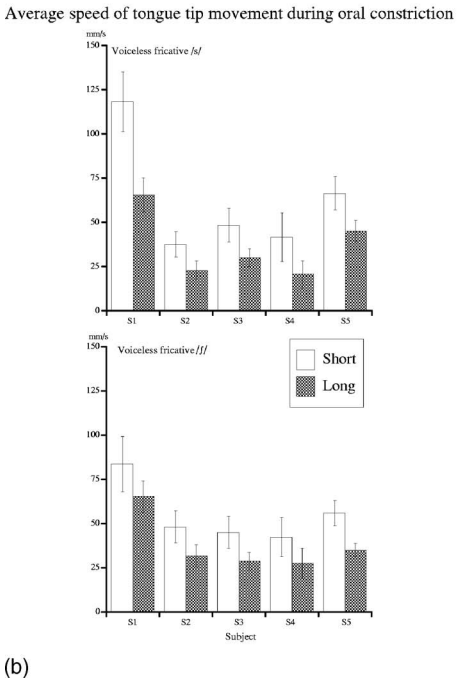


FIG. 4. Average speed of the tongue movement during the consonant (mean and standard deviation). (a) alveolar stops; (b) alveolar fricatives; and (c) velar stops. For the velar stops, the tongue body movement was measured, while for the other consonants, the tongue tip was used.

TABLE IV. Results of the *t*-tests for the average speed of the tongue during the oral closure/constriction.

	S1	S2	S3	S4	S5
/t, tt/	22.24	18.00	17.89	15.42	23.24
/d, dd/	26.68	23.81	17.45	9.53	17.09
/s, ss/	19.19	11.69	12.28	9.04	13.32
/ʃ, ʃʃ/	7.19	10.57	11.02	7.51	18.04
/k, kk/	10.31	4.66	4.72		31.7
/g, gg/	9.14	5.99	5.02		24.31

manner of articulation. The only exceptions were the voiceless fricative /ʃ/ and the voiced velar stop /g/ for subject S5, where the path was longer for the short consonant. The *t*-tests, summarized in Table III, showed that there was no significant difference between the length of the paths for the long and short consonants in the following cases: /t, tt/ for subjects S4 and S5; /s, ss/ for subjects S1, S2, and S4; /d, dd/ and /ʃ, ʃʃ/ for subjects S2 and S5. Given that the path was longer for the long than for the short consonants, correlations were made between the closure duration and the tongue movement path during the closure. All the correlations were positive, in particular for the short consonants. The patterns for the different consonant categories mirrored those for closure duration mentioned above. That is, the consonants with a longer duration also had a larger tongue movement.

2. Average speed of tongue movement

Figure 4 shows the average speed of the tongue body movement from the first to the second vowel. For all subjects and consonants, the average speed was slower for the long than for the short consonant. Table IV provides a summary of the *t*-tests, showing statistically significant differences for all subjects and consonants. There was no clear pattern for the different consonant caegories.

IV. DISCUSSION

This study examined tongue movement kinematics in Japanese long and short consonants. The hypothesis under investigation was based on the idea that speakers are constrained in the production of lingual consonants to maintain the contact between the tongue and the palate. Earlier work has shown that the tongue is moving during these consonants, although the influence of consonant duration on tongue movements has not been explored in any detail. Based on these earlier findings it was hypothesized that a speaker could either stop the tongue movement momentarily during a long consonant or slow down the movement speed. The results show that in all cases the speakers modulated the speed of the tongue, so that it was lower in the long than in the short consonants. There was no instance for any speaker where the tongue movement stopped completely.

If there is a difference in the average speed on the tongue movement between the long and short consonants, one might also predict that speakers modify the speed of the tongue movement within a category, long or short, as a function of consonant length. In particular, one would expect a negative relationship between consonant duration and average tongue movement speed. An examination of this relationship showed correlations to be mostly negative for the long consonants, but the correlations were not very strong. One reason that such a negative relationship was not found for the short consonants might simply be that their durations spanned a smaller range than those of the long consonants. Overall, the long consonants were produced with a longer tongue movement trajectory during the closure/constriction. Correlations were made between the closure duration and the tongue movement path during the closure to examine this relationship within the long and short consonants separately. All the correlations were positive, in particular for the short consonants.

Earlier work has shown a strong correlation between movement displacement and peak velocity in both speech and nonspeech movements (e.g., Cooke, 1980; Ostry *et al.*, 1983; Kelso *et al.*, 1985; Vatikiotis-Bateson and Kelso, 1993; Hertrich and Ackermann, 1997; Löfqvist and Gracco, 1997). The present study measured the average speed of the tongue movement during the acoustically defined closure/constriction for the consonant. There was, still, a very strong correlation between the magnitude of the tongue movement path and the average speed during the consonant, with almost all of the correlations higher than 0.7.

The present results show that speakers modify the speed of tongue movements in consonant production when the duration of the consonant is varied for linguistic purposes. They thus provide additional evidence for active, task-dependent control of tongue movement speed in speech, since the same was found for the movement from the first to the second vowel in a vowel-bilabial consonant-vowel sequence in Japanese, when the duration of the labial consonant was either long or short (Löfqvist, 2006). In this case, the speed of the tongue movement was lower during a long than during a short consonant. All speakers consistently showed this pattern, thus maintaining a similar, but not identical, pattern of coordination between the lip and tongue movements during long and short consonants.

These studies of the relationship between consonant duration and movement kinematics have used native adult speakers of Japanese. An interesting question is the emergence of the length contrast and its associated movement kinematics during speech development. Although there are methodological issues in the recording of tongue movements in children, lip and jaw movements can be readily analyzed (e.g., Green *et al.*, 2000; 2002). Another issue is how a learner of Japanese as a second language masters the length contrast and if the kinematic differences observed here in native speakers between long and short consonants are also

seen in second-language learners, in particular if there is a gradual change over time towards the kinematic patterns found here.

ACKNOWLEDGMENTS

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

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

Byrd, D. (**1995**). "Articulatory characteristics of single and blended lingual gestures," in *Proceedings of the XIIIth International Congress of Phonetic Sciences*, edited by K. Elenius and P. Branderud, Vol. **2**, pp. 438–441, Stockholm, Royal Institute of Technology and Stockholm University.

Cooke, J. (**1980**). "The organization of simple, skilled movements," in *Tutorials in Motor Behavior*, edited by G. Stelmach and J. Requin (North-Holland, Amsterdam), pp. 199–212.

Dembowski, J., Lindstrom, M. J., and Westbury, J. R. (**1998**). "Articulator point variability in the production of stop consonants," in *Neuromotor Speech Disorders: Nature, Assessment, and Management*, edited by M. P. Cannito, K. M. Yorkston, and D. R. Beukelman (Brookes, Baltimore), pp. 27–46.

Farnetani, E. (**1990**). "V-C-V lingual coarticulation and its spatiotemporal domain," in *Speech Production and Speech Modelling*, edited by W. Hardcastle and A. Marchal (Kluwer, Dordrecht), pp. 93–130.

Green, J., Moore, C., and Reilly, K. (**2002**). "The sequential development of jaw and lip control for speech," J. Speech Lang. Hear. Res. **45**, 66–79.

Green, J., Moore, C., Higashikawa, M., and Steeve, R. (**2000**). "The physiologic development of speech motor control: Lip and jaw coordination," J. Speech Lang. Hear. Res. **43**, 239–255.

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

Hertrich, I., and Ackermann, H. (**1997**). "Articulatory control of phonological vowel length contrasts: Kinematic analysis of labial gestures," J. Acoust. Soc. Am. **102**, 523–536.

Hiki, S., and Itoh, H. (**1986**). "Influence of palate shape on lingual articulation," Speech Commun. **5**, 141–158.

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.

Houde, R. (**1968**). *A Study of Tongue Body Motion during Selected Consonant Sounds* (Speech Communications Research Laboratory, Santa Barbara), SCRL Monograph 2.

Kelso, J. A. S., Vatikiotis-Bateson, E., Saltzman, E., & Kay, B. (**1985**). "A qualitative dynamic analysis of reiterant speech production: Phase portraits, kinematics, and dynamic modeling," J. Acoust. Soc. Am. **77**, 266–280.

Kent, R., and Moll, K. (**1972**). "Cinefluorographic analyses of selected lingual consonants," J. Speech Hear. Res. **15**, 453–473.

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

Löfqvist, A., and Gracco, V. L. (**1994**). "Tongue body kinematics in velar stop production: Influences of consonant voicing and vowel context," Phonetica **51**, 52–67.

Löfqvist, A., and Gracco, V. (**1997**). "Lip and jaw kinematics in bilabial stop consonant production," J. Speech Lang. Hear. Res. **40**, 877–893.

Löfqvist, A., and Gracco, V. (**2002**). "Control of oral closure in lingual stop consonant prodiction," J. Acoust. Soc. Am. **105**, 1864–1876.

Mooshammer, C., Hoole, P., and Kühnert, B. (**1995**). "On loops," J. Phonetics **23**, 3–21.

Munhall, K., Ostry, D., and Flanagan, R. (**1991**). "Coordinate spaces in speech planning," J. Phonetics **19**, 293–307.

Ostry, D., Keller, E., and Parush, A. (**1983**). "Similarities in the control of speech articulators and the limbs: Kinematics of tongue dorsum movements during speech," J. Exp. Psychol. Hum. Percept. Perform. **9**, 622–636.

Perkell, J. S. (**1969**). *Physiology of Speech Production: Results and Implications of a Quantitative Cineradiographic Study* (MIT Press, Cambridge, MA).

Perkell, J., M. 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.

Rubin, P. E. R., and Löfqvist, A. (**1996**). "HADES (Haskins Analysis Display and Experiment System)," Haskins Labs. Status Rep. Speech Res. (available at: http://www.haskins.yale.edu/MISC/DOCS/HADESRLn2.pdf). Viewed 5/11/07.

Schönle, P. (**1988**). *Elektromagnetische Artikulographie* (Electromagnet Articulography, Springer, Berlin).

Vance, T. (**1987**). *An Introduction to Japanese Phonology* (State University of New York Press, Albany, NY).

Vatikiotis-Bateson, E., and Kelso, J. A. S. (**1993**). "Rhythm type and articulatory dynamics in English, French, and Japanese," J. Phonetics **21**, 231–265.

Vaxelaire, B. (**1995**). "Single vs. double (abutted) consonants across speech rate: X-ray and acoustic data for French," in *Proceedings XIII Int. Conf. Phonetic Sci.*, Vol. **1**, 384–387, Stockholm, Royal Institute of Technology and Stockholm University.

Warner, N., and Arai, A. (**2001**). "Japanese mora-timing: A review," Phonetica **58**, 1–25.

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

Winer, B., Brown, R., and Michels, K. (**1991**) *Statistical Principles in Experimental Design* (McGraw-Hill, Boston).

Comparison of biomechanical modeling of register transitions and voice instabilities with excised larynx experiments

Isao T. Tokuda[a)]
School of Information Science, Japan Advanced Institute of Science and Technology, Ishikawa 923-1292, Japan

Jaromir Horáček
Institute of Thermomechanics, Academy of Sciences of the Czech Republic, Dolejškova 5, 182 00 Prague 8, Czech Republic

Jan G. Švec
Medical Healthcom Ltd., Voice Research Laboratory, Žitná 24, 120 00 Prague 2, Czech Republic

Hanspeter Herzel
Institute for Theoretical Biology, Humboldt University Berlin, Invalidenstrasse 43, D-10115 Berlin, Germany

(Received 13 May 2006; revised 26 April 2007; accepted 26 April 2007)

Voice instabilities were studied using excised human larynx experiments and biomechanical modeling. With a controlled elongation of the vocal folds, the experiments showed registers with chest-like and falsetto-like vibrations. Observed instabilities included abrupt jumps between the two registers exhibiting hysteresis, aphonic episodes, subharmonics, and chaos near the register transitions. In order to model these phenomena, a three-mass model was constructed by adding a third mass on top of the simplified two-mass model. Simulation studies showed that the three-mass model can vibrate in both chest-like and falsetto-like patterns. Variation of tension parameters which mimic activities of laryngeal muscles could induce transitions between both registers. For reduced prephonatory areas and damping constants, extended coexistence of chest and falsetto registers was found, in agreement with experimental data. Subharmonics and deterministic chaos were observed close to transitions between the registers. It is concluded that the abrupt changes between chest and falsetto registers can be understood as shifts in dominance of eigenmodes of the vocal folds. [DOI: 10.1121/1.2741210]

PACS number(s): 43.70.Aj, 43.70.Bk, 43.70.Gr [BHS] Pages: 519–531

I. INTRODUCTION

The careful excised larynx experiments of van den Berg (1957, 1959) paved the way for the myoelastic-aerodynamic theory of voice production. Two-mass models introduced by Ishizaka and Flanagan (1972) simulated the core mechanism of vocal fold vibrations—a phase shift of lower and upper edges of the vocal folds (Stevens, 1977). The resulting wave-like motion allows an efficient energy transfer from the airflow to the vibrating vocal folds (Titze, 1988). The combination of results for excised larynx experiments (Baer, 1981) and biomechanical modeling (Titze, 1973; Pelorson *et al.*, 1994; Story and Titze, 1995; Alipour-Haghighi *et al.*, 2000) has led to a reasonable understanding of the normal voice.

In addition to the normal voice with almost periodic vocal fold vibrations, a rich variety of voice instabilities has been discussed in the framework of nonlinear dynamics (Titze *et al.*, 1993). For example, subharmonics and chaos have been reported in newborn cries (Mende *et al.*, 1990) and in pathological voices (Herzel *et al.*, 1994). Moreover, nonlinear phenomena are widespread in animal voices (Wilden *et al.*, 1998; Fee *et al.*, 1998; Fitch *et al.*, 2002) and are exploited in contemporary vocal music (Neubauer *et al.*, 2004). In a case study of a patient with laryngeal paralysis, high-speed recordings of voice instabilities were related successfully to bifurcations in a biomechanical model (Mergell *et al.*, 2000). However, it is still quite difficult to model irregular vocal fold vibrations since a variety of mechanisms can lead to voice instabilities. Examples are left-right asymmetry (Ishizaka and Isshiki, 1976; Smith *et al.*, 1992; Tigges *et al.*, 1997), anterior-posterior modes (Neubauer *et al.*, 2001), excessively high subglottal pressure (Jiang *et al.*, 2001), interference with supraglottal resonances (Trendelenburg and Wullstein, 1935; Hatzikirou *et al.*, 2006), localized vocal fold lesions (Zhang *et al.*, 2004) or additional vibrating tissues such as vocal membranes (Mergell *et al.*, 1999). A quantitative comparison of experimentally observed voice instabilities and biomechanical modeling can profit from relatively simple model systems such as excised larynges. Indeed a remarkably rich repertoire of instabilities has been reported when subglottal pressure and vocal fold tension were varied systematically (van den Berg *et al.*, 1960; Berry *et al.*, 1996; Švec *et al.*, 1999; Horáček *et al.*, 2001; Jiang *et al.*, 2003).

The main focus of this paper is the transition between chest and falsetto registers and voice instabilities observed

a)Electronic mail: i.tokuda@biologie.hu-berlin.de

near the transition. Registers are one of the most important voice qualities for classifying the type of phonation. Up to now, registers have been modeled using sets of biomechanical parameters, which are chosen quite differently according to the phonation condition of each register (Story and Titze, 1995). Register transition can be realized by simulating a muscle activity that makes a bridge between such different parameter settings (Titze and Story, 2002). Little attention, however, has been paid to the hysteresis phenomenon of chest and falsetto and the voice instabilities observed during register transitions as observed experimentally (Berry *et al.*, 1996; Švec *et al.*, 1999; Horáček *et al.*, 2001). It should be noted that hysteresis implies coexistence of two registers within the same physiological condition of the vocal fold. Such situations have not yet been convincingly reproduced using mathematical models.

In this paper a three-mass model is introduced that focuses on very basic mechanisms of vocal fold vibrations. This model is used here to replicate as closely as possible the sudden chest-falsetto transitions and accompanying phenomena which were observed in experiments with excised human larynges when the vocal folds were symmetrically elongated and shortened. It turns out that even details of the spectrograms such as subharmonics near the register transitions can be reproduced by such a simple model. It will be shown that two-dimensional bifurcation diagrams similar to voice range profiles can be helpful to understand the complexity of voice instabilities. Finally, it is suggested that the findings are not sensitive to modeling details since the three-mass model represents just the core mechanisms of vibratory modes.

II. BIFURCATION SCENARIO ON REGISTER TRANSITION

In this section, some concepts of bifurcation theory are reviewed as a basic framework for understanding the register transitions. According to Titze (2000), registers are associated with quantal changes in voice quality due to gradual changes of physiological variables. For example, varying thyroarytenoid (TA) and cricothyroid (CT) muscle activities can induce abrupt register transitions. In the framework of nonlinear dynamics sudden qualitative changes due to slow and continuous parameter changes are termed bifurcations. In this paper, experimentally observed bifurcations are compared to model simulations. In order to introduce the relevant terminology, schematic one-dimensional and two-dimensional bifurcation diagrams are shown in Fig. 1. As early as 1918 Duffing observed that, at resonances of coupled oscillators, coexisting oscillation amplitudes can be found at the same value of a given control parameter (Duffing, 1918). Consequently, a slow increase of the parameter will lead to a jump of the amplitude at the value P_2 in Fig. 1(a). If the parameter is decreased slowly a jump to small amplitudes occurs at $P_1 < P_2$. This phenomenon is termed hysteresis and is a landmark of coexisting limit cycles. Such coexisting oscillatory regimes ("overlapping Arnold tongues") are generic features of coupled oscillators [see Berge *et al.* (1984) or Glass and Mackey (1988) for details]. Some of the dynamics of coupled oscillators can be illustrated by two-dimensional bifurcation diagrams such as in

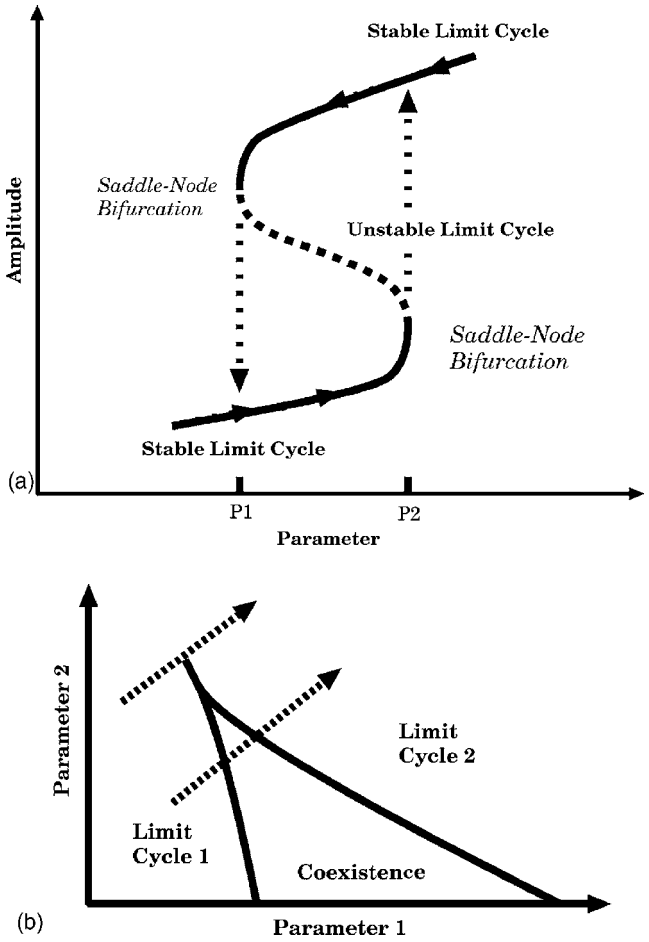


FIG. 1. (a) Schematic drawing of one-dimensional bifurcation diagram with coexisting stable limit cycles (solid lines) in the parameter range (P_1, P_2). The dashed line indicates an unstable limit cycle. The merging points of solid and dashed lines represent saddle-node bifurcations of limit cycles. (b) Schematic drawing of two-dimensional bifurcation diagram with a coexistence region of two distinct vibration patterns ("limit cycles"). A parameter variation along the lower arrow indicates sudden jumps in amplitude and frequency. Along the upper arrow a smooth transition is possible.

Fig. 1(b). Regarding *in vivo* register transitions the parameters might represent TA and CT muscles activities, which are considered as main control parameters. If these activities are increased, one can induce a transition from chest to falsetto registers (see Titze, 2000). For instance, in Fig. 10.11 of Titze (2000), two-dimensional plots are shown with TA and CT activities, which resemble the two-dimensional bifurcation diagram of Fig. 1. In the present study, fundamental frequency (F_0) and subglottal pressure were used as the control parameters. Such parameter controls are indicated by arrows in Fig. 1(b). The path along the lower arrow is associated with sudden breaks and hysteresis, and might represent spontaneous register changes of untrained singers or excised larynges. A more gradual transition can be achieved along the upper arrow via appropriate control of the two parameters, resembling *passagi* of trained singers.

The framework of bifurcation diagrams is helpful if the oscillations are much faster than parameter variations. Since variations of muscle activities *in vivo* or variations of the force in the present experiments are slow compared to vocal fold vibration changes, bifurcation theory is applicable. Reg-

ister breaks *in vivo* were found to be much quicker than muscular activation (Miller *et al.*, 2002) and thus their interpretation as bifurcations seem reasonable.

III. BIFURCATIONS IN EXCISED LARYNX EXPERIMENTS

In a study by Švec *et al.* (1999) the authors reanalyzed results of previous experiments with excised human larynx carried out by van den Berg *et al.* (1960) and van der Berg (1968). In the van den Berg experiment, the vibrating vocal folds were gradually symmetrically elongated and then shortened by applying an external force to the larynx while other parameters were held constant. The analysis revealed that chest-falsetto jumps can be seen as manifestations of bifurcations in the vocal-fold vibratory mechanism, i.e., the jumps in frequency accompanying the change of the register can arise without any sudden change of the tension of the vocal folds. Such sudden transitions reflect bifurcations of the underlying dynamical system.

The van den Berg experiment with excised human male larynges was repeated by Horáček *et al.* (2000, 2001). Tested larynges were fixed horizontally to a plate through which the airflow, heated to 37 °C and humidified, was delivered to the vocal folds. The air was passed through a tube, the dimensions of which corresponded to the volume of human subglottal space. The airflow rate was kept constant. Subglottal pressure in the tube was measured by pressure transducers. Neither supra-glottal space nor a vocal tract were included. Adduction of the vocal folds was kept constant. Longitudinal tension of the vocal folds, monitored by a force transducer, was smoothly increased or decreased by rotating the thyroid cartilage. Changes of the fundamental phonation frequency and of the vibration regimes were measured by microphone, pressure transducers, electroglottograph, and laser vibrometer. Furthermore, vocal fold oscillations were observed optically by strobovideoscopy and videokymography. More details on the experimental configuration are given in Horáček *et al.* (2000, 2001).

The focus of the previous studies was the chest-falsetto transition. Sudden jumps in the fundamental frequency [e.g., from 350 to 170 Hz (Horáček *et al.*, 2000), 600 to 200 Hz (Horáček *et al.*, 2001), or 208 to 80 Hz (Horáček *et al.*, 2004)] were associated with transitions from nearly harmonic vibrations without glottal closure to more intense oscillations with closure and strong higher harmonics. These vibration patterns resemble closely register transitions of the human voice (Hirano *et al.*, 1970; Hirano and Kakita, 1985) and thus careful excised experiments can elucidate the dynamical basis of voice registers.

Figure 2 displays a representative case where the spectrogram of the microphone signal is shown in the upper panel, and the corresponding slowly varying elongation force and subglottal pressure signals are shown in the middle and lower panels, respectively. The measurements were performed at a constant flow rate of 0.4 l/s. Consequently, changes of the vibration patterns are accompanied by changes of the subglottal pressure. Thus the force variations induce changes in fundamental frequency (see the spectrogram) and changes in pressure simultaneously. The first 25 s

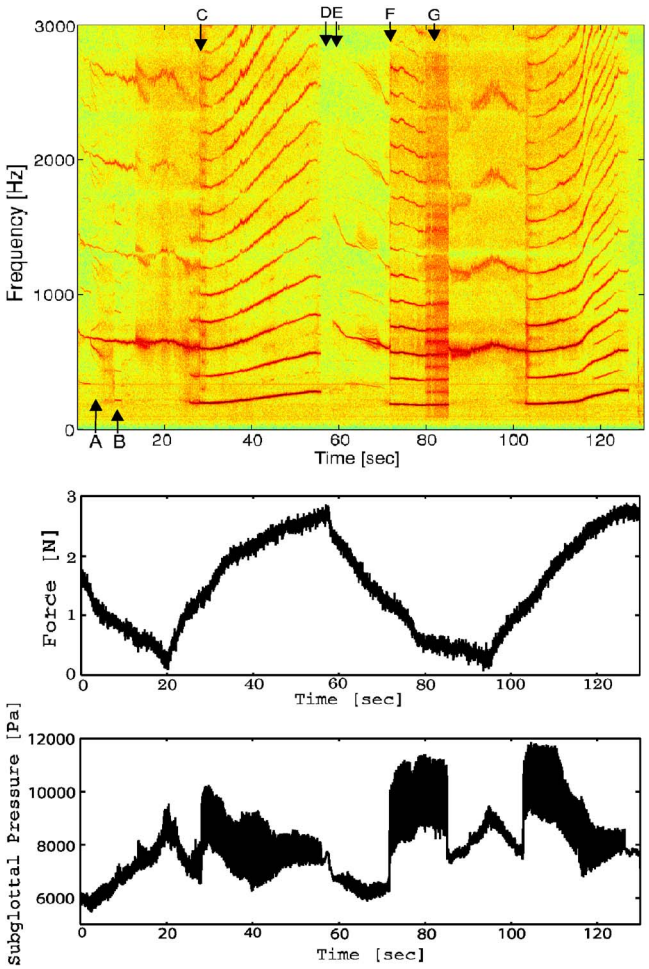


FIG. 2. (Color online) Upper graph: Spectrogram of the microphone signal from excised human larynx vibration; Middle graph: Force signal of vocal fold elongation; Lower graph: Subglottal pressure P_s.

represent falsetto-like vibrations with a fundamental frequency of about 600 Hz. Videokymography and EGG revealed absence of collisions of the vocal folds in this vibration regime (Horáček *et al.*, 2001). In the first 10 s additional spectral components are visible. Sidebands at about $t=5$ s (arrow "A") indicate a modulation, and around $t=8$ s (arrow "B") subharmonics at multiples of $F_0/3$ can be seen. At $t=25$ s (arrow "C"), a sudden transition to chest-like vibrations occurs. Large EGG amplitudes and videokymography revealed vocal fold closure and a pronounced wave-like vibration pattern (Horáček *et al.*, 2001). In this register, a clear increase of the fundamental frequency with the elongation force is found. Then a transition back to falsetto is observed around $t=60$ s (arrow "E"). This transition is accompanied by weak subharmonics at multiples of $F_0/2$ and a short interruption of phonation (an "aphonic episode;" arrow "D"). The falsetto regime again exhibits sidebands and it jumps back to chest-like vibrations around $t=70$ s (arrow "F"). Further transitions between chest and falsetto are again characterized by pronounced subharmonics. At about 82 s (arrow "G"), the subharmonic regime appears quite irregular. Thus the underlying dynamics might represent deterministic chaos.

The described falsetto-chest transitions are found at different values of the elongation force, in other words, the transitions exhibit hysteresis (Švec *et al.*, 1999; Horáček *et al.*, 2004). This phenomenon is related to the coexistence of different attractors representing coexistence of falsetto-like vibrations (an almost "sinusoidal limit cycle") and chest-like vibrations (a limit cycle with collisions).

In the following sections, a biomechanical model will be described that can reproduce the features illustrated in Fig. 2 such as register transitions with aphonia, subharmonics, hysteresis, and deterministic chaos.

IV. BIOMECHANICAL MODELING

The chest register of the human voice is characterized by pronounced wave-like vibrations of the whole vocal folds, i.e., by relatively large vibrating masses at medium vocal fold tension (Hirano *et al.*, 1970; Titze, 2000). In the falsetto register, a smaller portion of the vocal folds vibrates and a diminished closure of the glottis leads to weaker harmonics. In both registers, however, a wave-like vibration (sometimes called "mucosal wave") is visible (Berry *et al.*, 1996; Horáček *et al.*, 2001). These phase shifts of the lower and the upper portions of the vocal folds allow an energy transfer from the airflow to the vocal folds (Stevens, 1977; Titze, 1988). Consequently, both registers might be simulated by appropriate two-mass models that allow for such phase shifts and energy transfer.

There have been numerous studies on modeling vocal fold oscillations. Perhaps the most well known is the two-mass model by Ishizaka and Flanagan (1972). More advanced models can be classified into complex high-dimensional models (Alipour-Haghighi *et al.*, 2000) and simplified low-dimensional models (Pelorson *et al.*, 1994; Story and Titze, 1995; Titze and Story, 2002; Lous *et al.*, 1998; Sciamarella and d'Alessandro, 2004). An advantage of the complex models is that detailed anatomical and physiological structures of the vocal folds can be described. The drawback, however, is that many parameters are still not precisely known for such detailed modeling. A comprehensive bifurcation analysis becomes quite complicated with too many parameters. In this sense, a low-dimensional model was found to be more appropriate for the present study.

Considering the aerodynamics of glottal flow, Pelorson *et al.* (1994) took into account the formation of a free jet downstream of a moving separation point. Later Lous *et al.* (1998) introduced this flow-separation point to a biomechanical model, composed of two sets of three mass-less plates. Sciamarella and d'Alessandro (2004) utilized essentially the same model as Lous *et al.* (1998) and found coexistence of two vibratory patterns with different frequencies. However, the ratio of the frequencies, i.e., the characteristic leap interval (Miller *et al.*, 2002), was relatively small compared to the observations in the excised larynx experiments. Similarly, simulations of a related model of the bird syrinx exhibited only coexistence of oscillations with quite similar frequencies (Zaccarelli *et al.*, 2006). The Story-Titze three-mass model (1995) provides a faithful model to physiology, closely following the body-cover theory of Hirano (1974).

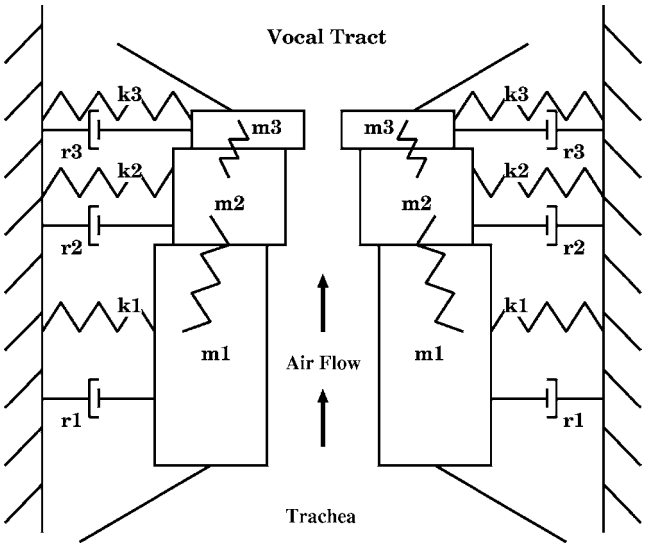


FIG. 3. Schematic illustration of the three-mass model.

This makes it possible to select the parameter values with close correspondence to physiological measurements. Titze and Story (2002) developed special rules for controlling the parameters of their three-mass model according to muscle activity, which successfully realize Hirano's four cases of phonations. However, chest and falsetto are associated in this model with quite different parameter settings (Story and Titze, 1995), which do not correspond to the findings from the present excised larynx experiments showing that chest and falsetto can coexist in the same parameter setting.

The aim of the present study is to demonstrate that a minimal model that focuses only on very basic mechanisms of vocal folds can capture the richness of bifurcations and replicate the characteristic chest-falsetto leap interval observed experimentally. A simple model was used here to keep the findings independent of the modeling details. The gross features of the register transitions and voice instabilities found in such a simple model are expected to be observed in more complex vocal fold models as well. A minimalist model may provide a reasonable starting point for further work. For example, extensions can be made such as modifying the effective vibrating masses, breaking left-right symmetry, adding sub- or supralaryngeal tubes, introducing more advanced aerodynamic formulas, and reconfiguring parameters to achieve closer correspondence to anatomy and physiology.

The present approach is based upon a three-mass model that can be considered as a combination of two two-mass models. The classical model of Ishizaka and Flanagan consists of two masses of $m_1=0.125$ g and $m_2=0.025$ g. In the present model, a third mass with $m_3=0.005$ g is added on top (see Fig. 3). The size of this additional mass is rather small so that it does not change the essence of the original two-mass configuration. The third mass has the following physiological function. In the excised larynx experiments, mucosal waves (defined, e.g, by Švec *et al.*, 2007) were visible in the videokymograms of both chest-like and falsetto-like registers (Berry *et al.*, 1996; Horáček *et al.*, 2001). In the falsetto-like register, the waves were visible only on the thin

upper medial portion of the folds and on the upper vocal fold surface. Addition of the third mass on the two-mass model divides the upper part of the vocal folds into two portions, which can vibrate out of phase. These phase differences can simulate the mucosal wave of the upper thin portion of the vocal folds as seen in the kymogram of the falsetto register (Horáček *et al.*, 2001). This approach enables us to adopt the standard parameter values, which have been established by Ishizaka and Flanagan (1972) and are subsequently referred to as the default parameter values in this paper. Following the simplifications of (1) neglecting the cubic nonlinearities of the oscillators, (2) neglecting the influence of vocal tract and subglottal resonances, and (3) neglecting the additional pressure drop at inlet and considering the Bernoulli flow only below the narrowest part of the glottis (Steinecke and Herzel, 1995), and assuming symmetry between the left and the right vocal folds, the model equations read

$$m_1\ddot{x}_1 + r_1\dot{x}_1 + k_1x_1 + \Theta(-a_1)c_1\left(\frac{a_1}{2l}\right) + k_{1,2}(x_1 - x_2) = ld_1P_1, \tag{1}$$

$$m_2\ddot{x}_2 + r_2\dot{x}_2 + k_2x_2 + \Theta(-a_2)c_2\left(\frac{a_2}{2l}\right) + k_{1,2}(x_2 - x_1) + k_{2,3}(x_2 - x_3) = ld_2P_2, \tag{2}$$

$$m_3\ddot{x}_3 + r_3\dot{x}_3 + k_3x_3 + \Theta(-a_3)c_3\left(\frac{a_3}{2l}\right) + k_{2,3}(x_3 - x_2) = ld_3P_3. \tag{3}$$

The dynamical variables x_i represent displacements of the masses m_i (lower mass: $i=1$, middle mass: $i=2$, upper mass: $i=3$), where the corresponding glottal areas are given by $a_i = a_{0i} + 2lx_i$ (a_{0i}: prephonatory area, l: length of the glottis). The constant parameters r_i, k_i, c_i, d_i represent damping, stiffness, collision stiffness, and thickness of the masses m_i, respectively, whereas $k_{i,j}$ represents coupling strength between two masses m_i and m_j. The collision function is approximated as $\Theta(x)=0$ $(x \leqslant 0)$; $\Theta(x)=\tanh(1000x)$ $(0<x)$. The pressures P_i, which act on the masses m_i, and the glottal volume flow velocity U are given by

$$P_1 = P_s\left[1 - \Theta(a_{\min})\left(\frac{a_{\min}}{a_1}\right)^2\right]\Theta(a_1), \tag{4}$$

$$P_2 = P_s\left[1 - \Theta(a_{\min})\left(\frac{a_{\min}}{a_2}\right)^2\right]\Theta(a_1)\Theta(a_2)\Theta(a_1 - a_3) \times \Theta(a_2 - a_3), \tag{5}$$

$$P_3 = 0, \tag{6}$$

$$U = \sqrt{\frac{2P_s}{\varrho}}a_{\min}\Theta(a_{\min}), \tag{7}$$

where $a_{\min}=\min(a_1,a_2,a_3)$ corresponds to the narrowest glottal area and ϱ represents the air density (Steinecke and Herzel, 1995).

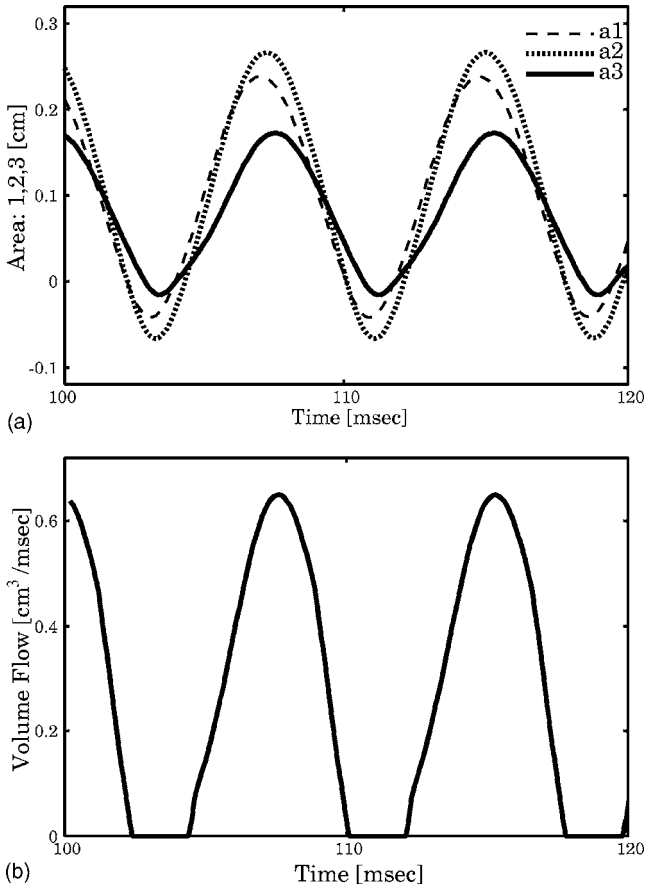


FIG. 4. (a) Chest-like vibration observed for the default parameter setting. Time series show glottal areas $a_1(t)$, $a_2(t)$, and $a_3(t)$ between the three masses. (b) Corresponding volume flow U.

V. SIMULATION RESULTS

A. Simulating chest-like and falsetto-like vibrations

This section shows simulation results of the biomechanical model. Figure 4(a) displays the vibration pattern of the chest-like phonation produced by the model. The default parameter values, initial conditions, and simulation details are compiled in the Appendix. The minor extension of the biomechanical model has little effect on the chest-like vibrations well known from the two-mass model. The phase difference is 23.5° between the lower and upper masses. Although this phase difference is somewhat smaller than the standard models (Story and Titze, 1995), it can be increased if the mutual couplings ($k_{1,2}$, $k_{2,3}$) or the subglottal pressure are reduced. The present simulation at the default parameter setting demonstrates that a relatively small phase difference is sufficient for robust self-sustained oscillations with a subglottal pressure of 0.008 g/cm ms^2. Complete glottal closure leading to the slightly skewed volume flow is seen in Fig. 4(b). This chest-like vibration has a frequency of 129.9 Hz.

In the three-mass model, the middle mass plays a particular role. In the chest register, it serves as a kind of valve to reduce the airflow during glottal opening compared to glottal closing. This is achieved by the phase shift to the lower mass oscillation discussed earlier. If the stiffness of the second mass is increased from the default value $k_2 = 0.008$ g/ms^2 to higher values, a qualitatively distinct vibra-

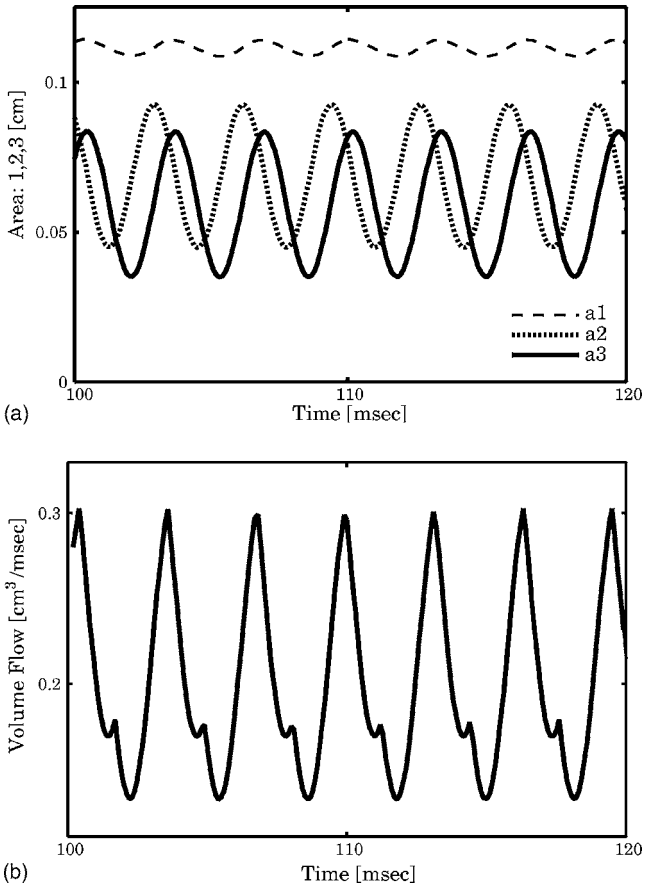


FIG. 5. (a) Falsetto-like vibration observed for the default parameter setting with $k_2=0.08$. Time series show glottal areas $a_1(t)$, $a_2(t)$, and $a_3(t)$ between the three masses. (b) Corresponding volume flow U.

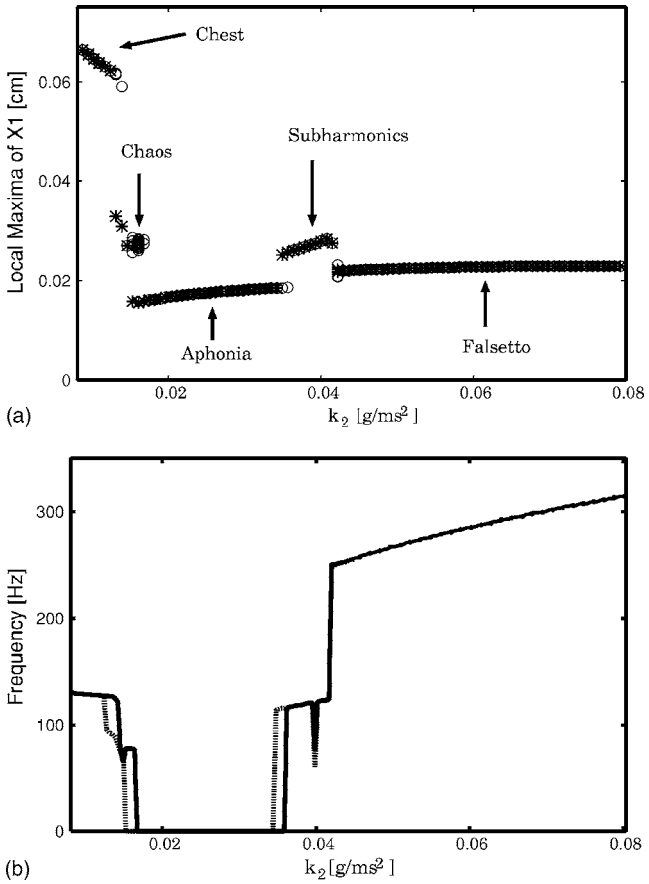


FIG. 6. (a) Bifurcation diagram that plots local maxima of the variable x_1 when changing the bifurcation parameter $k_2 \in [0.008, 0.08]$. The circles and the crosses correspond to increase and decrease of the bifurcation parameter, respectively. Note that "subharmonics" are indicated around $k_2=0.04$, since another drawing, which plotted maxima of x_2, showed two branches clearly at this region (not shown here). The points corresponding to "aphonia" represent a stable equilibrium, where these particular plots are no longer local maxima. (b) Frequencies corresponding to (a). The solid and dotted lines correspond to increase and decrease of the bifurcation parameter, respectively.

tion pattern can be established. For $k_2=0.08$ g/ms^2, Fig. 5(a) shows phase-shifted vibrations of the upper two masses whereas the lowest mass vibrates with small amplitude being displaced far from midline. The observed frequency of 314.5 Hz is much higher than the chest-like vibration. Due to the lack of the vocal fold collision, the volume flow wave form in Fig. 5(b) does not exhibit a closed phase. In this way, falsetto-like vibrations can be simulated.

A simple interpretation of the switch between chest and falsetto can be given as follows. For simplicity, let us assume that the three masses have no mutual connections, no damping, no effect from the pressure. Then each mass constitutes an independent oscillator with a frequency of $f_i = (1000/2\pi)\sqrt{k_i/m_i}$ Hz ($i=1,2,3$; the scaling factor is due to the unit g/ms^2 of the stiffness parameter). The default parameter setting yields $f_1=127.3$ Hz, $f_2=90.0$ Hz, $f_3=225.1$ Hz. In this case the similarity of the frequencies f_1 and f_2 leads to chest-like low-frequency vibrations. For $k_2=0.08$ g/ms^2, the frequency of the middle mass becomes $f_2=284.7$ Hz. Now the middle mass has a similar frequency as the upper mass leading to falsetto-like high frequency vibrations. In this way, switching between chest and falsetto could be understood as a frequency selection by the middle mass. Register transitions are generally accompanied by changes of many parameters, such as stiffness, mass, and geometry (see, e.g, Story and Titze, 1995). Since detailed changes of these parameters are not well known, only the stiffness parameter k_2 of the middle mass was varied for simplicity. In Sec. V B, a more detailed eigenfrequency analysis is presented to study correspondence between the eigenmodes and the observed registers.

Figure 6 displays a one-dimensional bifurcation diagram with transitions between chest and falsetto. Among dynamical variables x_1, x_2, x_3, or U, which yield qualitatively the same bifurcation structure, local maxima of the x_1 component are drawn. The stiffness parameter was varied gradually from $k_2=0.008$ g/ms^2 [chest-like phonation shown in Fig. 4(a)] to $k_2=0.08$ g/ms^2 [falsetto as in Fig. 5(a)]. In order to look for hysteresis, the parameter was varied in both directions which are indicated by the symbols and line styles in the plots. Below $k_2<0.01$ g/ms^2, chest-like vibrations were found with a low fundamental frequency of about 130 Hz and large amplitudes. Around $k_2=0.01$ g/ms^2, the vibrations became chaotic and switched to a stable steady state ("aphonia"). Increasing the stiffness to values about $k_2=0.03$ g/ms^2 led to subharmonics and subsequently to

falsetto-like dynamics with a fundamental frequency of about 300 Hz and small amplitudes. The comparison of solid lines (increase of k_2) and dotted lines (decrease of k_2) in the lower graph reveals that there was only a weak hysteresis at the transitions to aphonia. In summary, the model could produce register transitions with subharmonics, chaos, and aphonia. However, for the chosen parameters no coexistence region of chest and falsetto was found. Consequently, Sec. V C focuses on a modified parameter configuration exhibiting pronounced coexistence of registers as described experimentally (Berry *et al.*, 1996; Švec *et al.*, 1999; Horáček *et al.*, 2004).

B. Eigenfrequency analysis

To gain more insight into the chest-like and the falsetto-like oscillations observed in Sec. V A, an eigenfrequency analysis was carried out for the three-mass model. Without the pressure or collision terms, linearization of the model Eqs. (1)–(3) gives

$$\frac{d}{dt}\begin{pmatrix} \delta x_1 \\ \delta \dot{x}_1 \\ \delta x_2 \\ \delta \dot{x}_2 \\ \delta x_3 \\ \delta \dot{x}_3 \end{pmatrix} = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 & 0 \\ -\frac{k_1 + k_{1,2}}{m_1} & -\frac{r_1}{m_1} & \frac{k_{1,2}}{m_1} & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ \frac{k_{1,2}}{m_2} & 0 & -\frac{k_2 + k_{1,2} + k_{2,3}}{m_2} & -\frac{r_2}{m_2} & \frac{k_{2,3}}{m_2} & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & \frac{k_{2,3}}{m_3} & 0 & -\frac{k_3 + k_{2,3}}{m_3} & -\frac{r_3}{m_3} \end{pmatrix} \begin{pmatrix} \delta x_1 \\ \delta \dot{x}_1 \\ \delta x_2 \\ \delta \dot{x}_2 \\ \delta x_3 \\ \delta \dot{x}_3 \end{pmatrix}. \tag{8}$$

Oscillatory modes of the linearized equation are obtained by computing the eigenvalues λ_i of the above-presented matrix. The eigenfrequencies are calculated directly from the eigenvalues as $F_i = 1000|\Im(\lambda_i)|/(2\pi)$ Hz, where $\Im$ gives imaginary part of λ.

First, the default parameter setting will be analyzed that generates chest-like vibrations. The eigenvalues are obtained as $\lambda_{1,2} = -0.09 \pm 0.78$, $\lambda_{3,4} = -0.13 \pm 1.28$, $\lambda_{5,6} = -0.46 \pm 2.0$, which yield the eigenfrequencies of $F_{1,2} = 123.8$ Hz, $F_{3,4} = 204.3$ Hz, $F_{5,6} = 318.7$ Hz. To determine the eigenfrequency which corresponds to the chest-like vibration observed in the simulation, the associated eigenmodes (δx_1, $\delta \dot{x}_1$, δx_2, $\delta \dot{x}_2$, δx_3, $\delta \dot{x}_3$) can be compared with the real simulation. In Fig. 7(a), three eigenmodes are drawn by dotted lines. For comparison, the oscillatory mode of the simulation computed by ($\max x_1 - \min x_1$, $\max \dot{x}_1 - \min \dot{x}_1$, $\max x_2 - \min x_2$, $\max \dot{x}_2 - \min \dot{x}_2$, $\max x_3 - \min x_3$, $\max \dot{x}_3 - \min \dot{x}_3$) is simultaneously drawn by a solid line. Note that the mode vectors are normalized in such a way that the maximum component becomes unity. A clear correspondence can be seen between the first eigenmode and the simulation with a large amplitude of the lower mass oscillation. The corresponding eigenfrequency of $F_{1,2} = 123.8$ Hz provides a good approximation of the observed chest frequency of 129.9 Hz.

Now, the falsetto-like vibration is analyzed with the same default parameter setting except $k_2 = 0.08$ g/ms^2. The eigenvalues are obtained as $\lambda_{1,2} = -0.08 \pm 0.88$, $\lambda_{1,2} = -0.42 \pm 1.85$, $\lambda_{1,2} = -0.18 \pm 2.2$, which yield the eigenfrequencies of $F_{1,2} = 140.3$ Hz, $F_{3,4} = 294.7$ Hz, $F_{5,6} = 352.9$ Hz. Figure 7(b) draws the three eigenmodes together with the simulation mode. A good correspondence is found between the third mode and the simulation, whose mode is dominated by the upper two masses. The corresponding eigenfrequency of $F_{5,6} = 352.9$ Hz is slightly larger but comparable to the observed falsetto frequency of 314.5 Hz.

The deviations between the eigenfrequency and the simulation might be due to (i) linearization of the original equations or (ii) neglected pressure terms ("vacuum condition"). To quantify the latter effect, the eigenfrequency analysis including the pressure terms was carried out. The results are summarized in Table I and show that the analysis including the pressure terms is comparable to the vacuum case. Considering its simplicity and efficiency, the eigenfrequency analysis with vacuum condition turned out to be a helpful guide to extract the chest and falsetto modes from the three-mass model.

C. Simulating hysteresis at the register transition

In the preceding sections, the simplified two-mass model (Steinecke and Herzel, 1995) extended by a small third mass on top was analyzed. Simply by increasing the stiffness k_2 of the middle mass, the model could generate both the chest-like and falsetto-like registers as shown in Fig. 5(a). For the default parameters there was, however, an aphonic regime separating the two registers ["aphonia" in Fig. 6(a)]. Experimentally abrupt transitions between registers have been observed frequently (van den Berg *et al.*, 1960; Berry *et al.*, 1996; Švec *et al.*, 1999; Horáček *et al.*, 2004). Therefore, the parameter setting was changed to obtain parameter region

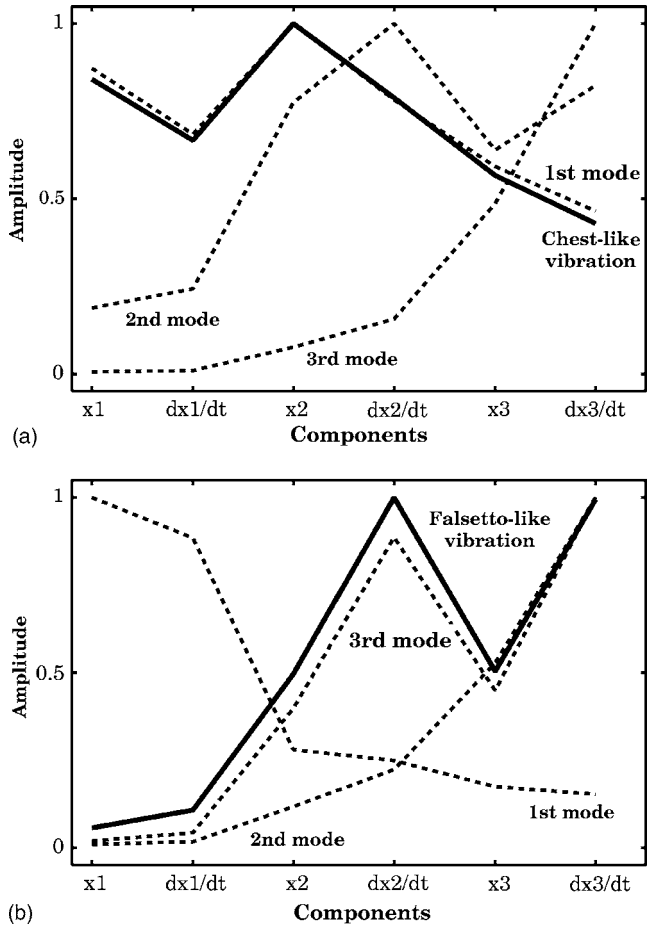


FIG. 7. Eigenfrequency analysis of (a) default parameter setting and (b) with an increased value of k_2=0.08. Normalized absolute components of three eigenmodes ($|\delta x_1|$, $|\delta \dot{x}_1|$, $|\delta x_2|$, $|\delta \dot{x}_2|$, $|\delta x_3|$, $|\delta \dot{x}_3|$) are plotted by dotted lines, whereas oscillatory mode of the real simulation is indicated by a solid line.

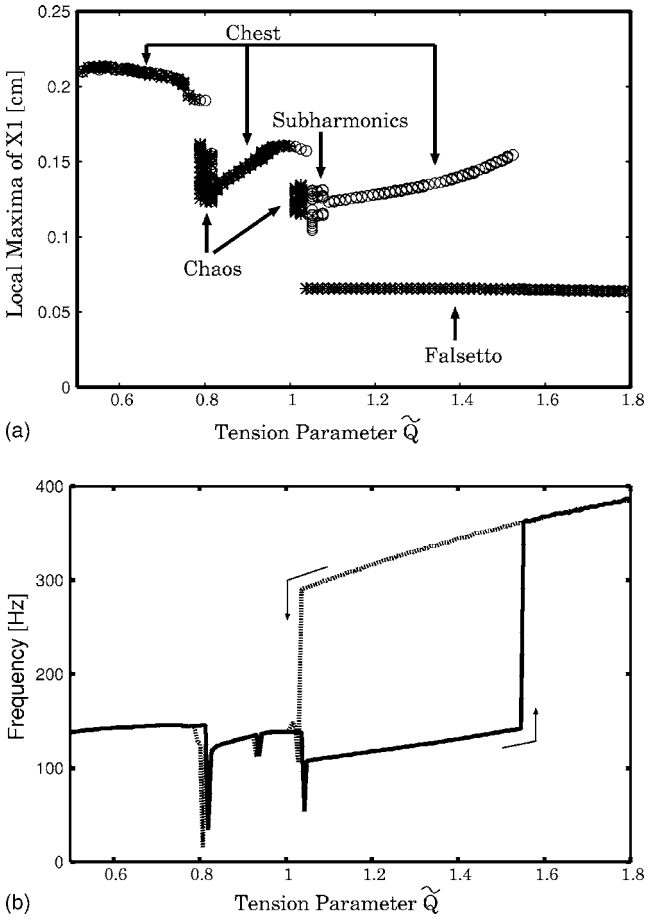


FIG. 8. (a) Bifurcation diagram that plots local maxima of the variable x_1 when changing the bifurcation parameter $\tilde{Q} \in [0.5, 1.8]$. The circles and the crosses correspond to increase and decrease of the bifurcation parameter, respectively. (b) Frequencies corresponding to (a). The solid and dotted lines correspond to increase and decrease of the bifurcation parameter.

with coexisting registers. It is known (Titze, 1988) that small prephonatory opening described by small a_{0i} values and low damping constants r_i support oscillations. Consequently, a_{01}, a_{02}, a_{03}, and r_1 were reduced to remove the aphonic regime between the registers. Furthermore the coupling constant $k_{1,2}$ was modified and a new tension parameter $\tilde{Q}$ [defined by Eqs. (A1) and (A2) in the Appendix] was introduced in a similar manner as in earlier papers (Smith *et al.*, 1992; Steinecke and Herzel, 1995). This new parameter scales the eigenfrequency of the second mass and is therefore an appropriate tuning parameter that governs the transitions from chest to falsetto and vice versa. The modified set of parameters is given in the Appendix.

TABLE I. Summary of eigenfrequency analysis with and without the pressure terms. The obtained eigenfrequencies are compared with the simulation results. Analysis with the pressure terms was carried out in the same manner as the one without the pressure terms (described in the main text), except that the model Eqs. (1)–(3) including the pressure terms were linearized around its equilibrium. The equilibrium was traced numerically by the shooting method.

	k_2=0.008 g/ms^2	k_2=0.08 g/ms^2
Eigenfrequency in vacuum condition	123.8 Hz	352.9 Hz
Eigenfrequency with pressure	116.1 Hz	295.5 Hz
Simulation	129.9 Hz	314.5 Hz

Figure 8(a) shows a one-dimensional bifurcation diagram, which plots local maxima of the x_1 component when the tension parameter is varied between $\tilde{Q} \in [0.5, 1.8]$. There is no longer an extended region of aphonia between the registers [compared with Fig. 6(a)]. Sudden jumps between the registers are accompanied here by subharmonics and deterministic chaos. The comparison of increasing versus decreasing tension reveals a large range of coexistence of chest-like and falsetto-like phonation: $1.05 < \tilde{Q} < 1.55$. This can be confirmed as the hysteresis in the spectrograms drawn in Fig. 9. The spectrogram obtained by increasing the parameter $\tilde{Q}$ is dominated by low-fundamental frequency vibrations in Fig. 9(a), whereas when decreasing the parameter $\tilde{Q}$ the spectrogram displays an enlarged range of high-fundamental frequency vibrations in Fig. 9(b). Directions of the switch between the low and high fundamental frequencies are more clearly indicated in Fig. 8(b).

These bifurcation diagrams illustrate the well-known complexity of coupled nonlinear oscillators (Berge *et al.*, 1984; Glass and Mackey, 1988). For a more comprehensive

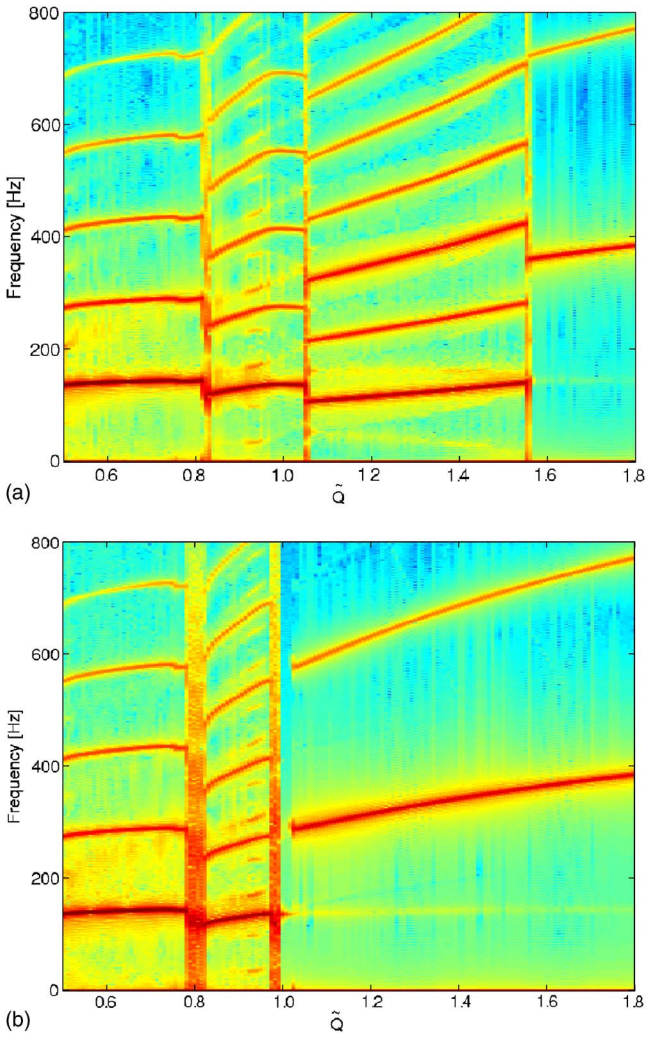


FIG. 9. (Color online) Spectrograms corresponding to Fig. 8. (a) and (b) correspond to the case of increasing and decreasing the bifurcation parameter $\tilde{Q}$, respectively.

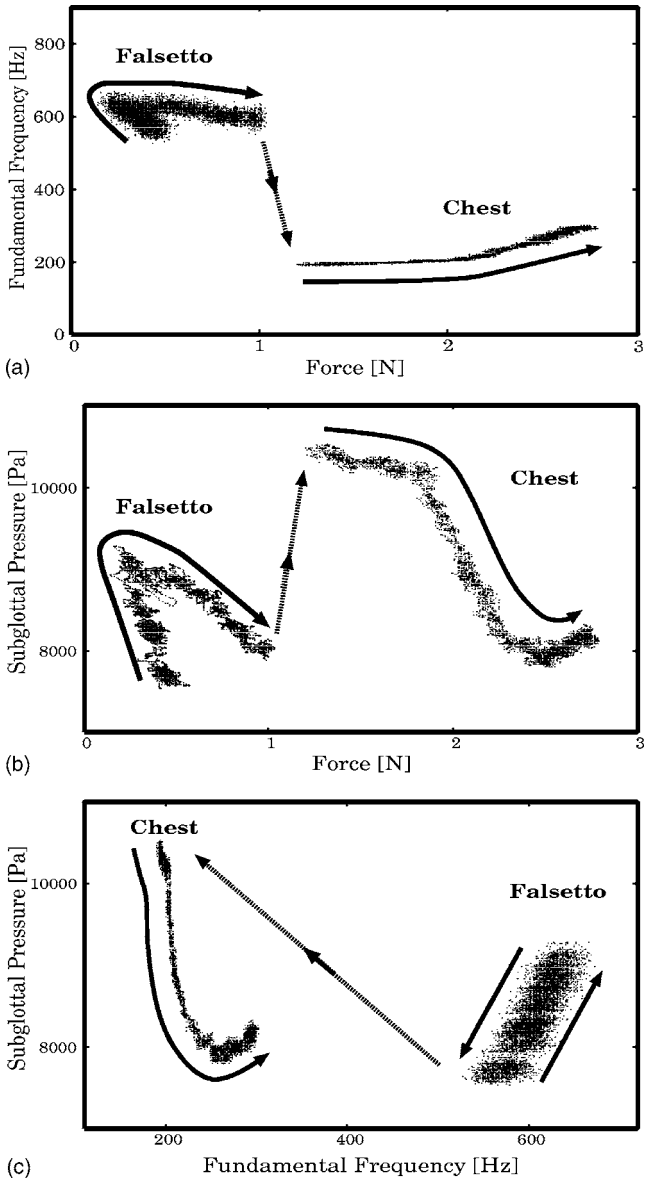


FIG. 10. Simultaneous plots of force, fundamental frequency, and subglottal pressure observed from the excised larynx experiment in the range from 88 to 126 s (see Fig. 2). The arrows indicate directions of dynamical change in the parameter space. (a) Force↔Fundamental frequency; (b) Force↔Subglottal pressure; (c) Fundamental frequency↔Subglottal pressure.

analysis of coupled oscillators, two-dimensional bifurcation diagrams are appropriate and thus Sec. V D is devoted to the study of parameter planes spanned by two parameters such as loudness and fundamental frequency.

D. Two-dimensional bifurcation diagram: Experiments and simulations

As discussed earlier, parameter variations may lead to rather complex bifurcation scenarios. In order to get a better understanding of the various transition phenomena, two-dimensional bifurcation diagrams are helpful. For example, voice instabilities due to unilateral laryngeal paralysis have been studied in the parameter plane of left-right frequency ratio and subglottal pressure. Moreover, voice range profiles can be considered as two-dimensional bifurcation diagrams with the fundamental frequency and sound pressure level as the bifurcation parameters (Schutte and Seidner, 1983; Pabon and Plomp, 1988).

Along these lines, the experimental data from Fig. 2 are reanalyzed. In the experiment, the elongation force was varied systematically, while the flow was kept constant. In this way, the force influenced the fundamental frequency and the subglottal pressure in a complicated manner. For the experimental data that exhibits falsetto-chest transition from 88 to 126 s, the fundamental frequency is extracted and plotted against the subglottal pressure and the force. Figure 10(a) shows that the fundamental frequency in the falsetto or chest registers is not strongly affected by the force. In Fig. 10(b), a monotonic dependence of the force and the subglottal pressure is observed. The register transition leads to an abrupt increase of the subglottal pressure.

In the present model, fundamental frequency and subglottal pressure serve as major control parameters. These parameters are plotted in Fig. 10(c). The falsetto-chest transition corresponds to a jump from high to low fundamental frequencies accompanied by increase of the subglottal pressure. This observation guided parameter variations in the modeling study.

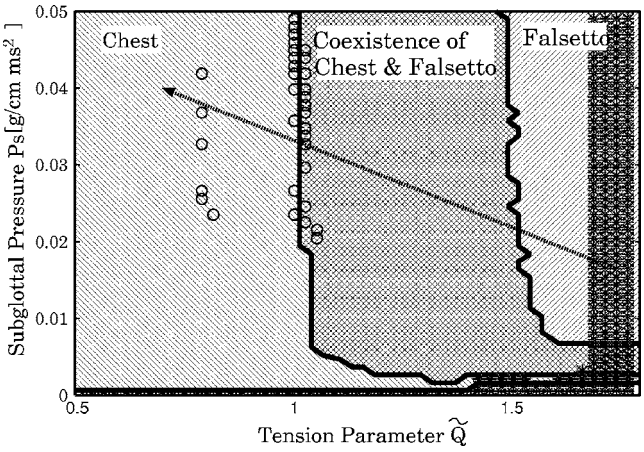


FIG. 11. Two-dimensional bifurcation diagram of $(\tilde{Q}, P_s) \in [0,5, 1.8] \times [0, 0.05]$. The three shaded regions correspond to the existence domain of chest register, falsetto register, and both. The circles indicate the existence of chaos, whereas the crosses indicate the existence of aphonia. The arrowed line indicates the bifurcation path exploited to produce the spectrogram in Fig. 12.

The present three-mass model is not realistic enough to reproduce the precise details of force-induced variations in the experiment but the main features of the excised larynx experiment can be simulated. Figure 11 shows a two-dimensional bifurcation diagram of the model. On the left side the chest register dominates, whereas on the right side the falsetto-like vibrations are found. In the range $1 < \tilde{Q} < 1.5$, coexistence of both registers is observed. Circles and crosses indicate parameter values with chaos and stable steady states ("aphonia"), respectively. The thick line with the arrow indicates the parameter variation corresponding to the spectrogram and bifurcation diagram in Figs. 12(a) and 12(b). For the first 50 s, decreasing $\tilde{Q}$ and increasing P_s induced transitions from aphonia (0–12 s) to falsetto (12–34 s), then to chaos (around 36 s) followed by chest (36–50 s) with subharmonics (39 s) and chaos (46 s). At $t = 50$ s the parameter variation is reversed and the backward bifurcation scenario is visible. The spectrogram, however, is not symmetric pointing to hysteresis of the transitions. For example, another subharmonic regime is observed around 68 s, the jump to falsetto is significantly delayed and there is finally no jump to aphonia, i.e., the falsetto register persists. The lower graph with the maxima of x_1 during these transitions is consistent with the interpretation of the spectrogram—subharmonics are characterized by multiple maxima and chaos leads to noisy clouds of points.

The simulations reveal that even the proposed simple three-mass model can reproduce many of the complex transitions observed experimentally. Between the chest-like and the falsetto-like registers a wide range of coexistence regime lies, which gives rise to sudden frequency jumps accompanied by hysteresis. Near the frequency jumps, voice instabilities due to subharmonics and deterministic chaos may occur. One of the important findings is that for almost all choices of bifurcation path that makes a bridge between the chest-like and the falsetto-like registers, it is inevitable to experience such frequency jumps and voice instabilities. This agrees with the excised larynx experiment.

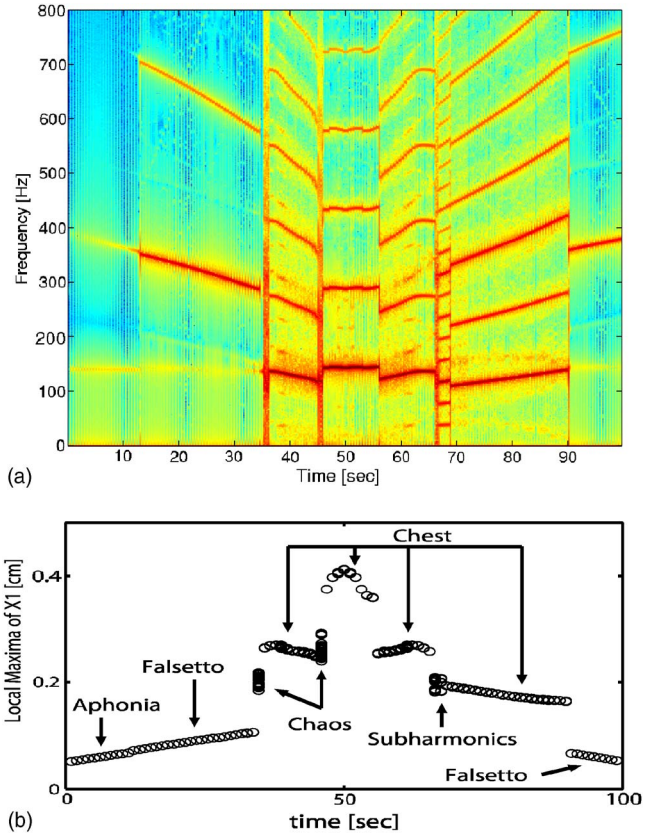


FIG. 12. (a): (Color online) Spectrogram of the three-mass model simulating a linear change from $(\tilde{Q}, P_s) = (1.8, 0.015)$ to (0.7,0.04) and then back to (1.8, 0.015). This bifurcation path in the $(\tilde{Q}, P_s)$ plane is drawn as the arrow in Fig. 11; (b) Corresponding bifurcation diagram. In the transition from "aphonia" to "falsetto," the corresponding circles represent stable equilibria and maxima of the limit cycle, respectively.

VI. SUMMARY AND DISCUSSION

It was the aim of this study to compare register transitions and complex bifurcation scenarios in excised larynx experiments with biomechanical simulations. The main features of the experimentally observed instabilities were large frequency jumps exhibiting hysteresis, aphonic episodes, subharmonics, and chaos near the register transitions. Keeping in mind the large heterogeneity of excised larynx experiments, it was not our intention to reproduce all details of spectrograms as shown in Fig. 2. For instance, fundamental frequency jumps between chest-like and falsetto-like vibrations are highly variable (Švec and Pešák, 1994) and thus the original parameters of two-mass models were kept with a relatively low fundamental frequency.

Here a three-mass model was designed that can vibrate per construction in a chest-like and falsetto-like mode. The chest register was essentially identical to oscillation of the original two-mass model (Ishizaka and Flanagan, 1972; Steinecke and Herzel, 1995) whereas the falsetto register is characterized by large phase-shifted oscillations of the upper two masses. A variation of tension parameters which mimic activities of laryngeal muscles could induce transitions between these registers. For reduced prephonatory areas and damping constants, extended coexistence of chest and falsetto was found, which corresponds to experimental data.

Whereas subharmonics and deterministic chaos were found to be related to left-right asymmetry or excessively high pressure in earlier studies (Steinecke and Herzel, 1995; Jiang *et al.*, 2001), the present observation required neither of such situations. The results predict specific new instabilities near the chest-falsetto transition that can be tested using appropriate high-speed recordings (Wittenberg *et al.*, 2000; Eysholdt *et al.*, 1996) or video-kymography (Švec and Schutte, 1996). Since this study was based upon a minimal model that focused on very basic mechanisms of the vocal folds, the phenomena found here can be expected to be quite common in many other vocal fold models, which can be physiologically more realistic. Dynamic behavior of any vibrating system can be characterized by extracting the specific eigenmodes that govern the resulting frequency and pattern of vibration (e.g, Titze and Strong, 1975; Berry *et al.*, 1994). As shown here, the mechanism of abrupt chest-falsetto register transitions could generally be understood as a spontaneous shift in dominance between different eigenmodes of the vocal folds. While the specific eigenmodes differ among various models, the underlying dynamic mechanism of the abrupt changes of vibration can be expected to be similar.

In some experiments using other human larynges, no clear falsetto-like vibration but the coexistence of two chest-like vibrations associated with similar frequencies was seen. In such experiments, voice instabilities seem to be induced essentially by left-right asymmetry. A future study will focus on modeling such examples using an asymmetric version of the present model.

The observed complexity of the signals produced by the simple three-mass model is not really surprising since the model can be regarded as a system of coupled oscillators representing a low-frequency chest mode and a high-frequency falsetto mode. It is known that coupled nonlinear oscillators generically exhibit synchronization ("Arnold tongues"), incommensurable frequencies ("tori," termed biphonation in voice research), frequency jumps, subharmonics, and chaos (Huygens, 1673; Berge *et al.*, 1984; Glass and Mackey, 1988). These features have been described extensively for the asymmetric two-mass model (Steinecke and Herzel, 1995).

It is appropriate to analyze coupled oscillators with the aid of two-dimensional bifurcation diagrams. In this way many instabilities can be traced back to transitions between different synchronization regimes. Interestingly, two-dimensional presentations of parameter variations have already some tradition in voice science. For instance, in Titze (2000), register transitions were discussed in the plane of CT and TA muscle activities. Another example is the widely used voice range profile. Here the oscillatory regions are explored in the plane of fundamental frequency and sound pressure level. If such an analysis includes registers and voice instabilities the similarity to two-dimensional bifurcation diagrams as in Figs. 10 and 11 is evident. Along these lines, a comparison of voice instabilities and biomechanical modeling seems possible also for patients with voice disorders.

In future studies it is planned to take into account more detailed physiological properties of the vocal folds. As demonstrated by Story and Titze (1995), the body-cover theory (Hirano, 1974) can be related to simplified models. Rules can be developed to associate muscle activity with the parameter configurations (Titze and Story, 2002). Towards such more realistic models, the body-cover structure with physiological parameters will be introduced to the model in a forthcoming study. Moreover, the role of sub- and supraglottal resonances should be studied in detail (Titze, 2000; Zhang *et al.*, 2006). The two-dimensional bifurcation diagrams of the present core model provide a starting point of more detailed studies.

ACKNOWLEDGMENTS

This work was supported by the Alexander-von-Humboldt Foundation and by the Grant Agency of the Academy of Sciences of the Czech Republic by Project No. IAA2076401 "Mathematical modeling of human vocal fold oscillations."

APPENDIX: MODEL PARAMETERS AND SIMULATION DETAILS

Here the default parameter values adopted from the simplified two-mass model (Steinecke and Herzel, 1995) used in Sec. V are listed:

$m_1 = 0.125$ g; $m_2 = 0.025$ g; $m_3 = 0.005$ g;

$d_1 = 0.25$ cm; $d_2 = 0.05$ cm; $d_3 = 0.01$ cm;

$r_1 = 0.02$ g/ms; $r_2 = 0.005$ g/ms;

$r_3 = 0.005$ g/ms; $k_1 = 0.08$ g/ms^2; $k_2 = 0.008$ g/ms^2;

$k_3 = 0.01$ g/ms^2; $k_{1,2} = 0.025$ g/ms^2;

$k_{2,3} = 0.01$ g/ms^2; $c_1 = 3k_1$; $c_2 = 3k_2$; $c_3 = 3k_3$;

$a_{01} = 0.05$ cm^2; $a_{02} = 0.05$ cm^2; $a_{03} = 0.05$ cm^2;

$l = 1.4$ cm; $\varrho = 0.00113$ g/cm^3;

$P_s = 0.008$ g/cm ms^2.

By changing the stiffness of the second mass to $k_2 = 0.08$ g/ms^2, the falsetto-like oscillation shown in Fig. 5(a) could be simulated.

To study hysteresis of the chest-falsetto registers, the following parameters were modified compared to the default setting: $r_1 = 0.01$ g/ms; $k_{1,2} = 0.05$ g/ms^2; $a_{01} = 0.01$ cm^2; $a_{02} = 0.01$ cm^2; $a_{03} = 0.01$ cm^2. To observe transitions between the chest and the falsetto registers, the tension parameter $\tilde{Q}$ was introduced to scale the eigenfrequency of the second mass as

$$k_2 = \tilde{Q} k_{20}, \tag{A1}$$

$$m_2 = m_{20}/\tilde{Q}, \tag{A2}$$

where the default parameters were $m_{20} = 0.025$ g and $k_{20} = 0.08$ g/ms^2.

The initial values for all simulations were set as $x_1=x_2=x_3=0.1$ and $\dot{x}_1=\dot{x}_2=\dot{x}_3=0$. To integrate the three-mass model Eqs. (1)–(3), the fourth-order Runge-Kutta method was applied with an integration step of $\Delta t=1/20$ ms. A smaller integration step was tried and it was confirmed that essentially the same results can be obtained.

To draw bifurcation diagrams, 20 local maxima of the first component $x_1(t)$ were plotted after discarding the first 100 maxima as transients. For the next parameter values, the final state of the preceding simulation was used as the initial condition. To obtain the two-dimensional bifurcation diagram, the parameters were varied in four different directions, namely, fixed $\tilde{Q}$ with increasing or decreasing P_s or fixed P_s with increasing or decreasing $\tilde{Q}$.

In Figs. 6(a), 8(a), and 12(b), the labels "chaos," "chest," "subharmonics," and "falsetto" were put according to the following rule. Local maxima of the x_1 component were defined as the Poincaré section, which transforms continuous-time dynamics to discrete-time mapping. With respect to the period number of the discrete mapping, dynamics were classified, e.g., period-1: chest or falsetto, more than 1 period: subharmonics, nonperiod: biphonation or chaos. If the frequency was less (more) than 200 Hz, the period-1 corresponded to chest (falsetto). Biphonation and chaos were classified by looking at the spectrogram.

The spectrograms were computed using the x_2 component with the following parameters. Sampling rate: 4 kHz; Window length: 2048 sample points; Overlap: 0; Windowing: Hanning. Note that other components such as $x_{1,3}$, $\dot{x}_{1,2,3}$, or glottal flow velocity U give essentially the same spectrogram.

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.

Baer, T. (**1981**). "Observation of vocal fold vibration: Measurement of excised larynges," in *Vocal Fold Physiology*, edited by K. N. Stevens and M. Hirano (University of Tokyo, Tokyo), pp. 119–133.

Berge, P., Pomeau, Y., and Vidal, C. (**1984**) *Order Within Chaos: Towards a Deterministic Approach to Turbulence* (Wiley, New York).

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

Berry, D. A., Herzel, H., Titze, I. R., and Story, B. H. (**1996**). "Bifurcations in excised larynx experiments," J. Voice **10**, 129–138.

Duffing, G. (**1918**). *Erzwungene Schwingungen bei Veränderlicher Eigenfrequenz und IHRE Technische Bedeutung* (Forced Oscillators with Variable Eigenfrequency and their Technical Meaning) (Vieweg, Braunschweig, 1918).

Eysholdt, U., Tigges, M., Wittenberg, T., and Proschel, U. (**1996**). "Direct evaluation of high-speed recordings of vocal fold vibrations," Folia Phoniatr (Basel) **48**, 163–170.

Fee, M. S., Shraiman, B., Pesaran, B., and Mitra, P. P. (**1998**). "The role of nonlinear dynamics of the syrinx in the vocalizations of a songbird," Nature (London) **395**, 67–71.

Fitch, T., Herzel, H., Neubauer, J., and Hauser, M. (**2002**). "Calls out of chaos: The adaptive significance of nonlinear phenomena in primate vocal production," Anim. Behav. **63**, 407–418.

Glass, L., and Mackey, M. C. (**1988**). *From Clocks to Chaos: The Rhythms of Life* (Princeton University Press, Princeton).

Hatzikirou, H., Fitch, W. T., and Herzel, H. (**2006**). "Voice instabilities due to source-tract interactions," Acta Acust. **92**, 468–475.

Herzel, H., Berry, D., Titze, I. R., and Saleh, M. (**1994**). "Analysis of vocal disorders with method from nonlinear dynamics," J. Speech Hear. Res. **37**, 1008–1019.

Hirano, M. (**1974**). "Morphological structure of the vocal cord as a vibrator and its variations," Folia Phoniatr (Basel) **26**, 89–94.

Hirano, M. and Kakita, Y. (**1985**). "Cover-body theory of vocal cord vibration," in *Speech Science*, edited by R. G. Daniloff (College Hill Press, San Diego), pp. 1–46.

Hirano, M., Vennard, W., and Ohala, J. (**1970**). "Regulation of register, pitch, and intensity of voice," Folia Phoniatr (Basel) **22**, 1–20.

Horáček, J., Švec, J. G., Veselý, J., and Vilkman, E. (**2000**). "Experimental study of the vocal fold vibration in excised larynx," in *Proceedings of Interaction and Feedbacks 2000*, edited by I. Zolotarev, Institute of Thermomechanics, Academy of Sciences, Prague, pp. 27–34.

Horáček, J., Švec, J. G., Veselý, J., and Vilkman, E. (**2004**). "Bifurcations in excised larynges caused by vocal fold elongation," in *Proceedings of the International Conference on Voice Physiology and Biomechanics*, edited by A. Giovanni, P. Dejonckere, and M. Ouaknine, Laboratory of Audio-Phonology, Marseille, pp. 87–89.

Horáček, J., Švec, J. G., Veselý, J., Vilkman, E., Klepáček, I., and Vetešník, A. (**2001**). "Measurement of the vocal-fold vibration behaviour in excised human larynges," in *Proceedings of the Second International Workshop on Models and Analysis of Vocal Emissions for Biomedical Applications*, Department of Electronics and Deaprtment of Physics, University of Firenze, Firenze, Italy.

Huygens, C. (**1673**). *Horologium Oscillatorium* (Apud F. Muguet, Parisiis, France) (English translation: *The Pendulum Clock*, Iowa State University Press, Ames, **1986**).

Ishizaka, K., and Flanagan, J. L. (**1972**). "Synthesis of voiced sounds from a two-mass model of the vocal cords," Bell Syst. Tech. J. **51**, 1233–1268.

Ishizaka, K., and Isshiki, N. (**1976**). "Computer simulation of pathological vocal-cord vibration," J. Acoust. Soc. Am. **60**, 1193–1198.

Jiang, J. J., Zhang, Y., and Ford, C. N. (**2003**). "Nonlinear dynamics of phonations in excised larynx experiments," J. Acoust. Soc. Am. **114**, 2198–2205.

Jiang, J. J., Zhang, Y., and Stern, J. (**2001**). "Modeling of chaotic vibrations in symmetric vocal folds," J. Acoust. Soc. Am. **110**, 2120–2128.

Lous, N. J., Hofmans, G. C., Veldhuis, R. N. J., and Hirschberg, A. (**1998**). "A symmetrical two mass vocal fold model coupled to vocal tract and trachea, with application to prothesis design," Acta Acust. **84**, 1135–1150.

Mende, W., Herzel, H., and Titze, I. R. (**1990**). "Bifurcations and chaos in newborn cries," Phys. Lett. A **145**, 418–424.

Mergell, P., Fitch, W. T., and Herzel, H. (**1999**). "Modeling the role of non-human vocal membranes on phonation," J. Acoust. Soc. Am. **105**, 2020–2028.

Mergell, P., Herzel, H., and Titze, I. R. (**2000**). "Irregular vocal fold vibration—High-speed observation and modeling," J. Acoust. Soc. Am. **108**, 2996–3002.

Miller, D. G., Švec, J. G., and Schutte, H. K. (**2002**). "Measurement of characteristic leap interval between chest and falsetto registers," J. Voice **16**, 8–10.

Neubauer, J., Edgerton, M., and Herzel, H. (**2004**). "Nonlinear phenomena in contemporary vocal music," J. Voice **18**, 1–12.

Neubauer, J., Mergell, P., Eysholdt, U., and Herzel, H. (**2001**). "Spatio-temporal analysis of irregular vocal fold oscillations, Biphonation due to desynchronization of spatial modes," J. Acoust. Soc. Am. **110**, 3179–3192.

Pabon, J., and Plomp, R. (**1988**). "Automatic phonetogram recording supplemented with acoustical voice-quality parameters," J. Speech Hear. Res. **31**, 710–722.

Pelorson, X., Hirschberg, A., van Hassel, R. R., Wijnands, A. P. J., and Auregan, Y. (**1994**). "Theoretical and experimental study of quasi-steady flow separation within the glottis during phonation," J. Acoust. Soc. Am. **96**, 3416–3431.

Schutte, H. K., and Seidner, W. (**1983**). "Recommendation by the union of European phoniatricians (UEP): Standardizing voice area measurement /phonetography," Folia Phoniatr (Basel) **35**, 286–288.

Sciamarella, D., and d'Alessandro, C. (**2004**). "On the acoustic sensitivity of a symmetrical two-mass model of the vocal folds to the variation of control parameters," Acta Acust. **90**, 746–761.

Smith, M. E., Berke, G. S., Gerrat, B. R., and Kreiman, J. (**1992**). "Laryngeal paralyses: Theoretical considerations and effects on laryngeal vibration," J. Speech Hear. Res. **35**, 545–554.

Steinecke, I., and Herzel, H. (**1995**). "Bifurcations in an asymmetric vocal fold model," J. Acoust. Soc. Am. **97**, 1571–1578.

Stevens, K. (**1977**). "Physics of laryngeal behavior and larynx mode," Phonetica **34**, 264–279.

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.

Švec, J. G., and Pešák, J. (**1994**). "Vocal breaks from the modal to falsetto

register," Folia Phoniatr (Basel) **46**, 97–103.
Švec, J. G., and Schutte, H. K. (**1996**). "Videokymography: High-speed line scanning of vocal fold vibration," J. Voice **10**, 201–205.
Švec, J. G., Schutte, H. K., and Miller, D. G. (**1999**). "On pitch jumps between chest and falsetto registers in voice: Data from living and excised human larynges," J. Acoust. Soc. Am. **106**, 1523–1531.
Švec, J. G., Šram, F., and Schutte, H. K. (**2007**). "Videokymography in voice disorders: What to look for?" Ann. Otol. Rhinol. Laryngol. **116**, 172–180.
Tigges, M., Mergell, P., Herzel, H., Wittenberg, T., and Eysholdt, U. (**1997**). "Observation and modelling glottal biphonation," Acta Acust. **83**, 707–714.
Titze, I. R. (**1973**). "The human vocal cords: A mathematical model. I.," Phonetica **28**, 129–170.
Titze, I. R. (**1988**). "The physics of small-amplitude oscillation of the vocal folds," J. Acoust. Soc. Am. **83**, 1536–1552.
Titze, I. R. (**2000**). *Principles of Voice Production*, 2nd ed. (National Center for Voice and Speech, Iowa City, IA).
Titze, I. R. (**2006**). *Myoelastic Arodynamic Theory of Phonation* (National Center for Voice and Speech, Iowa City, IA).
Titze, I. R., Baken, R. J., and Herzel, H. (**1993**). "Evidence of chaos in vocal fold vibration," in *Vocal Fold Physiology*, edited by I. R. Titze (Singular, San Diego), pp. 143–188.
Titze, I. R., and Strong, W. (**1975**). "Normal modes in vocal cord tissues," J. Acoust. Soc. Am. **57**, 736–744.
Titze, I. R., and Story, B. (**2002**). "Rules for controlling low-dimensional vocal fold models with muscle activation," J. Acoust. Soc. Am. **112**, 1064–1076.
Trendelenburg, W., and Wullstein, H. (**1935**). "Untersuchungen ueber die Stimmbandschwingungen (Research on vocal cord oscillations)," Sitzungsberd. Preuss. Akad. Wiss., Phys. Math. Kl. **XXi**, 399–426.
van den Berg, J. (**1957**). "Myoelastic-aerodynamic theory of voice production," J. Speech Hear. Res. **1**, 227–244.
van den Berg, J. (**1968**). "Sound productions in isolated human larynges," Ann. N.Y. Acad. Sci. **155**, 18–27.
van den Berg, J., and Tan, T. S. (**1959**). "Results of experiments with human larynxes," Pract. Otorhinolaryngol. (Basel) **21**, 425–450.
van den Berg, J., Vennard, W., Burger, D., and Shervanian, C. G. (**1960**). *Voice Production: The Vibrating Larynx Instruction Film* (University of Groningen, Groningen, The Netherlands).
Wilden, I., Herzel, H., Peters, G., and Tembrock, G. (**1998**). "Subharmonics, biphonation, and deterministic chaos in mammal vocalization," Bioacoustics **9**, 171–196.
Wittenberg, T., Tigges, M., Mergell, P., and Eysholdt, U. (**2000**). "Functional imaging of vocal fold vibration: Digital multi-slice high-speed kymography," J. Voice **14**, 422–442.
Zaccarelli, R., Elemans, C. P. H., Fitch, W. T., and Herzel, H. (**2006**). "Modeling bird songs: Voice onset, overtones and registers," Acta Acust. **92**, 741–748.
Zhang, Y., McGilligan, C., Vig, M., and Jiang, J. (**2004**). "Nonlinear dynamic analysis of voices before and after surgical excision of vocal polyps," J. Acoust. Soc. Am. **115**, 2270–2277.
Zhang, Z., Neubauer, J., and Berry, D. A. (**2006**). "The influence of subglottal acoustics on laboratory models of phonation," J. Acoust. Soc. Am. **120**, 1558–1569.

Comparison of speech production in upright and supine position

M. Stone,[a] G. Stock, K. Bunin, K. Kumar, and M. Epstein
Department of Biomedical Sciences, University of Maryland, Baltimore, Maryland 21201

C. Kambhamettu and M. Li
Department of Computer Science, University of Delaware

V. Parthasarathy and J. Prince
Department of Electrical Engineering, Johns Hopkins University

(Received 7 July 2006; revised 16 February 2007; accepted 20 February 2007)

Speech is usually produced in an upright sitting or standing posture. Measurements and judgments of speech may be made in conditions requiring a supine position, however. These conditions include MRI recordings, and oral procedures, such as, adjustments to dental appliances, medical and surgical procedures. It is of interest, therefore, to see whether gravity has strong or systematic effects on tongue behavior. In the present study, 13 subjects repeated several words, which contained extreme consonant and vowel tongue positions, during upright and supine condition. Ultrasound imaging provided midsagittal tongue contours, in each condition, for comparison. A neck brace was used to stabilize transducer placement and the palate was used as a physiological reference to register the data sets. Results showed a significant subject effect. In supine position the tongue was more posterior than upright for seven subjects, more anterior for two subjects and varied by phoneme for four subjects. However, there was no significant phoneme effect. The direction of change and the amount of change were not directly related. Most subjects had small upright-supine differences. The largest differences, less than 3 mm on average, were in the posterior tongue.
 [DOI: 10.1121/1.2715659]

PACS number(s): 43.70.Aj, 43.70.Jt [BHS] Pages: 532–541

I. INTRODUCTION

Gravitational force is one of the major external loads experienced during speech (Shiller *et al.*, 1999), but it is not well known how the tongue compensates for this load or how this compensation differs in upright versus supine position. Under ordinary circumstances speech is produced in an upright position. However, in situations like sleeping, dental procedures, and MRI recordings, the subject is supine. In sleep, tongue position is of interest due to its relationship with sleep apnea. Rose *et al.* (2002) concluded that alteration of head posture and tongue position alone had a significant effect on the severity of obstructive sleep apnea. In dentistry, body and head position are also a concern. The measurements and placement of an upper removable appliance are often done in supine position, and errors can affect tongue position and disturb tongue function (Reinicke *et al.*, 1998). In MRI, vocal tract recordings are made in supine condition but are used in "upright" models of speech production. The present study is particularly focused on the effects of gravity on the tongue in studies using MRI.

Several methods have been used to study the effects of gravitational orientation on tongue behavior. EMG measurements of muscle activity showed greater genioglossus posterior (GGP) activity for some vowels than others, and for inspiration than expiration. A comparison of these behaviors in upright and supine position showed more GGP activity in supine position (Myamoto *et al.*, 1997; Niimi *et al.*, 1994; Otsuka *et al.*, 2000; Sauerland and Mitchell, 1975). It is likely that increased GGP activation results in a variety of tongue modifications, including maintenance of the upright pharyngeal position, a more anterior tongue position, or a more posterior position, as long as the airway is sufficiently open. These tongue modifications may be subject and/or task specific. Cephalograms have confirmed a variety of responses to gravity. Patients with sleep apnea maintained their upright tongue position when moved to supine condition, whereas nonapneic snorers had significant superior-posterior tongue movement in supine position (Myamoto *et al.*, 1997).

The introduction of MRI to the study of speech affords great advancement in the ability to visualize and quantify the behavior of the tongue (cf. Narayanan *et al.*, 1995; Stone *et al.*, 2001). Unfortunately, the supine data collection of MRI does not perfectly reflect upright speech and may require transformation to simulate upright positions if these data are to be compared to upright data sets. A better understanding of gravitational effects will allow better utilization of MRI data in speech analysis and modeling. Several MRI studies have examined sustained phonemes. Badin *et al.* (2002) found that sustained consonants and vowels had backward displacement of the tongue in MRI images when compared to cineradiographs. Engwall (2006) further found that gravity affected the posterior tongue to a greater extent than the an-

[a]Electronic mail: mstone@umaryland.edu

terior tongue. Kitamura *et al.* (2005) used Open MRI to compare upright and supine tongue positions for three subjects saying five sustained vowels. The results showed subject differences in the effects of gravity orientation. One subject had an almost identical position of the tongue body and posterior pharyngeal wall in both conditions. The other two subjects showed posterior displacement of both the tongue and the posterior pharyngeal wall. The authors proposed differences in head, rather than tongue position, as the major cause. However, the observed tongue retraction was smaller for front vowels, and these two subjects reduced jaw opening in supine position. The authors proposed stabilization of the anterior tongue against the hard palate for front vowels to account for this difference.

Engwall (2006) used MRI to compare supine and prone speech tasks. This work used three-dimensional reconstructions of supine and prone tongue surfaces for vowels and fricatives sustained for 5 seconds, and midsagittal contours of VCVs collected at nine frames per second. Overall, differences seen in supine versus prone condition were small and varied with phoneme. The effects of gravity were larger for sustained phonemes than speech articulations. In supine condition, the tongue was more posterior especially for front vowels reducing the tongue shape differences between /i/, and the back vowels /ɑ/, /u/.

Tiede *et al.* (2000) used x-ray microbeam measures of the tongue and jaw, to study two subjects saying vowels, /bV/ syllables, and phrases, in sitting and supine position. Results for the sustained vowels reflected subject differences. One subject had a slightly posterior jaw, and a slightly upward and anterior tongue when supine. The second subject had virtually no mandibular displacement, and a slightly posterior tongue when supine. During movements, however, the two subjects behaved similarly across gravity conditions. For jaw movements, both subjects showed reduced range of motion, and reduced opening/closing gestures, when supine. For tongue movements, both had minimal variability in the region of the constriction when comparing supine to upright. Differences in tongue position were somewhat dependent on the type of task. As with Engwall (2006) tongue position in static and /bV/ vowels was more variable and more affected by gravity than those in phrases. Although the jaw and tongue positions were variable between upright and supine posture, the output acoustics were relatively similar regardless of body orientation. The authors concluded that in supine position articulatory adaptation was idiosyncratic and designed to maintain acoustics, rather than preserve articulatory trajectories.

Shiller *et al.* (1999) modeled the effects of gravity on jaw displacement based on the hypothesis that no compensations to gravity are made. They predicted that in supine position the jaw would shift backward horizontally and rotate away from occlusion. These predictions were compared with empirical data from the nonhigh front vowels (/æ/ and /ɛ/). Five subjects produced CVCs with various consonants in upright, supine and prone positions to compare with the simulations. Measurements were made of jaw motion (Optotrak - Northern Digital, Canada) and vowel formant frequencies. Formant analysis indicated a significant upright-supine difference for /æ/ but not /ɛ/. Jaw measurements, averaged across all five subjects, indicated a significant difference in jaw rotation, but not horizontal jaw translation for five of the six syllables in supine condition. The lack of backward jaw translation indicated that differences in supine jaw motion are not due to gravity alone. The authors concluded that subjects do not completely compensate for gravitational load.

Taken together these studies indicate that the effects of supine position on tongue behavior are both local and global, are lesser for time-varying tokens than sustained phonemes, and vary among subjects. The studies primarily examined vowels. Lingual consonants are also of interest and may exhibit a smaller gravitational effect at their vocal tract constriction, than the vowels due to their tighter tongue palate contact (similar to Badin *et al.*, 2002).

The present study aims to replicate a cine-MRI session and determine potential effects in midline tongue contours due to gravity. In static MRI studies, where a single sound is sustained, the weak RF emissions from hydrogen protons are summed over time as the subject holds still. Cine-MRI requires multiple repetitions of a single utterance while the emissions of each time frame are summed in an ensemble fashion. The technique was developed and refined in a smaller study that examined the effect of gravitational orientation on tongue position during words and pauses for a small number of subjects (Stone *et al.*, 2002b). That study found backward rotation was the most salient feature in upright/supine comparisons for two subjects, and that the center of this rotation differed by phoneme, with /i/ being most anterior. There was no acoustic effect of the backward tongue rotation.

In the present study, tongue shapes are expected to differ locally and globally due to gravitational orientation and also to differ among subjects. To determine whether these expectations are true, ultrasound images presenting a variety of lingual tongue contours during consonants and vowels were taken from words spoken in upright and supine condition. Ultrasound allows tongue data to be collected in both supine and upright conditions, and provides good imaging of the pharyngeal region, which is of major interest in this study. In addition, midsagittal ultrasound contours of the tongue parallel those observed in midsagittal MRI, despite the loss of tissue visibility in the extreme anterior and posterior tongue (Stone *et al.*, 2003). Finally, the data were collected from continuous speech, which is more realistic than sustained utterances, and seems from previous work to have a lesser response to gravity. Ultrasound images are thus ideal for exploring the effects of gravity on tongue surface shape.

The goals of the present study were to measure the effects of gravity on tongue position in speech behaviors, and to determine whether task and subject effects interact with gravity orientation. Two responses to gravity are possible. First, the tongue might be positioned more posteriorly in supine orientation with little or no compensation to gravity, as modeled by Shiller *et al.* (1999). This would occur if the posterior displacement of the tongue did not significantly affect speech acoustics or airway patency. Second, the tongue might be positioned more anteriorly (this includes

TABLE I. Subject and task listing. Thirteen subjects said the words "bang," "dash," and "golly," except as noted in the table.

			Bang	Dash		Golly
Subject	Age	Sex	ŋ	dæ	ʃ	gɑli
1	16	f	×	×	×	×
2[a]	26	f	He rang	Dack	Highchair	×
3	25	f	He rang	Dack	Highchair	×
4	24	m	×	×	×	×
5	19	m	×	×	×	×
6	27	f	×	×	×	×
7	48	m		×	×	×
8	22	f	×			×
9[b]	24	m	×	×	×	×
10	29	f	×	×	×	×
11	16	m	Anger		Highchair	×
12	24	m	×	×	×	×
13	16	m	×	×	×	×

[a]Native speaker of Brazilian Portuguese.
[b]Native speaker of Russian.

maintaining the original posture) to protect the airway (Tiede *et al.*, 2000; Kitamura *et al.*, 2005). Strategies may vary with subject and phoneme.

II. METHOD

A. Subjects

Thirteen subjects participated in the study. They included seven males and six females, with ages ranging from 16 to 48 years old (mean age 24). All but two were native speakers of U.S. English; those two were native speakers of Portuguese and Russian and were fluent in English (see Table I).

B. Speech materials

The target speech tasks were designed to include a wide range of oral positions and to mimic tasks used in an MRI data session. Therefore the tasks were short words that could be repeated at one-second intervals and included the consonants /ŋ/, /d/, /ʃ/, /g/, /l/, and the vowels /i/, /ɑ/, /æ/. The words were "bang," "golly," and "dash." Due to slightly different protocols, the /æ/ and /ʃ/ were taken from "dack" (/dæk/ and "high chair" for some subjects. The /æ/ was always taken from the nonnasal context (dack or dash). Each word was repeated seven times. To eliminate possible "list-effects" on the first and last repetitions, the five medial repetitions were measured. Not all subjects said the same words and Table I lists the words for each subject. Of the 13 subjects, three words were missing due to errors in data collection. Thus the corpus consisted of 39 words, 37 vowels, and 61 consonants.

C. Data collection

An ultrasound machine (Acoustic Imaging, Inc., Phoenix, AZ, Model No. A15200S) was used to collect midsagittal images of the tongue from each subject during the speech tasks in two gravity conditions: upright (UP) and supine (SUP). A 2.0–4.0 MHz multifrequency convex-curved linear array transducer that produced 30 wedge-shaped scans per second was used. In order to create similar upright and supine recording conditions, the ultrasound transducer was fitted to a cervical collar in a midsagittal orientation. The collar was positioned around the subject's neck, in the upright position, so that the transducer was under the chin, and then adjusted until the shadows of the hyoid and jaw were as close as possible to being equidistant from the edges of the scan. This is meant to normalize transducer position across subjects. The transducer was immobile for both supine and upright data collection (Fig. 1, col. 1). This inhibited jaw motion severely, but allowed consistent transducer positioning across conditions. A microphone, attached to the dental chair, recorded speech.

The subjects were instructed to repeat the words seven times in each gravity condition to the beat of a metronome. The metronome was set at one beat per second to mimic an MRI data collection, where repetition time is one second. The entire recording procedure lasted 20–30 minutes. Both the ultrasound image sequences and the synchronized acoustic signal were digitized using the ADVC1394 (Canopus Inc.) I/O card. The ultrasound image sequences were saved as a series of jpeg files at 29.97 frames per second. The acoustic data were saved as a monaural wave file at a sampling rate of 44.1 kHz. All data were backed-up by simultaneous recording on videotape.

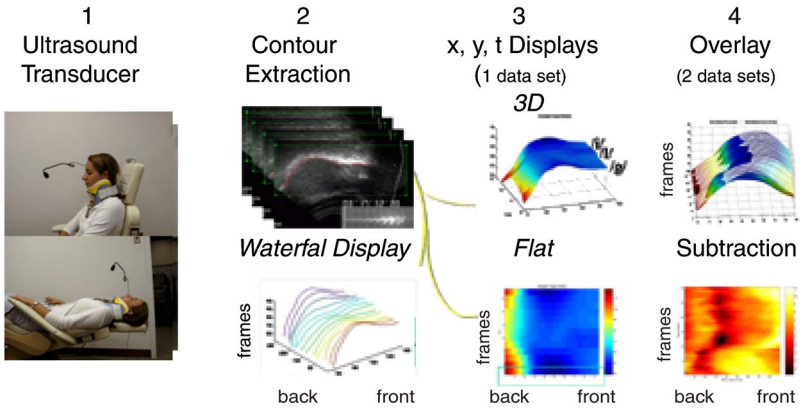


FIG. 1. (Color online) Data collection and analysis. (1) Transducer in neck brace; (2) tongue contour extraction and time-motion (waterfall) display; (3) time-motion displays as 3D surface and flat sequence; and (4) overlay of two contour sequences and their difference.

D. Palate alignment

The first four subjects were collected in upright position first. When put into supine position it was possible that some had moved their heads backwards. All subsequent subjects were measured in supine position first and instructed to keep the head immobile when moving into upright position. When supine was collected first the neck brace appeared to adequately inhibit head movement in the speakers. Validation was done on the supine-first method by measuring distances and angles between points on the head and neck brace on a single subject. These measurements indicated less than 3 mm error between the gravity conditions. However, it was believed that some subjects may have moved their head in the superior-inferior direction. Therefore, a more valid method was developed to align palates, described in detail in Epstein and Stone (2005). In brief, the supine palatal contour was overlaid and aligned to the upright contour using rigid body transformation (x-, y-translation and in-plane rotation). The palate transformation parameters for each subject were saved and used to transform the supine tongue contours into the upright orientation.

Palate shape is not normally visible in ultrasound. However, during a swallow the palate is visible because the sound reaches the palatine bone when the tongue touches the palate. When the ultrasound wave reaches the palatine bone it reflects back to the transducer. The reflected palate appears as a white line, which can be traced like the tongue. Palatal contours for each subject were extracted from spontaneous dry swallows in upright and supine condition, and the pair of palate contours was rigidly aligned. In addition, by measuring the palate during several frames of the swallow, the velum was captured in multiple positions as it moved upward into a maximally closed position. Overlaying the velar contours reveals the junction of the velum and the hard palate. The palatovelar junction is a tissue point that was used to normalize the location of tongue constrictions across subjects (for details see Epstein and Stone, 2005).

The quality of the UP-SUP palate alignment was determined for each subject by two calculations: maximum absolute error and root-mean-squared (RMS) error. Error measurements were made by comparing the nearest-neighbor points in the x, y directions. The RMS errors provide a global estimate of error, and the maximal absolute error indicates how the palatal differences compare to the ultrasound measurement error (max=0.7 mm, Unser and Stone, 1992). RMS errors were used rather than average distances, because they give increased weight to larger distances making differences more noticeable. After alignment seven subjects had maximum errors below measurement error (0.7 mm), for five others it was less than 1.3 mm. For the RMS errors, 10 subjects were below measurement error, two were at 0.8 mm. The remaining subject had a maximum error of 2.9 mm and an RMS error of 1.5 mm. This final subject (No. 11) had two entirely different palate shapes; one may have been the tongue. Since one swallow was collected per subject, a better tracing could not be obtained. However, because parts of the palatal contours were well aligned visually, his data were included. His results were among the larger differences, but not the most extreme.

E. Tongue data preprocessing and analysis

Frame numbers (i.e., temporal location) for each word, phoneme, and repetition were extracted from the speech wave. For the consonants and vowels, tongue surface contours were measured and analyzed in several steps using the Maryland Tongue Analysis Package (MTAP) (www.speech.umaryland.edu). First, tongue contours were extracted from the ultrasound images using EdgeTrak, which semiautomatically extracts and tracks tongue contours (and palate contours) for each subject and word (Li *et al.*, 2005a) (Fig. 1, col. 2). Second, the supine and upright palate contours were registered by rigidly transforming the supine palates to the upright orientation. These transformation parameters were then applied to all the supine tongue contours. All subsequent analyses were done using palatally aligned supine data. Third, before averaging the five repetitions of each word, temporal alignment was done using features in the ultrasound contours.

Spatial alignment (rigid body registration) also was done to reduce random variation between replicates introduced by subject imprecision, without modifying tongue shape (Li *et al.*, 2005b). It was assumed that inter-repetition variability was due to unintended speaker error (noise), and that the subject intended to produce the same token in each repetition. Inter-repetition variability varies by speaker and tongue location. The current wisdom is that variability is minimal at constriction locations and (Perkell and Nelson, 1982), because constrictions are acoustically important; acoustically unimportant regions have greater variability. Previous ultrasound studies have found average-RMS differences for replicate curves (data set: /a, i, u, e, o/) to be 1.36 mm or less (Morrish *et al.*, 1985). However, differences at the location of maximum variability (data set: /a, i, u/) were 1.4 cm or less (Stone *et al.*, 1983). Large maximum and small average variability are consistent with fairly small overall variability in multiple replicates. Our unpublished examinations of list-type replications support this as well. Rigid body alignment, based on least-squares minimization, would optimally reduce larger variations while minimally affecting smaller ones.

The fourth preprocessing step averaged the five spatiotemporally aligned repetitions for each gravity condition. The average dataset was considered to be a better exemplar of the word than any individual token. Moreover, MRI data are typically summed from many repetitions, and thus contain smoothing similar to the ultrasound average. All subsequent analyses used the average contours except range of motion.

In order to visualize and statistically analyze the data sets, they had to be of the same size. Therefore, for each subject, upright-supine word pairs were cut or extended to the same length, with the software Surfaces, which employs thin-plate splines (kriging) to extend a contour (Parthasarathy *et al.*, 2005). The resulting contour sequence for each word is represented as a 3D object, that is, a spatiotemporal

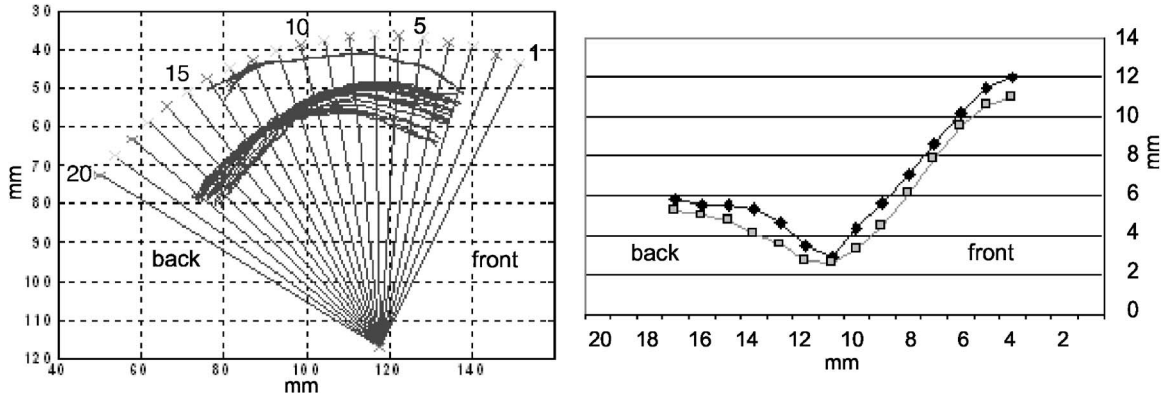


FIG. 2. (a) Tongue contours for one repetition of "dash" (upright), and the subject's palate trace (top), are intersected by up to twenty radii. The palate trace shows multiple velar positions. (b) Motion functions show the ROM at each radius for five repetitions dash in upright (black diamond) and supine (gray square) condition. Radius 1 is anterior; measured values are in millimeters.

"surface" consisting of *x*, *y*, and *t* values (see Fig. 1, col. 3). Height values are color-coded. From the upright and supine surfaces, the target phoneme contours were extracted and overlaid. UP-SUP contour differences were visualized by subtraction (Fig. 1, col. 4) and compared globally using RMS error. The tongue was then divided into six equal length regions and the average UP-SUP distance calculated for each. In ultrasound images the tongue tip and root are imaged less clearly, causing measurement error. This error is compounded by the use of kriging (spline extension) to equalize contour lengths for the comparisons. As the ends (regions 1 and 6) may have visibility and extrapolation errors, regions 2 and 5 were used as the extreme posterior and anterior tongue regions respectively. As transducer angle cannot be perfectly controlled in the neck brace, some subjects' data were collected at a more posterior angle than others. To compare constriction location across subjects the location of the velar junction, a tissue point on the hard palate, was used as a reference point across subjects.

For each subject, range of motion (ROM) was calculated in both upright and supine condition for five repetitions of each word. Differences in ROM between the two gravity conditions are of interest as they reflect tongue mobility effects. The need to maintain an open airway in supine position could reduce ROM. It is also possible that gravitational pull would increase backward movement while forward positioning was maintained for acoustic reasons, resulting in larger supine ROM. ROM was examined at multiple locations on the tongue surface by creating a polar grid template for each subject. Each template was based on the subject's entire dataset, that is, all the tongue contours for 3 (or 4) words × 5 repetitions × 2 gravity conditions. The tongue contours were overlaid and a vertex placed several mm below and midway between the two most extreme endpoints. Originating from this vertex, 20 equally spaced radii were positioned to intersect with the tongue surface; the first and last radii were aligned with the most extreme endpoints in the data. Figure 2(a) shows tongue contours for one repetition of dash drawn on such a template. To calculate ROM at each radius, the minimally displaced contour was subtracted from the maximally displaced contour (max-min difference). The max-min differences were averaged for each word, and the averages used to compare the effects of gravity, subject and word on the extent of tongue surface motion.

After all UP-SUP comparisons were made, the supine contours were "corrected " to match their upright counterparts using rigid body and rigid-plus-affine (hereafter, affine) alignment methods. The goal was to determine whether global transformations were sufficient to correct the effects of gravity on the tongue, and whether additional affine correction (homogeneous scale, stretch and shear) improved the match.

Acoustic analysis was performed on the upright and supine speech waves for nine subjects, and five repetitions, of the words dash (or dack) and golly using PRAAT. The first two formants were extracted from the midpoints of each vowel by visual inspection of the spectrogram, with LPC tracking for assistance. In addition, formants were extracted at vowel onsets or offsets to capture consonant information for /ŋ/, /d/, /g/, and at the onset and midpoint of /l/. For the acoustic analysis five repetitions of the sustained vowels /u/, /a/, and /i/ were also available. The formants were extracted at the midpoint of the vowel for each repetition. These sounds did not have visually observable formant variability and were used to determine whether sustained sounds were different from continuous speech. Because the ultrasound machine is fairly noisy, the acoustic wave was as well, and neither formant bandwidths nor fricative noise were measured.

F. Comparative statistics

Acoustic data were available for 198 upright-supine comparisons (nine subjects × 11 phonemes × 2 formants). Thirty of these comparisons had missing data, leaving 168 comparisons. Confidence intervals (at 95% confidence) were calculated for F1 and F2 in each gravity condition, and significant differences determined based on overlap.

The ultrasound contours were examined in three ways. (1) Paired t-tests compared the ROM of each word in the upright and supine condition. (2) Two one-way ANOVA's for repeated measures examined the UP-SUP differences in global contour (RMS error), and the pharyngeal zone, due to subject and phoneme. If a significant difference was found, a

Tukey HSD test was used to extract the significant comparisons. (3) Paired t-tests were used to compare the upright contours to the corrected supine contours on a point-to-point basis.

III. RESULTS

This study examined several global and local features of midsagittal tongue contours that might be affected by gravity. (1) Vowel formants were used to compare acoustic consequences of gravity on target consonants and vowels. (2) Range of motion was calculated to determine whether supine motions were more limited than upright. (3) Global contour differences were compared for each phoneme. (4) Pharyngeal contour differences were compared for each phoneme. (5) The effect of gravity on the location of the consonant constriction was measured. (6) The supine contours were corrected to upright using rigid and affine transformations. These global and local effects were considered with respect to subjects and tasks.

A. Acoustic effects due to gravity

The confidence intervals of the 168 comparisons showed 13 significant differences between upright and supine at the $p=0.05$ level. This number is consistent with chance, and indicates that the physiological effects of gravity were negligible acoustically. Interestingly, 9 of the 13 significant differences occurred in sustained vowels.

B. Range of motion in upright and supine condition

ROM was calculated for each of the 39 words in the upright and supine condition to examine tongue motion. The other measurements were applied to single frames extracted from the image-sequence to represent the maximum position of each phoneme.

The words differed in ROM based on phonemic content. Golly had the largest average ROM in both gravity conditions (mean=13.3 mm-UP, 13.6 mm-SUP) due to the large difference in tongue position for the high back /g/, low back /ɑ/ and high front /i/. A smaller ROM was seen for dash (8.3 mm, 8.5 mm), which had no back phonemes. The smallest was seen for bang (6.5 mm, 6.1 mm), with only two lingual phonemes. Figure 2(a) displays a set of overlaid tongue contours for one repetition of upright dash, and a palate contour showing several velar positions. The lowest tip position (radius 4) occurs in /ae/, the highest one in /d/, with the /ʃ/ in between. ROM is maximal anteriorly and minimal at a pivot region toward the back. The ROM calculated at each radius can be plotted with respect to the radius point (index) to generate a "motion function." Motion functions for the word dash are shown in Fig. 2(b) for the upright tongue contours presented previously in Fig. 2(a) (black diamonds) and for the corresponding supine contours (white squares). Motion functions were generated for each word in the UP-SUP conditions and indicated, across gravity conditions, a pattern similar to that in Fig. 2(b).

The motion functions shown in Fig. 2(b) reflect the large anterior ROM and a rotational movement during the word. At radius 11 is the motion minimum, which depicts the fulcrum of the tongue rotation. The difference in average (and SD) ROM for all the radii is very small [upright 9.64 mm (3.1), supine 9.47 mm (3.3)]. For each word, the average ROM values were compared using a matched pairs t-test and the differences were non-significant ($p=0.05$). Of the 39 subject/word comparisons, 27 had average ROMs that were larger in upright condition; 21 of these differences were less than 1 mm. The remaining 12 words had ROMs that were larger in supine condition; 8 of these were less than 1 mm.

C. Upright-supine differences in tongue contour

Subject and phoneme differences were compared globally by calculating RMS difference between each pair of UP-SUP contours. The distances ranged from 0.7 to 6.9 mm (Table II). A one-way ANOVA for repeated measures determined that subject had a significant effect on the RMS difference ($F=10.724, p\leqslant 0.0001$), but phoneme did not ($F=0.488, p\leqslant 0.841$). A Tukey HSD test was used to determine which subjects grouped together. In Table II, the bars indicate groups that are not significantly different. The single horizontal bar at the bottom indicates no phoneme differences. The vertical bars classify the majority of subjects into a single group with small UP-SUP RMS differences (mean $\leqslant$2.5 mm), and other groups with larger UP-SUP differences. Two subjects (1,3) bad significantly greater UP-SUP differences than almost all the other subjects. These two subjects were among the four whose data were collected upright first. Repeat measurements and visual inspection of the raw images indicated a backwardly displaced tongue relative to the palate in supine condition. To artifactually create such a difference in tongue position relative to the palate reference, the transducer or head would need to rotate sufficiently that the angle of recording was vastly different. A largely different vertex location and radii angle could create such an error. Such a movement was not observed during data collection, however, it cannot be ruled out.

Subject and phoneme differences also were compared in the pharyngeal zone since gravity might have its greatest effect on this region. The pharyngeal zones varied in location by subject, and were determined by the velar junction to be 2, 3, or 4. The pharyngeal differences ranged from 0.2 to 8.9 mm. About a quarter of the differences (27 of 97) were less than 1 mm (Table III). A one-way ANOVA for repeated measures determined that subject had a significant effect on the pharyngeal contour difference ($F=12.728, p\leqslant 0.0001$), and phoneme did not ($F=0.600, p\leqslant 0.755$). The result of the Tukey HSD test is indicated by the black bars on the side and bottom of Table III. The single bar at the bottom indicates that phoneme differences were not significant. The bars on the right indicate that ten subjects were not significantly different from each other; they had from 1 to 3 mm differences in pharyngeal tongue position. Again subjects 1 and 3 had the largest posterior differences, over 6 mm.

Although the predominant response to the supine gravity condition was to move the pharyngeal tongue posteriorly, subject specific patterns showed that this was not universal, and that the size of the response did not reflect the direction of motion. The largest group had small, nonsignificant, pha-

TABLE II. Upright-supine tongue contour differences calculated using RMS error, in ascending order of subject mean. Means that are not within the same bar are significantly different. Values in millimeters.

Subject	ng	d	sh	g	l	a	ae	i	Mean (SD)
2	1.5	0.7	1.6	1.8	1.8	2	0.7	1.8	1.5 (0.5)
8	0.9			2.6	1.4	1.7		1.6	1.6 (0.6)
10	0.5	1.2	0.9	2.8	2.3	2.2	1.3	3.6	1.9 (1.1)
7		1.8	2.7	1.7	0.9	2.3	1.9	2	1.9 (0.6)
6	0.4	2.5	1.5	0.7	1.9	2.4	3.5	2.7	2.0 (1.0)
13	1	3.1	0.8	1.9	2.1	2.9	1.4	4.1	2.2 (1.1)
9	2.9	1.2	3.9	1.4	1.8	1.2	2.2	2.4	2.1 (0.9)
12	3.7	1.5	2.3	2.9	1.4	3.3	3.2	1.6	2.5 (0.9)
5	1	1.6	2.2	1.3	2.3	3.2	2.4	6.2	2.5 (1.6)
4	6.8	5.7	5.1	1.2	2.1	2.4	4.1	4	3.9 (1.9)
11	5.9		3.1	4.1	4.6	3.6		4.5	4.3 (1.0)
1	4.2	6	5.4	5.3	6.3	4.2	4.5	4.4	5.0 (0.8)
3	6.9	5	4.7	4.7	4.3	5.3	3.8	6.5	5.2 (1.1)
Mean	2.98	2.75	2.85	2.49	2.55	2.8	2.78	3.1	
SD	2.48	1.93	1.6	1.43	1.55	1.1	1.1	1.5	

ryngeal differences (subjects 2, 5, 6, 7, 8, 9, 10, 11, 12, and 13). Five of these moved backward (2, 5, 7, 9, and 11), one moved forward (7), and four varied by task (8, 10, 12, and 13). A second group (1 and 3) moved backward considerably in supine position (mean=6 mm). A single subject (4), had a moderate mean difference (4 mm), and moved forward in supine condition. Figure 3 displays typical contours in upright (black) and supine (gray) conditions. Backward and forward motion included translation (A,C) and rotation (B). Differences were largest posteriorly.

D. The effect of gravity on the consonant constriction

It is well documented that the effects of continuous speech, such as coarticulation, rate, multiple repetitions, have

TABLE III. Upright-supine contour differences in pharyngeal region of tongue, in ascending order of subject mean. Means that are not within the same bar are significantly different. Values in millimeters.

Subject	ng	d	sh	g	l	a	ae	i	Mean (SD)
6	0.2	1.6	1.6	0.4	0.2	0.8	0.7	2.4	0.9 (0.80)
8	0.9			1.1	1.3		0.4	2.6	1.3 (0.82)
7		2	2.3	0.7	0.2	1.9	1	0.8	1.3 (0.79)
2	0.4	0.4	1.6	1.4		0.6	2.5	2.5	1.3 (0.92)
13	0.2	1.4	0.3	1.5	2.9	1.7	3	0.7	1.5 (1.07)
10	0.6	0.8	0.8	2.3	2.4	1.1	2.2	1.8	1.5 (0.75)
9	0.7	0.7	6.9	0.2	1.5	1.3	0.3	3	1.8 (2.24)
12	3.3	0.7	1	1.1	2.2	3.7	0.8	2.8	1.9 (1.21)
5	0.4	0.6	3	3	2.6	3.1	3.4	3.7	2.5 (1.26)
11	1.9		1.7	3.1	4		2.7	5.7	2.8 (1.72)
4	3.7	7.4	6.7	1.5	2	5.6	1.7	5.8	4.3 (2.38)
1	3.9	8.9	7.7	5.4	7.8	6.4	5	5.9	6.4 (1.66)
3	9.3	5.2	5	6.4	6.1	3.8	6.5	8.8	6.4 (1.87)
Mean	2.13	2.7	3.22	2.16	2.77	2.73	2.32	3.58	2.61
SD	2.66	3.02	2.64	1.89	2.26	1.97	1.86	2.34	1.89

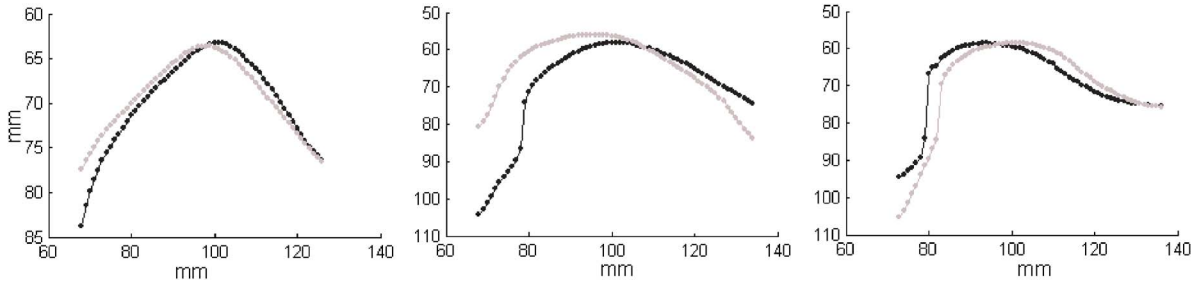


FIG. 3. (Color online) A comparison of upright (black) and supine (gray) contours for three subjects saying /g/ demonstrates backward translation, backward rotation and forward translation. (A) Subject 5 /æ/; (B) Subject 1 /ʃ/; and (C) Subject 8 /g/.

their smallest effect on vocal tract area at the location of the consonant constriction (cf. Perkell *et al.*, 1992). In the present study it was hypothesized that the UP-SUP difference would be minimal at the location of the consonant constriction to preserve this acoustically important feature. The velar junction and the minimal UP-SUP difference for each subject and phoneme were studied to determine whether the minimum difference was at the constriction location. Since the tongue tip is not visible in ultrasound images, the anterior constriction locations may not be well depicted.

Figure 4 depicts the number of times that the minimal upright/supine distance occurred at a particular zone for each consonant. The sounds are ordered from top-to-bottom in the order /l, d, ʃ, s, ŋ/. Some phonemes had uniform UP-SUP differences throughout the tongue, and often these differences were quite small. Distances of less than or equal to 1.0 mm throughout the tongue (zones 2–5) were considered to have no-change and were not plotted. The number of times each phoneme had no-change (out of 13 per phoneme) was 7 for /ŋ/, 6 for /d/, 6 for /ʃ/, 4 for /g/ and 2 for /l/. Thus, for some phonemes, such as /ŋ/, there were very few token differences to consider. For the remaining 36 comparisons, consonants with more anterior constriction locations were found to have minima at more anterior segments. Although the data were generally consistent with preserving size of the constriction, the pharyngeal zone differed across subjects and there were not enough data to do a statistical analysis, or provide unambiguous results.

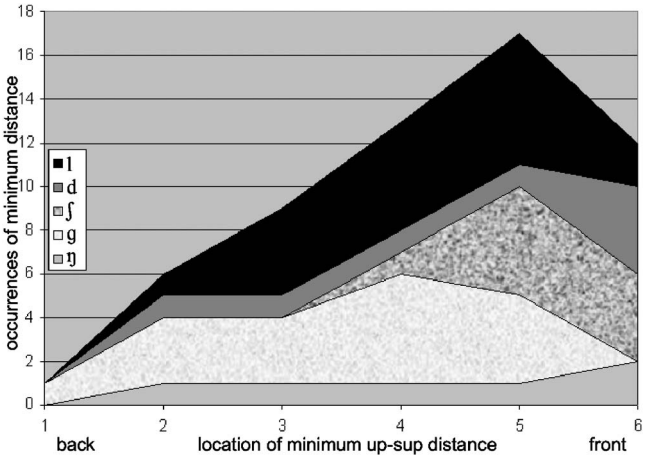


FIG. 4. Number of occurrences of minimal up-sup distance at each zone for each phoneme. Phonemes with anterior constrictions have more frequent occurrences of anterior minima. Order of phonemes from top to bottom is: /l, d, ʃ, s, and ŋ/.

E. Correcting the supine data for the effects of gravity

Once the UP-SUP differences were documented for each subject and phoneme, the next step was to correct the supine contours into upright ones to see how complex a transformation was needed. Correction of the supine contours into the upright ones was performed using two methods: rigid alignment (i.e., rotation and translation), and affine alignment (i.e., rigid plus homogeneous stretch, shear and scale). A one-way ANOVA found that there was no subject effect on the RMS distances for either the rigid or affine correction. That is, regardless of whether the correction was affine or rigid, the corrected supine contours were very similar to the original upright contours for all subjects. In only one subject did the RMS error show contour differences of more than 0.4 mm (see Table IV). A matched-pair t-test compared the effects of phoneme, within each subject, on transformation method. Although the affine transformation significantly improved the rigid alignment ($p=0.05$) for 5 subjects (3, 4, 6, 10, and 12), these improvements were on the order of 0.2 mm or less, and are probably not functionally meaningful.

TABLE IV. RMS error for rigid and affine transformation methods of aligning upright and supine contours (mm).

Subject	Rigid	Affine
1	0.8	0.7
2	0.5	0.5
3	0.6	0.4
4	0.7	0.5
5	0.4	0.4
6	0.6	0.4
7	0.5	0.6
8	0.7	0.5
9	0.7	0.7
10	0.4	0.3
11	0.8	0.7
12	1.1	0.7
13	0.7	0.8
Mean	0.66	0.54
SD	0.18	0.16

IV. DISCUSSION

Three factors appeared to contribute to the supine tongue displacements: preservation of speech quality, airway protection, and subject preference.

The first finding was that acoustic spectra (formant frequencies) were preserved despite the varying tongue responses to gravity. The number of significant differences was due to chance and occurred mostly on steady state sounds. This is consistent with the findings of others, who found tongue position was more sensitive to gravity in steady sounds than continuous speech (Tiede *et al.*, 2000; Engwall, 2006).

Airway protection appeared to be a consideration in tongue displacement as well. Only two subjects had large backward displacements in supine condition. The small backward translation in the other 11 is consistent with airway protection being a factor in tongue positioning. The fairly large number of subjects reveals this as a consistent behavior despite between-subject differences.

The presented data were midsagittal. It was not possible to position the transducer in the neck brace to collect comparable coronal or parasagittal slices across subjects. Our belief is that parasagittal tongue differences would be smaller than midsagittal. Lateral tissue is attached to bony structures or other muscles (e.g., pharyngeal constrictors) whose mass would resist anterior-posterior motion of the lateral tongue. In general we have found that tongue motions are smaller in the para- than the midsagittal planes during speech. In the present data, the consistent formants across gravity conditions is interesting, and suggests that either these fairly small motions are not occurring in acoustically sensitive regions, or that the front back trade-offs create cavity sizes that produce comparable formant patterns, or that some parasagittal changes are indeed compensating for the gravitational changes.

The supine tongue positions indicated the use of all three possible compensatory strategies: backward displacement (minimal or no compensation), maintenance of upright position (sufficient compensation), and forward positioning (strong compensation). Backward displacement in supine position was the most common response, seen in seven of thirteen subjects. It is possible that different strategies would have been observed if the jaw had been free to move. However, Shiller *et al.* (1999) found no anterior-posterior jaw displacement in supine position. A stable jaw position implies that tongue changes are not correlated to jaw changes. The range of backward displacements was consistent with two strategies: little or no opposition to gravity (e.g., Subjects 1 and 3), and "sufficient" opposition needed to maintain some feature, such as, airway opening or phoneme quality (e.g., Subjects 2, 5, 7, 9, and 11). Two additional subjects showed a greater opposition to gravity by positioning the tongue forward in supine position, clearly maintaining an open airway. These nine subjects used different strategies for supine speech (rotation, translation, local displacement), but each applied his/her strategy fairly consistently across phonemes. Four other subjects varied in a non-systematic manner across phonemes. These variable strategies suggest a variety of speaker dependent compensations to the change in gravity, which are motivated by constraints of airway protection, acoustic clarity and user idiosyncracies.

There was mild support for preservation of the constriction location consistent with Tiede *et al.* (2000) and Engwall (2006). However, the missing tongue tip and the large number of upright-supine comparisons in which the entire contour differed by less than 1 mm prevents strong interpretation of this finding.

The corrections used on supine contours suggest control strategies used to adjust the tongue to the changed gravitational orientation. Rigid body transformation implies a muscle synergy in which the tongue moves as a single entity to maintain shape in supine position. This is possible if orthogonal muscles, such as verticalis/GG versus transverse, contract simultaneously to form isometric stiffening. Orthogonal muscle contractions would reduce or prevent local shape changes.

Affine transformation implies a homogeneous, but less rigid response to the gravitational changes, so that the tongue can stretch or shear uniformly. This would occur if the additional GGP activity found in supine position (Sauerland and Mitchell, 1975) changed the shape of the posterior tongue. However, the present data showed that tongue shape is often constant; rigid body transformation accounts for the bulk of the difference between upright and supine tongue position. This suggests that, like the muscles that stabilize the middle ear bones, GGP contraction is usually exactly sufficient to overcome the airway pressures of supine position. In addition GGP's orthogonality to transverse would contribute further to the rigid body aspect of airway preservation. Muscle contraction patterns would be needed to confirm these hypotheses.

V. CONCLUSION

This study confiRMS several findings seen in previous studies including: subject specific strategies for tongue displacement in supine position; preservation of formant frequencies in continuous speech in supine position; possibly greater preservation of tongue position at the constriction location than at other regions of the vocal tract. The study also showed that ultrasound can determine a speaker's strategy for use in subsequent tongue correction. As MRI speech data sets tend to be short it is possible to collect and analyze the identical data set on MRI and ultrasound.

For some subjects there is enough variability between upright and supine contours that the datasets cannot be used interchangeably, for example, in some vocal tract models or when considering the effects of a supine procedure on speech or muscle tissue. However, in such cases rigid, affine, or local transformations may be useful to transform supine data to an upright position.

ACKNOWLEDGMENT

The research was supported in part by NIH Grant No. R01-DC01758 from the NIDCD.

Badin, P., Bailly, G., Reveret, L., Baciu, M., and Segebarth, C. (**2002**). "Three-dimensional linear articulatory modeling of tongue, lips and face, based on MR1 and video images," J. Phonetics **30**, 533–553.

Engwall, O. (**2006**). "Assessing MRI measurements: Effects of sustenation, gravitation and coarticulation," *Speech production: Models, Phonetic Processes and Techniques*, edited by J. Harrington and M. Tabain (Psychology Press, New York), pp. 301–314.

Epstein, M., and Stone, M. (**2002**). "Validation of upright - supine methods," Working Papers and Technical Reports, No. 2002-01, Vocal Tract Visualization Laboratory, University of Maryland Dental School, Baltimore, MD, pp. 1–7.

Epstein, M., and Stone, M. (**2005**). "The tongue stops here: Ultrasound imaging of the palate," J. Acoust. Soc. Am. **118**, 2128–2131.

Kitamura, T., Takemoto, H., Honda, K. Shimada, Y., Fujimoto, I., Syakudo, Y., Masaki, S., Kuroda, K., Oku-uchi, N., and Senda, M. (**2005**). "Difference in vocal tract shape between upright and supine postures: Observations by an open-type MRI scanner," Acoust. Sci. & Tech. **26**, 465–468.

Li, Kambhamettu, and Stone. (**2005a**).

Li, *et al.* (**2005b**).

Morrish, K., Stone, M., Shawker, T., and Sonies, B. C. (**1985**). "Distinguishability of tongue shape during vowel production," J. Phonetics **13**, 89–204.

Myamoto, K., Ozbek, M. M., Lowe, A. A., and Fleetham, J. A. (**1997**). "Effect of body position on tongue posture in awake patients with obstructive sleep apnea," Thorax **52**, 255–259.

Narayanan, S., Alwan, A. A., and Haker, K. (**1995**). "An articulatory study of fricative consonants using magnetic resonance imaging," J. Acoust. Soc. Am. **98**, 1325–1347.

Niimi, S., Kumada, M., and Niitsu, M. (**1994**). "Functions of tongue-related muscles during production of the five Japanese vowels," Ann. Bull. RILP **28**, 33–39.

Otsuka, R., Ono, T., Ishiwata, Y., and Kuroda, T. (**2000**). "Respiratory related genioglossus electromyographic activity in response to head rotation and changes in body position," Angle Orthod. **70**, 63–69.

Parthasarathy, V., Stone, M., and Prince, J. L. (**2005**). "Spatiotemporal visualization of the tongue surface using ultrasound and kriging (SURFACES)," Clin. Linguist. Phonetics **19**, 529–544.

Perkell, J., *et al.* (**1992**).

Perkell, J., and Nelson, W. (**1982**). "Articulatory targets and speech motor control: A study of vowel production," *Speech Motor Control*, edited by S. Grillner, B. Lindblom, J. Lubker, and A. Persson (Pergamon, New York), pp. 187–204.

Perkell, J., and Zandipour, M. (**2002**). "Economy of effort in different speaking conditions. II. Kinematic performance spaces for cyclical and speech movements," J. Acoust. Soc. Am. **112**, 1642–1651.

Reinicke, C., Obijou, N., and Trankmann, J. (**1998**). "The palatal shape of upper removable appliances. Influence on tongue position in swallowing," J. Orofac. Orthop. **59**, 2002–2007.

Rose, E. C., Staats, R., Lehner, M., and Jonas, I. E. (**2002**). "Cephalometric analysis in patients with obstructive sleep apnea. Part I: diagnostic value," J. Orofac. Orthop. **63**, 143–153.

Sauerland, E. K., and Mitchell, P. (**1975**). "Electromyographic activity of intrinsic and extrinsic muscles of the human tongue," Tex Rep. Biol. Med. **33**, 445–455.

Shiller, D. M., Ostry, D. J., and Gribble, P. L. (**1999**). "Effects of gravitational load on jaw movements in speech," J. Neurosci. **19** 9073–9080.

Stone, M., Sonies, B., Shawker, T., Weiss, G., and Nadel, L. (**1983**). "Analysis of real-time ultrasound images of tongue configuration using a grid-digitizing system," J. Phonetics **11**, 207–218.

Stone, M., Davis, E., Douglas, A., Ness Aiver, M., Gullapalli, R., Levine, W., and Lundberg, A. (**2001**). "Modeling the motion of the internal tongue from tagged cine-MRI images," J. Acoust. Soc. Am. **109**, 2974–2982.

Stone, M., Lundberg, A., and Grouse, U. (**2002a**). "Exploring muscle interactions using tagged cine-MRI," presented at Speech Motor Control Conference, Williamsburg, VA, March 14.

Stone, M., Sutton, M. Parthasarathy, V., Prince, J., Li, M., and Kambhamettu, C. (**2002b**). "Effects of upright and supine orientation on tongue position during silence," Joint Meeting of the Acoustical Society of America, Iberoamerican Congress of Acoustics and Mexican Congress on Acoustics. Cancun, Mexico. December 2–6.

Stone, M., Parthasarathy, V., Prince, J. L., and Iskarous, K. (**2003**). "Using ultrasound and tagged Cine-MRI to infer the functional relationship between tongue internal strains and surface shapes," *Proceedings of the International Conference on Phonetic Sciences*, Barcelona, Spain, August 3–9.

Tiede, M. K., Masaki, S., and Vatikiotis-Bateson, E. (**2000**). "Contrasts in speech articulation observed in sitting and supine conditions," *Proceedings of the 5th Seminar on Speech Production*. Bavaria, pp. 25–28.

Unser, M., and Stone, M. (**1992**). "Automated Detection of tongue surfaces in sequences of ultrasound images," J. Acoust. Soc. Am. **91**, 3001-3007.

Speech recognition as a function of high-pass filter cutoff frequency for people with and without low-frequency cochlear dead regions

Vinay[a)]
All India Institute of Speech and Hearing, Manasagangothri, Mysore-570006, India

Brian C. J. Moore
Department of Experimental Psychology, Cambridge University, Downing Street, Cambridge CB2 3EB, England

(Received 28 July 2006; revised 25 January 2007; accepted 12 March 2007)

Regions in the cochlea with no (or very few) functioning inner hair cells and/or neurons are called "dead regions" (DRs). The recognition of high-pass filtered nonsense syllables was measured as a function of filter cutoff frequency for hearing-impaired people with and without low-frequency (apical) cochlear DRs. The diagnosis of any DR was made using the TEN(HL) test, and psychophysical tuning curves were used to define the edge frequency (f_e) more precisely. Stimuli were amplified differently for each ear, using the "Cambridge formula." For subjects with low-frequency hearing loss but without DRs, scores were high (about 78%) for low cutoff frequencies, remained approximately constant for cutoff frequencies up to 862 Hz, and then worsened with increasing cutoff frequency. For subjects with low-frequency DRs, performance was typically poor for the lowest cutoff frequency (100 Hz), improved as the cutoff frequency was increased to about $0.57f_e$, and worsened with further increases. These results indicate that people with low-frequency DRs are able to make effective use of frequency components that fall in the range $0.57f_e$ to f_e, but that frequency components below $0.57f_e$ have deleterious effects. The results have implications for the fitting of hearing aids to people with low-frequency DRs.
 [DOI: 10.1121/1.2722055]

PACS number(s): 43.71.Ky, 43.66.Ts, 43.71.Gv [KWG] Pages: 542–553

I. INTRODUCTION

Cochlear hearing loss is often associated with loss of function of the hair cells within the cochlea (Engström, 1983; Schuknecht, 1993; Borg *et al.*, 1995). Sometimes, the inner hair cells (IHCs) at certain places along the basilar membrane may be completely nonfunctioning or even missing, which means that there is no transduction of basilar-membrane vibration at those places. In addition, the auditory neurons associated with certain places may be nonfunctioning or degenerated, which again means that information about basilar-membrane vibration at those places is not transmitted to the brain. Places where there are no or very few functioning IHCs and/or neurons are referred to as "dead regions" (DRs; Moore and Glasberg, 1997; Moore, 2001), although the concept has been given a variety of other names (Troland, 1929; Gravendeel and Plomp, 1960). DRs appear to be relatively common in cases of severe or profound hearing loss, with a prevalence greater than 50% when the hearing loss exceeds 70 dB (Aazh and Moore, 2007; Vinay and Moore, 2007).

In anatomical studies, the extent of a DR can be defined in terms of "place" or distance along the basilar membrane. A given place can in turn be related to the characteristic frequency (CF) for that place, using a frequency-place map (Greenwood, 1961, 1990; Moore, 1986; Schuknecht and Gacek, 1993). Thus, the extent of a DR can be defined in terms of the CFs of the IHCs or neurons immediately adjacent to the DR (Moore, 2001). When a DR extends from low CFs up to a certain value (i.e., there is a low-frequency DR), or from high CFs down to a certain value (i.e., there is a high-frequency DR), the extent of the DR can be defined in terms of a single edge frequency, denoted here f_e.

A test for diagnosing DRs in the clinic was described by Moore *et al.* (2000). The test is based on a masking noise, called "threshold-equalizing noise" (TEN), which is spectrally shaped so that, for normal-hearing subjects, the masked threshold of a pure tone is almost independent of frequency. A similar noise was used earlier by Langenbeck (1965). The TEN level is specified as the level in a one-ERB_N (132 Hz) wide band centered at 1000 Hz, where ERB_N denotes the average equivalent rectangular bandwidth of the auditory filter as determined for young normal-hearing listeners at moderate sound levels (Glasberg and Moore, 1990; Moore, 2003). For normal-hearing listeners, the signal level at masked threshold is approximately equal to the noise level/ERB_N. For a hearing-impaired listener without a DR, the signal is detected via IHCs and neurons with CFs close to the signal frequency. In such a case, the threshold in the TEN is only 2–5 dB higher than normal (Moore, 2001, 2004).

When a person has a DR, and the signal produces maximum basilar-membrane vibration in the DR, the signal is

[a)]Address for correspondence: Vinay, 231, Shri Vinyasa, 9th cross, 5th main, Vijayanagar First Stage, Mysore 570 017, Karnataka, India. Electronic mail: shrivinyasa@gmail.com

detected, if sufficiently intense, via IHCs/neurons with CFs remote from the signal frequency; in other words, off-place listening occurs. In such a case, the signal threshold in the TEN is expected to be markedly higher than normal. Moore *et al.* (2000) proposed the following "rule:" if the threshold in the TEN is 10 dB or more above the TEN level/ERB_N, and the TEN produces at least 10 dB of masking, this is indicative of a DR at the signal frequency. These are referred to as the TEN-test criteria. In the original version of the TEN test (Moore *et al.*, 2000), all levels were specified and calibrated in dB SPL. In a more recent version of the test, which is the one used in this paper, levels are calibrated in dB HL, and the test is referred to as the TEN(HL) test (Moore *et al.*, 2004). However, the same criteria apply for both tests.

An alternative but more time-consuming way of diagnosing DRs is to measure psychophysical tuning curves (PTCs), which are obtained using a fixed signal presented at a low sensation level. The level of a narrow-band masker required to mask the signal is plotted as a function of the masker center frequency. When the tip of the PTC is shifted away from the signal frequency, this is taken as indicating a DR at the signal frequency. The frequency at the tip of the PTC is assumed to indicate the value of f_e (Thornton and Abbas, 1980; Florentine and Houtsma, 1983; Turner *et al.*, 1983; Moore and Alcántara, 2001). Although it has been reported that there can be discrepancies between the results of the TEN test and PTCs (Summers *et al.*, 2003), these discrepancies seem to be caused by the influence of beats and combination tones on the PTCs (Kluk and Moore, 2005). When precautions are taken to minimize such effects, there is generally a good correspondence between the value of f_e estimated using the TEN test and the value of f_e estimated using PTCs (Moore *et al.*, 2000; Huss and Moore, 2003; Kluk and Moore, 2005, 2006). In the studies of Kluk and Moore (2005, 2006), 20 subjects met the TEN-test criteria for a dead region in at least one ear (often both ears) and there was only one case (one ear) for which the PTC did not show a shifted tip.

This paper is concerned with the effect of low-frequency DRs on speech intelligibility, and with changes in intelligibility when speech is high-pass filtered so as to remove some of the frequency components falling within a DR. We start with a brief review of previous work assessing the effects of filtering on speech intelligibility for people with DRs. Most previous work on this topic has been focused on subjects with high-frequency DRs. There have been many reports suggesting that people with moderate-to-severe hearing loss at high frequencies often do not benefit from amplification of high frequencies, or even perform more poorly when high frequencies are amplified (Hogan and Turner, 1998; Moore *et al.*, 1985; Murray and Byrne, 1986; Ching *et al.*, 1998; Turner and Cummings, 1999; Amos and Humes, 2001); however, no test for the presence of DRs was conducted in those studies.

In two recent studies, subjects were tested for the presence of DRs using both PTCs and the TEN test (Vickers *et al.*, 2001; Baer *et al.*, 2002). All subjects had high-frequency hearing loss, but some had high-frequency DRs and some did not. Generally, the subjects with DRs had more severe high-frequency hearing losses than those without DRs. The speech stimuli were vowel-consonant-vowel (VCV) nonsense syllables. Prior to presentation via earphones, the stimuli were subjected to the frequency-dependent amplification prescribed by the "Cambridge" formula (Moore and Glasberg, 1998), which is roughly equivalent to a half-gain rule for mid-range frequencies. The goal of the amplification was to restore audibility as far as possible, while avoiding excessive loudness. Subjects were tested using broadband speech (upper frequency limit 7500 Hz) and speech that was low-pass filtered with various cutoff frequencies.

In one study (Vickers *et al.*, 2001), all stimuli were presented in quiet. In another study (Baer *et al.*, 2002), the stimuli were presented in steady speech-shaped noise, using a speech-to-noise ratio chosen to give a moderate reduction in performance relative to that measured in quiet. The general pattern of results was similar for the two studies. For subjects without DRs, performance improved progressively with increasing cutoff frequency. This indicates that these subjects were able to make use of high-frequency information. For subjects with DRs, scores tended to increase with increasing cutoff frequency up to 1.5 f_e to 2 f_e. For higher cutoff frequencies, scores remained roughly constant or decreased slightly (not significantly). These results indicate that the subjects with DRs extracted little or no information from frequencies that were well inside the DR.

Vestergaard (2003) presented data broadly consistent with the findings summarized above. He tested experienced users of hearing aids, using the TEN test to diagnose DRs. He measured speech intelligibility using low-pass filtered word lists. Subjects listened through their own hearing aids, all of which provided substantial amplification at frequencies where there was hearing loss. For subjects with no DR or high-frequency DRs starting at 3 kHz or above, performance tended to improve progressively with increasing cutoff frequency. For subjects with high-frequency DRs starting in the range 0.75–1.5 kHz, the pattern was less consistent, but the mean scores showed no clear change with increasing cutoff frequency above 1 kHz.

Mackersie *et al.* (2004) compared speech intelligibility for two groups of subjects, with and without high-frequency DRs, diagnosed using the TEN test. The audiometric thresholds were matched for the two groups. Subjects were tested while wearing a hearing aid that provided substantial gain in frequency regions of hearing loss. Speech identification scores were obtained for unfiltered stimuli and stimuli that were low-pass filtered at f_e, 1.41 f_e and 2 f_e. Filter settings for the ears without suspected DRs were the same as settings for the threshold-matched counterparts. In quiet and in low levels of noise (signal-to-noise ratios close to 15 dB), scores were significantly higher for the wide-band (unfiltered) stimuli than for the filtered stimuli, and performance was similar for the ears with and without DRs. In high levels of noise (signal-to-noise ratios close to 0 dB), mean scores were highest for the wide-band speech for the ears without DRs, but performance reached an asymptote at a lower cutoff frequency for the ears with DRs. In other words, when sub-

stantial noise was present, the subjects with DRs did not benefit from amplification of frequency components well inside the DR.

Overall, these results support the idea that subjects with extensive high-frequency DRs do not make as effective use of audible speech information at high frequencies as subjects without such DRs. Furthermore, the results support the idea that increasing the audibility of speech for frequencies well inside a high-frequency DR does not lead to concomitant increases in speech intelligibility.

There has been much less research on the perception of speech by people with low-frequency DRs. Thornton and Abbas (1980) measured the intelligibility of word lists for three subjects who had been diagnosed as having low-frequency DRs on the basis of PTCs with upwards-shifted tips and on the basis of masking effects produced by a high-frequency tone. For the broadband speech presented at 80 dB HL, scores ranged from 56% to 88%, while scores for a control group of five normal-hearing subjects ranged from 94% to 100%. When the speech was high-pass filtered at 3000 Hz, scores worsened and were similar for the two groups (34% to 46% for the subjects with DRs and 16% to 38% for the control group). Thus, the subjects with low-frequency DRs were able to use the high-frequency information in the speech as well as or better than the control group. When the speech was low-pass filtered at 1500 Hz, performance worsened dramatically (12%–44%) for the group with DRs, but scores for the control group remained high (76% to 84%). Finally, when the speech was low-pass filtered at 1500 Hz and presented simultaneously with noise that was high-pass filtered at 1500 Hz, scores dropped even further for the group with DRs (0% to 22%), but remained high for the control group (64% to 80%). The effect of the noise suggests that, for the group with DRs, low-frequency speech information was partly obtained via IHCs and neurons tuned to middle or high frequencies.

Van Tasell and Turner (1984) presented a case study of a person with a low-frequency DR, diagnosed on the basis of PTCs with upward-shifted tips. The speech stimuli were monosyllabic words, sentences, and nonsense syllables. A control group of normal-hearing subjects was also tested. Speech recognition scores of all subjects were close to 100% correct for all stimuli presented unfiltered at a moderate intensity level. When stimuli were low-pass filtered, the performance of the hearing-impaired subject fell below that of the control group, but remained considerably above chance. A further worsening in the hearing-impaired subject's recognition of nonsense syllables resulted from the addition of a high-pass masking noise, indicating that his performance in the filtered quiet condition was attributable largely to information carried via IHCs and neurons with medium to high CFs.

Halpin *et al.* (1994) measured the recognition of words in 14 subjects with low-frequency hearing loss. Seven were diagnosed as having "true" hearing at low frequencies, and the other seven were diagnosed as having low-frequency DRs using a test based on the masking of low-frequency tones by a high-frequency tone situated at the border of the region of hearing loss. Speech recognition scores were generally lower (10% to 88%) for the group with DRs than for the group without DRs (84% to 100%). Obtained scores were compared to scores predicted from the articulation index (ANSI, 1969). The predictions were based either on the "true" audiograms or on "hypothetical" audiograms where the hearing loss at low frequencies was set to a large value to simulate low-frequency DRs. For the subjects without any DR, the best correspondence between obtained and predicted scores was obtained using the "true" audiograms for the predictions. For the subjects with DRs, the best correspondence was obtained using the "hypothetical" audiograms. Halpin *et al.* concluded that, for broadband speech, subjects with low-frequency DRs extracted little or no information from low-frequency components in the speech.

The present study was designed to provide much more quantitative information about the ability of subjects with low-frequency DRs to use information from frequency components of the speech falling within the DR. To do this, the recognition of VCV nonsense syllables was measured for speech that was high-pass filtered with a large number of cutoff frequencies. The results were intended to provide guidelines as to the appropriate form of amplification for people with low-frequency DRs. For comparison purposes, we also tested a group of subjects with low-frequency hearing loss but without any DR.

II. METHOD

A. Subject selection

Twenty-eight subjects with sensorineural hearing loss were selected on the basis of their having reasonably flat audiograms with a hearing loss at low frequencies (averaged over the frequencies 500, 750, and 1000 Hz) of 40 dB or more. Their ages ranged from 17 to 85 years (mean age =47.9 years; 15 males and 13 females). All were attending an audiology clinic, some for the first time and some for a repeat visit. None of the subjects had any history of ear discharge or fullness in the ear. Audiological reevaluation was done for subjects who had attended the clinic previously. Subjects generally attended for the purpose of being fitted with a hearing aid. This usually happened after the audiological evaluation, following clearance by an ENT specialist. None of the subjects had nonauditory neural conditions. The diagnosis of sensorineural hearing loss was based on an air-bone gap in audiometric threshold less than or equal to 10 dB for frequencies from 250 to 4000 Hz (Madsen OB922 audiometer with Telephonics TDH39 headphones and Radio ear B-71 bone vibrator) and normal results for immittance measurements (Grason-Stadler Tympstar). Thresholds were measured using the modified Hughson-Westlake procedure described by Carhart and Jerger (1959).

The presence or absence of DRs was initially diagnosed using the TEN(HL) test (Moore *et al.*, 2004), which has test-tone frequencies of 500, 750, 1000, 1500, 2000, 3000, and 4000 Hz (see below for details). For all subjects who met the TEN(HL)-test criteria at any frequency, or for whom the masked threshold was higher than the output limit of the system (86 dB HL; see below for details), PTCs were mea-

TABLE I. Data for the test ears of subjects diagnosed as having DRs. The first four columns show, from left to right, subject identification number (ID), ear, age, and gender. In the other columns, the upper entry in each cell shows the audiometric threshold (dB HL) and the lower entry shows masked thresholds in TEN(HL) with a level of 70 dB/ERB_N. NR indicates "no response." The bold numbers indicate frequencies falling in a DR. Asterisks indicate the approximate value of f_e.

				Frequency (Hz)								
ID	Ear	Age	Gender	250	500	750	1000	1500	2000	3000	4000	8000
1	Right	50	F	70	**75**	**75**	**75***	65	65	65	70	70
					NR	**NR**	**NR**	74	72	74	78	
2	Right	45	F	70	**70**	**75**	**75**	**75**	**75***	65	65	60
					NR	**NR**	**NR**	**NR**	**NR**	74	74	
3	Right	65	F	70	**75**	**75**	**75***	60	60	60	55	60
					NR	**NR**	**NR**	74	76	76	76	
4	Left	18	M	75	**75**	**75**	**80***	65	60	60	60	80
					NR	**NR**	**NR**	76	74	78	78	
5	Right	65	F	80	**80**	**80**	**80***	75	70	75	75	80
					NR	**NR**	**NR**	78	76	76	78	
6	Left	18	F	50	**75**	**75**	**75***	70	70	75	75	90
					NR	**NR**	**NR**	76	78	78	78	
7	Left	55	M	60	**75**	**80**	**85**	**80***	75	75	75	NR
					NR	**NR**	**NR**	**NR**	76	76	76	
8	Right	64	F	65	**70**	**70**	**70***	60	60	60	65	60
					86	**NR**	**NR**	76	76	78	78	
9	Right	38	M	80	**80**	**80**	**80**	**75***	55	55	55	80
					NR	**NR**	**NR**	**86**	78	76	74	
	Left			80	**85**	**80**	**80**	**75***	55	55	55	75
					NR	**NR**	**NR**	**NR**	74	74	76	
10	Right	49	F	75	**80**	**80**	**75**	**75***	65	65	65	NR
					NR	**NR**	**NR**	**NR**	76	76	78	
11	Right	84	M	70	**65**	**70**	**80**	**80**	**80***	65	65	105
					84	**82**	**NR**	**NR**	**NR**	76	78	
	Left			70	**70**	**80**	**85**	**85**	**85***	70	70	80
					84	**NR**	**NR**	**NR**	**NR**	74	78	
12	Right	38	M	75	**80**	**80**	**75**	**70***	55	60	65	80
					NR	**NR**	**NR**	**86**	74	76	78	
13	Left	57	F	70	**80**	**80**	**80**	**85**	**75**	**75***	70	80
					NR	**NR**	**NR**	**NR**	**NR**	**NR**	78	
14	Left	24	M	80	**80**	**80**	**85**	**80**	**80***	65	60	70
					NR	**NR**	**NR**	**NR**	**NR**	74	74	
15	Right	22	M	70	**70**	**75**	**85**	**85**	**85***	70	70	65
					84	**NR**	**NR**	**NR**	**NR**	78	78	
	Left			65	**70**	**75**	**80**	**80**	**85***	70	65	70
					84	**86**	**NR**	**NR**	**NR**	76	78	
16	Left	19	F	70	**75**	**80**	**85**	**85**	**85**	**85***	70	70
					NR	**NR**	**NR**	**NR**	**NR**	**NR**	78	

sured to confirm the diagnosis and to define the value of f_e more precisely (see below for details). Three signal frequencies were chosen, in the region of the value of f_e indicated by the results of the TEN(HL) test. PTCs were also measured for three subjects who did not meet the TEN(HL)-test criteria at any frequency. Sixteen subjects were diagnosed as having a low-frequency DR in one or both ears. An ear was selected for further testing if the value of f_e was above 750 Hz (it was anticipated that a DR with f_e at 750 Hz or below would have little effect on speech intelligibility), and if the DR appeared to be continuous rather than "patchy." In some cases only one ear was tested due to the subject only being available for a limited time. Information about the 19 test ears of the subjects with DRs is given in Table I. Twelve subjects were diagnosed as not having a DR. Two were available for only a limited time and were tested using one ear only. The remainder were tested using both ears. Information about the 22 test ears of these subjects is given in Table II. It is noteworthy that the subjects without low-frequency DRs had audiometric thresholds at low frequencies that were 70 dB HL or less, while the subjects with low-frequency DRs had audiometric thresholds at low frequencies that were 70 dB HL or more. However, it has been shown that the audiometric threshold does not provide a reliable indication of the presence or absence of a DR, although the prevalence of DRs increases with increasing hearing loss (Aazh and Moore, 2007; Vinay and Moore, 2007).

Subjects S3–S13 from the group with DRs and subjects S24–S28 from the group without DRs had previous experience with (mostly unilaterally fitted) hearing aids. The duration of hearing loss was approximately as follows: 3–10 years for S2, S4, S6, S9, S10, S12, S14–S18, S20, S22, S26, and S27; 10–20 years for S1, S3, S5, S7, S8, S13, S19, S21, and S28; more than 20 years for S23 and S25; unknown for S11 and S25. The cause of hearing loss was

TABLE II. As in Table I but for the test ears of subjects for whom no DR was diagnosed.

				Frequency (Hz)								
ID	Ear	Age	Sex	250	500	750	1000	1500	2000	3000	4000	8000
17	Right	18	M	40	40	40	45	45	50	50	55	55
					72	70	74	76	74	74	74	
	Left			40	40	40	40	50	55	55	50	60
					70	70	74	72	72	74	74	
18	Right	34	F	45	55	55	60	55	50	50	60	60
					70	70	70	70	72	72	72	
	Left			35	35	40	50	50	50	45	40	55
					70	70	72	70	72	70	70	
19	Right	54	M	50	50	55	60	65	60	65	60	70
					72	72	70	74	76	74	74	
	Left			55	55	55	60	60	65	65	60	65
					70	70	70	70	74	78	76	
20	Right	50	F	35	55	60	60	60	60	60	65	85
					70	70	70	70	70	72	76	
21	Right	63	M	40	40	45	50	45	45	50	55	60
					70	72	70	70	72	76	76	
	Left			35	50	50	50	50	45	50	55	65
					74	72	70	72	72	72	74	
22	Right	17	F	55	65	60	60	60	60	65	70	60
					74	74	76	74	72	74	78	
	Left			50	55	55	55	55	55	60	65	60
					74	74	74	72	74	76	76	
23	Right	75	F	50	50	50	50	50	50	50	55	60
					70	70	70	70	70	72	74	
24	Right	85	M	55	60	60	65	65	65	65	70	65
					72	72	72	70	74	74	76	
	Left			60	55	60	60	65	70	70	70	65
					72	72	76	74	74	78	76	
25	Right	84	M	40	50	50	50	55	55	50	50	60
					74	72	72	74	74	72	74	
	Left			35	40	40	40	45	50	50	45	50
					70	72	70	70	70	74	72	
26	Right	40	M	50	65	60	60	60	60	65	65	65
					74	72	72	72	74	74	76	
	Left			40	60	55	50	55	60	65	70	65
					70	74	72	74	74	74	78	
27	Right	42	M	55	65	65	65	60	60	60	60	65
					70	74	74	74	76	74	76	
	Left			60	65	70	70	65	65	70	70	70
					74	74	74	76	76	74	78	
28	Right	67	M	30	40	45	45	45	45	45	45	55
					70	70	70	70	70	70	70	
	Left			35	40	45	45	45	45	45	50	50
					72	70	70	70	72	70	74	

genetic (as indicated by family history) for S15, but was unknown for the other subjects; no subject was diagnosed with Ménière's disease.

B. Application of the TEN(HL) test

The TEN(HL)-test CD was replayed via a Philips 729 K CD player connected to Madsen OB922 audiometer equipped with TDH39 earphones. The level of the signal and the TEN(HL) were controlled using the attenuators in the audiometer. The TEN(HL) level was 70 dB HL/ERB_N. The signal level was varied in 2-dB steps to determine the masked thresholds, as recommended by Moore *et al.* (2004). A "no response (NR)" was recorded when the subject did not indicate hearing the signal at the maximum output level of the audiometer, which was 86 dB HL for the signals derived from the TEN(HL) CD. The presence or absence of a DR at a specific frequency was based on the criteria suggested by Moore *et al.* (2004). If the masked threshold in the TEN(HL) was 10 dB or more above the TEN(HL) level/ERB_N, and the TEN(HL) elevated the absolute threshold by 10 dB or more, then a DR was assumed to be present at the signal frequency. If the masked threshold in the TEN was less than 10 dB above the TEN(HL) level/ERB_N, and the TEN elevated the absolute threshold by 10 dB or more, then a DR was assumed to be absent. If the masked threshold in the TEN(HL) was 10 dB or more above the TEN(HL) level/ERB_N, but the TEN(HL) did not elevate the absolute threshold by 10 dB or more, then the result was considered inconclusive.

C. Measurement of PTCs

PTCs were measured using a Maico 53 dual-channel audiometer equipped with TDH39 headphones. The audiom-

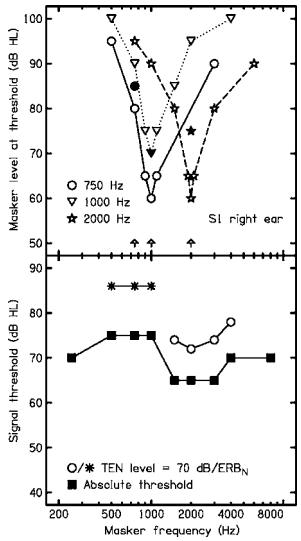


FIG. 1. Results of the TEN(HL) test and PTCs for the right ear of S1. The lower panel shows the results of the TEN(HL) test. The signal threshold is plotted as a function of signal frequency. Filled squares indicate absolute thresholds. Open circles and asterisks show thresholds obtained in the presence of TEN(HL) with a level of 70 dB/ERB$_N$. The asterisks indicate that the TEN(HL) test criteria were met and/or that the masked threshold was greater than the output limit of the system; for this ear, both were the case. The upper panel shows PTCs for signal frequencies of 750, 1000, and 2000 Hz, as indicated in the key. Filled symbols show the signal level and frequency (also indicated by the up-pointing arrows), and open symbols show the masker level required for threshold. Note that the PTC for a signal frequency of 750 Hz has a tip at 1000 Hz.

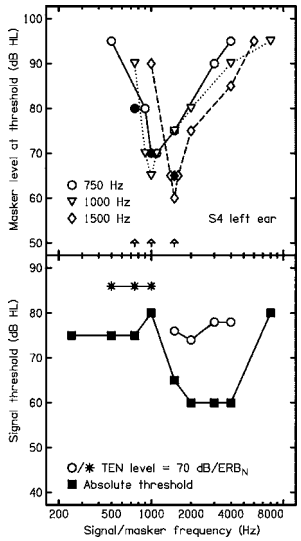


FIG. 2. As in Fig. 1, but for the left ear of S4.

eter was set to dual-frequency mode. The signal tone was generated in one channel. It was presented at 10 dB SL and was pulsed on and off in a regular sequence (0.25 s on, 0.25 s off). A narrow-band noise masker was selected in the other channel. The noise conformed to the specifications given in ANSI-S3.6 (2004), and had a bandwidth between $\frac{1}{3}$ and $\frac{1}{2}$ oct. This relatively large bandwidth is required to reduce the influence of beats on the PTCs (Kluk and Moore, 2004, 2005). The two channels were mixed in order to present the tone and noise to the same ear. The subjects were asked to respond when they could hear the tone in the presence of the noise. The minimum noise level required to mask the tone was determined by manual adjustment of the noise level. This was repeated for several masker frequencies placed at, below, and above the signal frequency. The number of masker frequencies used varied across subjects, depending on the time for which they were available. However, the masker frequencies were always closely spaced around the tip in order to define more precisely the value of f_e.

D. Measurement of the intelligibility of filtered speech

Speech intelligibility was measured using vowel-consonant-vowel (VCV) nonsense syllables, spoken by a man. Recordings were made in a sound-treated room using a Sony D-500 microphone and a computer with Audio Lab software (32 bit, Mono, 32 000-Hz sampling rate). The vowels /a/, /i/ and /u/ were used (same vowel at start and end of a given syllable) in combination with each of 21 consonants: /p, b, m, t, d, h, ʈ, ɖ, ɳ, k, g, r, l, ɭ, ç, ʝ, j, θ, ʃ, s, h/. This gave 63 tokens per list. Ten lists were recorded, each with the stimuli in a different order.

Speech stimuli with a nominal level of 65 dB SPL were amplified according to the Cambridge formula (Moore and Glasberg, 1998), which is intended to provide approximately equal loudness/ERB$_N$ over the speech frequency range from 500 to 4000 Hz. The formula specifics the insertion gain (IG) at frequency f as

$$\mathrm{IG}(f) = 0.48\mathrm{HL}(f) + \mathrm{INT}(f), \quad (1)$$

where HL(f) is the hearing loss in dB at f, and INT(f) is an intercept whose value depends upon frequency. The necessary gains were determined separately for each ear of each subject. Following amplification, stimuli were high-pass filtered with one of the following cutoff frequencies: 100, 300, 430, 610, 862, 1220, 1725, 2440, 3450, and 4880 Hz. A Tucker-Davis Technologies (TDT) System 3 was used to digitize the stimuli and to implement the Cambridge formula filter (Camfilter) and high-pass filter. The analog output was presented to one ear only using Sennheiser HD580 headphones. The Camfilter was designed using the FIR2 function in Matlab and contained 251 taps. The Camfilter took into account the frequency response of the HD580 headphones as measured on a KEMAR manikin (Burkhard and Sachs, 1975). The high-pass filter was designed using either the FIR1 function or the Remez function in Matlab and contained 479 taps. The response of the high-pass filter typically fell by about 50 dB when the frequency was decreased from the nominal cutoff frequency to 85 Hz below it.

Each subject was initially given examples of the stimuli, including stimuli with different cutoff frequencies. Following each presentation, the subject was instructed to write down the consonant heard and was asked to make a response even if they thought they were guessing. These results were not scored. Once subjects were familiar with the stimuli and were comfortable with the task, testing proper commenced. For testing proper, two lists were administered per ear per subject for each cutoff frequency. The choice of cutoff fre-

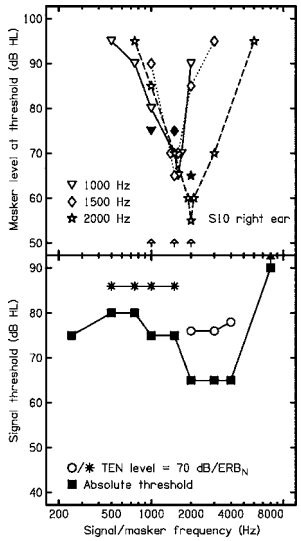


FIG. 3. As in Fig. 1, but for the right ear of S10. In this case, the signal frequencies for the PTCs were 1000, 1500, and 2000 Hz. The PTC for the 1000-Hz signal frequency has a tip at 1500 Hz.

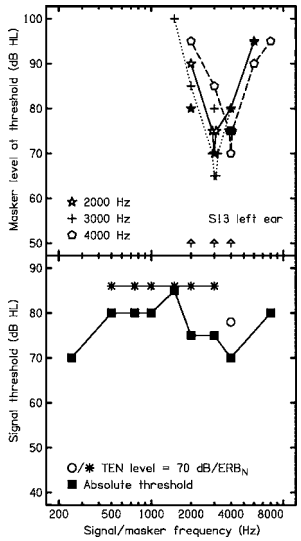


FIG. 4. As in Fig. 1, but for the left ear of S13. In this case, the signal frequencies for the PTCs were 2000, 3000, and 4000 Hz. The PTC for the 2000-Hz signal frequency has a tip at 3000 Hz.

quencies for each subject depended upon the time that each subject was available, but at least eight cutoff frequencies were used for each ear tested. When time was limited, the cutoff frequencies were concentrated around the estimated value of f_e. Testing was conducted in two sessions during each of which each cutoff frequency was used once. The order of cutoff frequencies was randomized for the first session and the order was reversed for the second session in order to counterbalance effects of practice and fatigue. Scores presented are the means (and standard deviations) for each cutoff frequency across the two sessions.

III. RESULTS

A. TEN(HL) test and PTCs

For the ears diagnosed as having low-frequency DRs, it was often the case that, when conducting the TEN(HL) test, the output limit of the system was reached for some frequencies within the assumed DR. The results were inconclusive for one or more test frequencies, because the masked threshold in the TEN(HL) was not 10 dB above the absolute threshold, for 14 out of the 19 ears. However, for most of these ears there were at least two signal frequencies for which the TEN(HL)-test criteria were met, i.e., the signal level at the output limit was more than 10 dB above the absolute threshold. PTCs were measured for all subjects for whom the TEN(HL)-test criteria were met for at least one frequency. For signal frequencies at which the results of the TEN(HL) test were inconclusive, the presence of a low-frequency DR was always confirmed by at least one PTC with a shifted tip. For all ears for which the TEN(HL)-test criteria were met, at least one PTC was measured with a shifted tip. The frequency at the shifted tip was taken as the best estimate of the value of f_e. The estimated value of f_e for each subject is given in Figs. 7–10, which are presented later.

Some examples of the results are shown in Figs. 1–4; the lower panels show the results of the TEN(HL) test and the upper panels show the results of the PTC measurements. For the right ear of S1 (Fig. 1), the TEN(HL)-test criteria were met for signal frequencies of 500, 750, and 1000 Hz and were not met for higher frequencies. The PTCs for signal frequencies of 1000 and 2000 Hz both had tips at their respective signal frequencies, but the PTC for a signal frequency of 750 Hz (circles) was shifted upwards to 1000 Hz. This suggests that the right ear of S1 had a low-frequency DR with f_e close to 1000 Hz. Note that the PTC for the signal frequency of 1000 Hz did not show a shifted tip, despite the TEN(HL)-test criteria being met at 1000 Hz. This can happen when the signal frequency used to measure the PTC falls very close to f_e.

For the left ear of S4 (Fig. 2), the TEN(HL)-test criteria were met for signal frequencies of 500 and 750 Hz, but the results for the signal frequency of 1000 Hz were inconclusive, as the masked threshold was above the output limit of the system and at the limit the signal level was less than 10 dB above the absolute threshold. The PTCs for signal frequencies of 1000 and 1500 Hz had tips at their respective signal frequencies, but the PTC for the signal frequency of 750 Hz was shifted upwards to between 1000 and 1100 Hz, suggesting a low-frequency DR with f_e=1000 or 1100 Hz. The value of f_e was taken as 1000 Hz in this case.

For the right ear of S10 (Fig. 3), the TEN(HL)-test criteria were met for signal frequencies of 1000 and 1500 Hz and were not met for higher frequencies. The results for signal frequencies of 500 and 750 Hz were inconclusive, as the masked threshold was above the output limit of the system and at the limit the signal level was less than 10 dB above the absolute threshold. The PTCs for signal frequencies of 1500 and 2000 Hz had tips at their respective signal frequencies. However, the PTC for the signal frequency of 1000 Hz had a tip at 1600 Hz. These results suggest a low-frequency DR with f_e=1500 or 1600 Hz; the value was taken as 1500 Hz in this case.

For the left ear of S13 (Fig. 4), the TEN(HL)-test criteria were met for signal frequencies of 2000 and 3000 Hz. The results for lower signal frequencies were inconclusive, as the

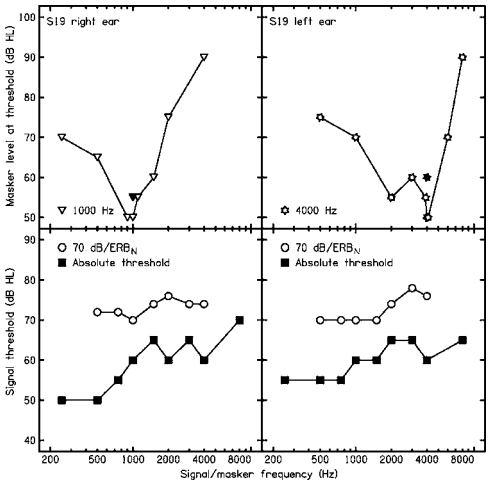


FIG. 5. As in Fig. 1, but showing the results for each ear of a subject (S19) without any DR. The TEN test criteria were not met for any signal frequency, and the PTCs have tips at the signal frequency.

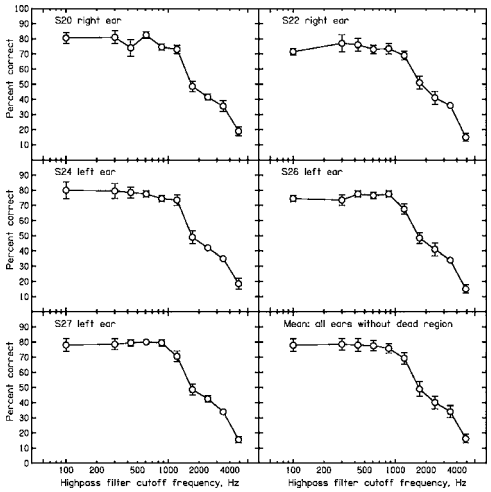


FIG. 6. Results of the VCV test for five ears of five subjects without any DR, and for the mean across all ears without any DR (bottom-right panel). The percent correct score is plotted as a function of high-pass filter cutoff frequency. For the individual ears, error bars indicate ± one standard deviation (SD) across repeated runs. For the mean data across ears, the error bars indicate ± one standard deviation (SD) across ears.

masked threshold was above the output limit of the system and at the limit the signal level was less than 10 dB above the absolute threshold. The TEN(HL)-test criteria were not met for the signal frequency of 4000 Hz. The PTCs for signal frequencies of 3000 and 4000 Hz had tips at their respective signal frequencies. However, the PTC for the signal frequency of 2000 Hz had a tip at 3000 Hz. These results suggest a low-frequency DR starting at 3000 Hz.

The estimated values of f_e for all subjects, based on the tip frequencies of the PTCs with shifted tips, are given in Figs. 7–10, which are presented later. Comparison of these values with the TEN(HL)-test results shown in Table I indicates a good correspondence between the TEN(HL)-test results and the PTCs.

For subjects for whom no DR was diagnosed using the TEN(HL) test, because the masked threshold was less than 10 dB above the TEN(HL) level per ERB_N for all signal frequencies, PTCs were not routinely measured. However, PTCs were measured for three subjects to check that the tips were always at the signal frequency, regardless of what the signal frequency was. An example of the results is shown in Fig. 5. PTCs were measured for a signal frequency of 1000 Hz for the right ear and 4000 Hz for the left ear. Each of the PTCs had a tip at the signal frequency. This was the case for all PTCs measured using ears that were diagnosed as not having a DR.

B. VCVs

The pattern of VCV scores for the subjects without DRs was very consistent across subjects. Figure 6 shows examples of the results for five ears of five subjects, chosen because they all had a hearing loss of about 60 dB at low frequencies. The bottom-right panel shows the results averaged across all 22 ears tested. As the high-pass filter cutoff frequency was increased, performance remained approximately constant at about 78% until the cutoff frequency was greater than 862 Hz. Further increases in cutoff frequency led to a progressive worsening in scores, and scores reached 10%–20% for the highest cutoff frequency (4880 Hz).

The results for the subjects with low-frequency DRs are shown in Figs. 7–10. In each panel, the up-pointing arrow indicates the value of f_e. For the lowest cutoff frequency, scores were markedly lower (30%–55%) than for the subjects without DRs. As the cutoff frequency was increased, scores initially either stayed approximately constant (e.g., S10 right ear; Fig. 8, bottom-left panel) or increased slightly (e.g., S1 and S2, right ears; Fig. 7, top two panels). Then, scores for nearly all ears increased up to the point where the cutoff frequency was somewhat below f_e. In other words, removal of the lower-frequency components that fell within the DR led to improved speech identification. With further increases in cutoff frequency, the scores for all ears decreased. For all ears, the cutoff frequency at which scores were the highest, f_{max}, fell below f_e. Based on a Wilcoxon matched-pairs test, the difference between f_{max} and f_e was highly significant ($p < 0.001$). The average ratio f_{max}/f_e was

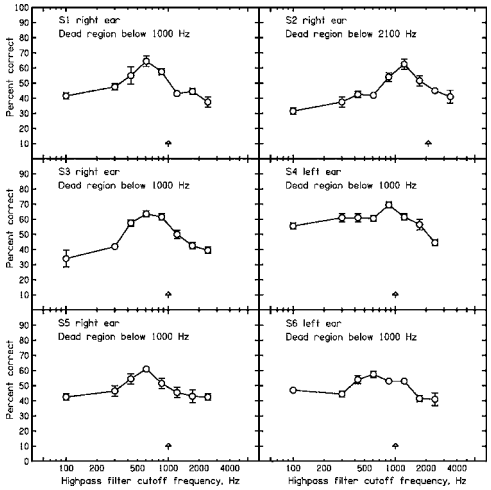


FIG. 7. Results of the VCV test for six ears of six subjects with low-frequency DRs. The up-pointing arrows indicate the estimated edge frequency of the DR, f_e. Otherwise as in Fig. 6.

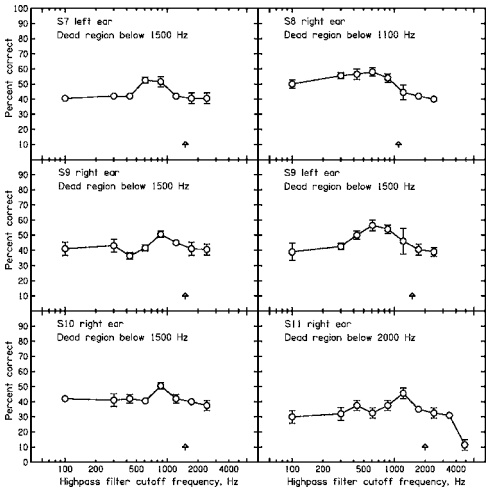


FIG. 8. Results of the VCV test for six ears of five subjects with low-frequency DRs. Otherwise as in Fig. 7.

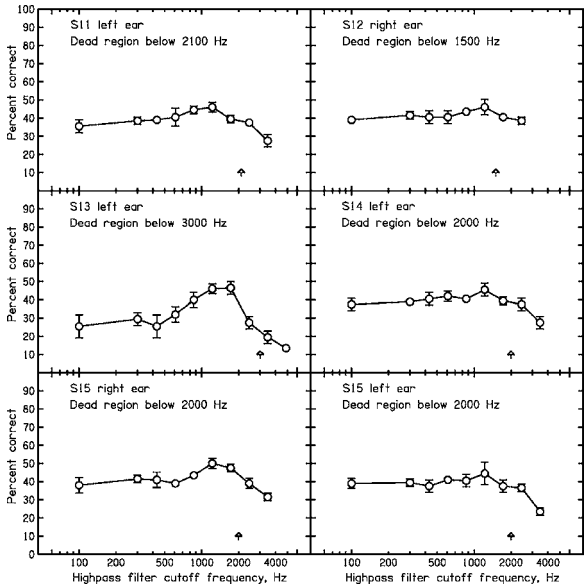


FIG. 9. Results of the VCV test for six ears of five subjects with low-frequency DRs. Otherwise as in Fig. 7.

1.76, with a standard deviation (SD) of 0.33. Put another way, the speech identification scores of the subjects with DRs were maximal when the cutoff frequency was $f_e/1.76$ or $0.57f_e$. It should be noted that the value of $f_{\max}$ was not fixed in frequency, but varied with f_e. The Pearson correlation between $f_{\max}$ and f_e was 0.875 ($n=19$), which is statistically significant ($p<0.01$).

It should be remembered that all of the speech stimuli were subjected to amplification according to the Cambridge formula (Moore and Glasberg, 1998). All of the subjects with low-frequency DRs received some benefit for intelligibility from the presence of amplified frequency components below f_e. The mean score at $f_{\max}$ was 52.2% (SD=8.3%) while the mean score when the cutoff frequency was at f_e (determined by interpolation) was 42.8% (SD=9.7%). The difference between scores at $f_{\max}$ and at f_e ranged from 3% to 23% across ears, with a mean of 10.4% (SD=4.7%).

To assess the extent to which the Cambridge formula amplification restored the audibility of the speech at low frequencies, consider a typical case with a DR where the absolute threshold at 500 Hz was 80 dB HL. Since the "normal" monaural absolute threshold (minimum audible pressure) at 500 Hz is about 7 dB SPL (Moore *et al.*, 1997), the threshold of 80 dB HL corresponds to about 87 dB SPL. It seems reasonable to assume that frequency components in speech are integrated over a bandwidth of at least $\frac{1}{3}$ oct for a hearing-impaired subject (even if the components are detected at the "wrong" place). For speech with an overall level of 65 dB SPL, the root-mean-square (rms) level in a $\frac{1}{3}$ oct band around 500 Hz is 58.8 dB SPL (ANSI, 1997). For a hearing loss of 80 dB at 500 Hz, the insertion gain recommended by the Cambridge formula is 30.4 dB. Hence the rms level of the $\frac{1}{3}$ band around 500 Hz would be 89.2 dB SPL after amplification, which is about 2 dB above the absolute threshold. Since the effective dynamic range of speech extends from about 15 dB above to 15 dB below the rms level (ANSI, 1997), this means that about 57% $\left(\frac{17}{30}\times 100\right)$ of the speech would have been audible in the band around 500 Hz. Calculations based on other amounts of hearing loss and other low frequencies indicate that the amplification provided would typically have been sufficient to provide audibility of 50% or more for the subjects with DRs. For the subjects without DRs the audibility at low frequencies was typically greater than 95%. Hence differences in audibility may partly account for why intelligibility was lower for the subjects with DRs than for those without DRs. However, audibility alone cannot account for the finding that, for the subjects with DRs, performance worsened when the high-pass filter cutoff frequency was decreased below about $0.57f_e$.

We used sequential information analysis (SINFA) (Wang and Bilger, 1973) to determine if the pattern of information about the phonetic features of voicing, manner, and place (Miller and Nicely, 1955) was affected by the high-pass filter cutoff frequency differently for the subjects with and without DRs. This was investigated using the stimulus-response matrices for subjects without DRs and for subjects with DRs with f_e close to 1000 Hz. The percentage of input information transmitted (IT) for a given feature was determined using the SINFA Analysis Suite "FIX" (Mike Johnson, Department of Phonetics and Linguistics, University College London; see www.phon.ucl.ac.uk.resource.htm). The SINFA analysis was conducted separately for each ear in the two groups, and then the results were averaged across the ears within each group.

The results for the two groups are shown in Fig. 11. For the subjects without any DR (bottom), the IT was highest for voicing and lowest for place. With increasing cutoff fre-

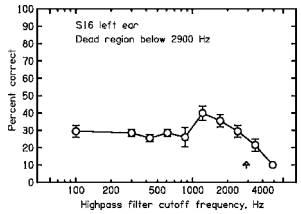


FIG. 10. Results of the VCV test for the left ear one subject with a low-frequency DR. Otherwise as in Fig. 7.

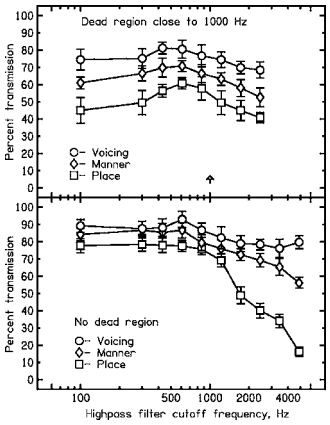


FIG. 11. Results of the SINFA analysis for subjects with DRs with f_e close to 1000 Hz (top) and for subjects without any DR (bottom). Error bars indicate ± one standard deviation (SD) across ears. See text for details.

quency, the IT for voicing declined only slightly, the IT for manner declined somewhat more, and the IT for place declined dramatically, consistent with the idea that IT for place depends on spectral information over a reasonably wide frequency range. For the subjects with DRs (bottom), the ordering of IT values for the different features was the same as for the subjects without DRs. However, for low cutoff frequencies, there was a decrease in transmission of all features, with IT for place decreasing the most. It appears that providing speech information at frequencies below about $0.57f_e$ impairs the use of spectral information. Possible reasons for this are discussed below.

IV. DISCUSSION

For the subjects with low-frequency DRs, speech intelligibility initially increased as the high-pass filter cutoff frequency was decreased from f_e to about $0.57f_e$. This is consistent with the conclusion reached from previous studies, that information from frequency components of speech falling within a low-frequency DR can be extracted via neural channels tuned to higher frequencies (Thornton and Abbas, 1980; Van Tasell and Turner, 1984).

The pattern of results found here is in some ways a mirror image of the pattern found by Vickers *et al.* (2001) and Baer *et al.* (2002). They tested subjects with *high-frequency* DRs as a function of the cutoff frequency of a *low-pass* filter. Like us, they found a benefit for speech intelligibility of amplifying frequency components that were somewhat inside the DR; optimum performance was usually obtained for a cutoff frequency of $1.5f_e$ to $2f_e$. However, in contrast to us, they did not usually find a worsening in performance when frequency components well within the DR were amplified; for most of their subjects, performance remained roughly constant when the cutoff frequency of the low-pass filter was above about $2f_e$. A worsening was found for only a few subjects, and it was not statistically significant. It is possible that Vickers *et al.* (2001) and Baer *et al.* (2002) failed to find a deleterious effect of amplifying frequency components well inside the DR because the high-frequency gain used in their studies was not sufficient to restore full audibility of the high-frequency components (Rankovic, 2002), although audibility was at least partially restored (Baer *et al.* 2002; Moore, 2002). In our study, the gain used would have been sufficient to provide audibility of 50% or more for the low-frequency components, as described earlier.

There are several theoretical reasons why people with DRs might extract little or no information from frequency components of speech that fall well within a DR, and why those components can sometimes have a deleterious effect. These reasons include the following:

(1) The frequency components are received through the "wrong" place in the cochlea. For example, if there is a low-frequency DR, amplified low-frequency components will be detected and analyzed via places in the cochlea that are tuned to higher frequencies. This mismatch between frequency and place may lead to difficulty in interpreting the information derived from the low frequencies, if the mismatch is sufficiently large. However, people given a cochlear implant experience such a mismatch, but they usually learn to understand speech via the implant and can even learn to compensate for an extra mismatch that is deliberately introduced (Dorman and Ketten, 2003). Similarly, simulations of cochlear-implant processing using normal-hearing listeners indicate that it is possible to compensate for a mismatch with extended training (Rosen *et al.*, 1999; Faulkner *et al.*, 2006). Eleven of the subjects in the group with DRs had previous experience listening via hearing aids that amplified low-frequency sound, so one might have expected that, for these subjects, any learning to compensate for the mismatch would already have taken place. Nevertheless, all subjects showed a deterioration in performance when the cutoff frequency was decreased below about $0.57f_e$. A possible explanation is that compensation for a frequency-place mismatch may be difficult or impossible when the mismatch is too large. Another possibility is that compensation may be more difficult when there is a frequency-place mismatch for some frequencies (those within the DR) but not for adjacent frequencies (those in the "live" region) (Baskent and Shannon, 2006).

(2) If the components falling in a DR are amplified sufficiently to make them audible, all of the components will be detected and analyzed via the same restricted region of the cochlea. This may lead to some form of masking or interference. For example, if there is a low-frequency DR, the lowest frequency components in broadband speech (below $0.57f_e$) may interfere with the analysis of frequency components between $0.57f_e$ and f_e; all of these components would be detected at the place tuned just above f_e. Clearly, some degree of overlap can be tolerated and even be beneficial, since the components between $0.57f_e$ and f_e led to improved intelligibility. However, too much information falling within a single "channel" may be a "bad thing," leading to a reduced ability to extract temporal information (Carlyon, 1996).

(3) Information in speech, such as information about formant frequencies, may partly be coded in the time patterns of the neural impulses (phase locking). The analy-

sis of temporal information may normally be done on a place-specific basis. For example, the neural machinery required to "decode" temporal information about frequencies around 1000 Hz may be restricted to neural channels with CFs close to 1000 Hz (Loeb *et al.*, 1983; Srulovicz and Goldstein, 1983; Moore and Carlyon, 2005). Studies of the perception of pitch by people with DRs are consistent with this idea (Huss and Moore, 2005a, b). When there is a mismatch between the frequencies of the speech components and the place where they are detected, the temporal decoding mechanisms required to analyze those speech components may not operate effectively.

Another possibility is connected with the fact that the level of the speech at low frequencies was higher for the subjects with DRs than for the subjects without DRs, since the gain used at low frequencies was greater for the former than for the latter. It is possible that, for the subjects with DRs, the high gain applied to frequency components below $0.57f_e$ led to upward spread of masking and hence reduced speech intelligibility. One way of assessing this possibility would be to use higher low-frequency gains for the subjects without DRs, although this might be difficult in practice, since the stimuli might then be unacceptably loud.

We consider now the implications of our results for the fitting of hearing aids to people with low-frequency DRs. As described in the Introduction, Halpin *et al.* (1994) suggested that hearing aids fitted to people with low-frequency DRs should be adjusted so that frequency components falling within the DR are not be amplified. If our results can be generalized to the perception of speech in everyday life, they lead to a somewhat different recommendation: frequency components with frequencies down to about $0.57f_e$ should be amplified, but components with frequencies lower than that should not be amplified. However, further research is needed before a firm recommendation can be made, as it is not known whether the pattern of results would be similar for sentence materials or under conditions where visual (speech reading) information was available.

Of course, application of this recommendation requires a test to determine whether or not a DR is present, and if so what the value of f_e is. The value of f_e can be estimated using the TEN(HL) test, but this gives only an approximate value. Also, the results of the TEN(HL) test can be inconclusive for people with severe or profound hearing loss, due to limitations in the levels that can be used, as found here and in previous work (Moore *et al.*, 2003; Markessis *et al.*, 2006; Vinay and Moore, 2007). Generally, such difficulties are encountered for frequencies where the hearing loss is 80 dB or more. Therefore, it may be better to check the diagnosis and to estimate f_e more precisely using PTCs. In clinical practice, the "fast" method with a sweeping masker may be useful for this purpose (Sek *et al.*, 2005; Kluk and Moore, 2006).

V. SUMMARY AND CONCLUSIONS

The identification of VCV syllables was measured as a function of high-pass filter cutoff frequency for subjects with low-frequency DRs and subjects with low-frequency hearing loss but without DRs. Initial diagnosis of DRs was based on the TEN(HL) test. The diagnosis was confirmed by measuring PTCs, which were also used to estimate the edge frequency of the DR, f_e, more precisely. There was a good correspondence between the results of the TEN(HL) test and the PTCs; for all cases where the TEN(HL)-test criteria for a DR were met, there was at least one PTC with a shifted tip. The VCV stimuli were amplified using the Cambridge formula to ensure audibility; the required frequency-gain characteristic was determined separately for each ear of each subject. For subjects with low-frequency hearing loss but without DRs, scores were high (about 78%) for low cutoff frequencies, remained approximately constant for cutoff frequencies up to 862 Hz, and then worsened with increasing cutoff frequency. For subjects with low-frequency DRs, performance was typically poor for the lowest cutoff frequency (100 Hz), improved as the cutoff frequency was increased to about $0.57f_e$, and worsened with further increases. The cutoff frequency leading to the best performance was highly correlated with the value of f_e. These results indicate that subjects with low-frequency DRs are able to make effective use of frequency components that fall in the range $0.57f_e$ to f_e, but that frequency components below $0.57f_e$ have deleterious effects. It is tentatively recommended that, when fitting hearing aids to people with low-frequency DRs, gain should be applied for frequencies down to about $0.57f_e$, but gain should not be applied for frequencies lower than that.

ACKNOWLEDGMENTS

This work of the second author was supported by the Medical Research Council (UK). We thank Tom Baer for designing the digital filters used in this study; Brian Glasberg for help with statistical analysis; Tom Baer, Karolina Kluk, Ken Grant, Chris Turner, and an anonymous reviewer for helpful comments on earlier versions of this paper; and the Director, All India Institute of Speech and Hearing, for providing the facilities to carry out this work.

Aazh, H., and Moore, B. C. J. (**2007**). "Dead Regions in the cochlea at 4 kHz in elderly adults: Relation to absolute threshold, steepness of audiogram, and pure tone average," J. Am. Acad. Audiol **18**, 97–106.

Amos, N. E., and Humes, L. E. (**2001**). "The contribution of high frequencies to speech recognition in sensorineural hearing loss," in *Physiological and Psychophysical Bases of Auditory Function*, edited by D. J. Breebaart, A. J. M. Houtsma, A. Kohlrausch, V. F. Prijs, and R. Schoonhoven (Shaker, Maastricht).

ANSI (**1969**). "ANSI S3.5. Methods for the calculation of the articulation index," (American National Standards Institute, New York).

ANSI (**1997**). "ANSI S3.5-1997, Methods for the calculation of the speech intelligibility index," (American National Standards Institute, New York).

ANSI (**2004**). "ANSI S3.6-2004 Specification for audiometers," (American National Standards Institute, New York).

Baer, T., Moore, B. C. J., and Kluk, K. (**2002**). "Effects of lowpass filtering on the intelligibility of speech in noise for people with and without dead regions at high frequencies," J. Acoust. Soc. Am. **112**, 1133–1144.

Baskent, D., and Shannon, R. V. (**2006**). "Frequency transposition around dead regions simulated with a noiseband vocoder," J. Acoust. Soc. Am. **119**, 1156–1163.

Borg, E., Canlon, B., and Engström, B. (**1995**). "Noise-induced hearing loss-Literature review and experiments in rabbits. Morphological and electrophysiological features, exposure parameters and temporal factors, variability and interactions," Scand. Audiol. **24**(Suppl. 40), 1–147.

Burkhard, M. D., and Sachs, R. M. (**1975**). "Anthropometric manikin for acoustic research," J. Acoust. Soc. Am. **58**, 214–222.

Carhart, R., and Jerger, J. F. (**1959**). "Preferred method for clinical determination of pure-tone thresholds," J. Speech Hear Disord. **24**, 330–345.
Carlyon, R. P. (**1996**). "Encoding the fundamental frequency of a complex tone in the presence of a spectrally overlapping masker," J. Acoust. Soc. Am. **99**, 517–524.
Ching, T., Dillon, H., and Byrne, D. (**1998**). "Speech recognition of hearing-impaired listeners: Predictions from audibility and the limited role of high-frequency amplification," J. Acoust. Soc. Am. **103**, 1128–1140.
Dorman, M. F., and Ketten, D. (**2003**). "Adaptation by a cochlear-implant patient to upward shifts in the frequency representation of speech," Ear Hear. **24**, 457–460.
Engström, B. (**1983**). "Stereocilia of sensory cells in normal and hearing impaired ears," Scand. Audiol. Suppl. **19**, 1–34.
Faulkner, A., Rosen, S., and Norman, C. (**2006**). "The right information may matter more than frequency-place alignment: simulations of frequency-aligned and upward shifting cochlear implant processors for a shallow electrode array insertion," Ear Hear. **27**, 139–152.
Florentine, M., and Houtsma, A. J. M. (**1983**). "Tuning curves and pitch matches in a listener with a unilateral, low-frequency hearing loss," J. Acoust. Soc. Am. **73**, 961–965.
Glasberg, B. R., and Moore, B. C. J. (**1990**). "Derivation of auditory filter shapes from notched-noise data," Hear. Res. **47**, 103–138.
Gravendeel, D. W., and Plomp, R. (**1960**). "Perceptive bass deafness," Acta Oto-Laryngol. **51**, 549–560.
Greenwood, D. D. (**1961**). "Critical bandwidth and the frequency coordinates of the basilar membrane," J. Acoust. Soc. Am. **33**, 1344–1356.
Greenwood, D. D. (**1990**). "A cochlear frequency-position function for several species—29 years later," J. Acoust. Soc. Am. **87**, 2592–2605.
Halpin, C., Thornton, A., and Hasso, M. (**1994**). "Low-frequency sensorineural loss: Clinical evaluation and implications for hearing aid fitting," Ear Hear. **15**, 71–81.
Hogan, C. A., and Turner, C. W. (**1998**). "High-frequency audibility: Benefits for hearing-impaired listeners," J. Acoust. Soc. Am. **104**, 432–441.
Huss, M., and Moore, B. C. J. (**2003**). "Tone decay for hearing-impaired listeners with and without dead regions in the cochlea," J. Acoust. Soc. Am. **114**, 3283–3294.
Huss, M., and Moore, B. C. J. (**2005a**). "Dead regions and noisiness of pure tones," Int. J. Audiol. **44**, 599–611.
Huss, M., and Moore, B. C. J. (**2005b**). "Dead regions and pitch perception," J. Acoust. Soc. Am. **117**, 3841–3852.
Kluk, K., and Moore, B. C. J. (**2004**). "Factors affecting psychophysical tuning curves for normally hearing subjects," Hear. Res. **194**, 118–134.
Kluk, K., and Moore, B. C. J. (**2005**). "Factors affecting psychophysical tuning curves for hearing-impaired subjects," Hear. Res. **200**, 115–131.
Kluk, K., and Moore, B. C. J. (**2006**). "Detecting dead regions using psychophysical tuning curves: a comparison of simultaneous and forward masking," Int. J. Audiol. **45**, 463–476.
Langenbeck, B. (**1965**). *Textbook of Practical Audiometry* (Edward Arnold, London).
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.
Mackersie, C. L., Crocker, T. L., and Davis, R. A. (**2004**). "Limiting high-frequency hearing aid gain in listeners with and without suspected cochlear dead regions," J. Am. Acad. Audiol **15**, 498–507.
Markessis, E., Kapadia, S., Munro, K. J., and Moore, B. C. J. (**2006**). "Modification of the TEN test for cochlear dead regions for use with steeply sloping high-frequency hearing loss," Int. J. Audiol. **45**, 91–98.
Miller, G. A., and Nicely, P. E. (**1955**). "An analysis of perceptual confusions among some English consonants," J. Acoust. Soc. Am. **27**, 338–352.
Moore, B. C. J. (**1986**). "Parallels between frequency selectivity measured psychophysically and in cochlear mechanics," Scand. Audiol. Suppl. **25**, 139–152.
Moore, B. C. J. (**2001**). "Dead regions in the cochlea: Diagnosis, perceptual consequences, and implications for the fitting of hearing aids," Trends Amplif. **5**, 1–34.
Moore, B. C. J. (**2002**). "Response to 'Articulation index predictions for hearing-impaired listeners with and without cochlear dead regions' [J. Acoust. Soc. Am. **111**, 2545–2548 (2002)] (L)," J. Acoust. Soc. Am. **111**, 2549–2550.
Moore, B. C. J. (**2003**). *An Introduction to the Psychology of Hearing*, 5th ed. (Academic, San Diego).
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 Carlyon, R. P. (**2005**). "Perception of pitch by people with cochlear hearing loss and by cochlear implant users," in *Pitch Perception*, edited by C. J. Plack, A. J. Oxenham, R. R. Fay, and A. N. Popper (Springer, New York).
Moore, B. C. J., and Glasberg, B. R. (**1997**). "A model of loudness perception applied to cochlear hearing loss," Aud. Neurosci. **3**, 289–311.
Moore, B. C. J., and Glasberg, B. R. (**1998**). "Use of a loudness model for hearing aid fitting. I. Linear hearing aids," Br. J. Audiol. **32**, 317–335.
Moore, B. C. J., Glasberg, B. R., and Baer, T. (**1997**). "A model for the prediction of thresholds, loudness and partial loudness," J. Audio Eng. Soc. **45**, 224–240.
Moore, B. C. J., Glasberg, B. R., and Stone, M. A. (**2004**). "New version of the TEN test with calibrations in dB HL," Ear Hear. **25**, 478–487.
Moore, B. C. J., Killen, T., and Munro, K. J. (**2003**). "Application of the TEN test to hearing-impaired teenagers with severe to profound hearing loss," Int. J. Audiol. **42**, 465–474.
Moore, B. C. J., Laurence, R. F., and Wright, D. (**1985**). "Improvements in speech intelligibility in quiet and in noise produced by two channel compression hearing aids," Br. J. Audiol. **19**, 175–187.
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.
Murray, N., and Byrne, D. (**1986**). "Performance of hearing-impaired and normal hearing listeners with various high-frequency cut-offs in hearing aids," Aust. J. Audiol. **8**, 21–28.
Rankovic, C. M. (**2002**). "Articulation index predictions for hearing-impaired listeners with and without cochlear dead regions (L)," J. Acoust. Soc. Am. **111**, 2545–2548.
Rosen, S., Faulkner, A., and Wilkinson, L. (**1999**). "Adaptation by normal listeners to upward spectral shifts of speech: implications for cochlear implants," J. Acoust. Soc. Am. **106**, 3629–36.
Schuknecht, H. F. (**1993**). *Pathology of the Ear*, 2nd ed. (Lea and Febiger, Philadelphia).
Schuknecht, H. F., and Gacek, M. R. (**1993**). "Cochlear pathology in presbycusis," Ann. Otol. Rhinol. Laryngol. **102**, Suppl. 158, 1–16.
Sek, A., Alcántara, J. I., Moore, B. C. J., Kluk, K., and Wicher, A. (**2005**). "Development of a fast method for determining psychophysical tuning curves," Int. J. Audiol. **44**, 408–420.
Srulovicz, P., and Goldstein, J. L. (**1983**). "A central spectrum model: a synthesis of auditory-nerve timing and place cues in monaural communication of frequency spectrum," J. Acoust. Soc. Am. **73**, 1266–1276.
Summers, V., Molis, M. R., Musch, H., Walden, B. E., Surr, R. K., and Cord, M. T. (**2003**). "Identifying dead regions in the cochlea: psychophysical tuning curves and tone detection in threshold-equalizing noise," Ear Hear. **24**, 133–142.
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.
Troland, L. T. (**1929**). "The psychophysiology of auditory qualities and attributes," J. Gen. Psychol. **2**, 28–58.
Turner, C. W., and Cummings, K. J. (**1999**). "Speech audibility for listeners with high-frequency hearing loss," Am. J. Audiol. **8**, 47–56.
Turner, C. W., Burns, E. M., and Nelson, D. A. (**1983**). "Pure tone pitch perception and low-frequency hearing loss," J. Acoust. Soc. Am. **73**, 966–975.
Van Tasell, D. J., and Turner, C. W. (**1984**). "Speech recognition in a special case of low-frequency hearing loss," J. Acoust. Soc. Am. **75**, 1207–1212.
Vestergaard, M. (**2003**). "Dead regions in the cochlea: implications for speech recognition and applicability of articulation index theory," Int. J. Audiol. **42**, 249–261.
Vickers, D. A., Moore, B. C. J., and Baer, T. (**2001**). "Effects of lowpass filtering on the intelligibility of speech in quiet for people with and without dead regions at high frequencies," J. Acoust. Soc. Am. **110**, 1164–1175.
Vinay, and Moore, B. C. J. (**2007**). "Prevalence of dead regions in subjects with sensorineural hearing loss," Ear Hear. **28**, 231–241.
Wang, M. D., and Bilger, R. C. (**1973**). "Consonant confusions in noise: A study of perceptual features," J. Acoust. Soc. Am. **54**, 1248–1266.

Sentential, lexical, and acoustic effects on the perception of word boundaries

Sven L. Mattys[a] and James F. Melhorn
Department of Experimental Psychology, University of Bristol, 12A Priory Road, Bristol, Avon BS8 1TU, United Kingdom

(Received 26 May 2006; revised 2 April 2007; accepted 3 April 2007)

This study investigates the effects of sentential context, lexical knowledge, and acoustic cues on the segmentation of connected speech. Listeners heard near-homophonous phrases (e.g., /plʌmpaɪ/ for "plum pie" versus "plump eye") in isolation, in a sentential context, or in a lexically biasing context. The sentential context and the acoustic cues were piloted to provide strong versus mild support for one segmentation alternative (plum pie) or the other (plump eye). The lexically biasing context favored one segmentation or the other (e.g., /skʌmpaɪ/ for "scum pie" versus *"scump eye," and /lʌmpaɪ/, for "lump eye" versus *"lum pie," with the asterisk denoting a lexically unacceptable parse). A forced-choice task, in which listeners indicated which of two words they thought they heard (e.g., "pie" or "eye"), revealed compensatory mechanisms between the sources of information. The effect of both sentential and lexical contexts on segmentation responses was larger when the acoustic cues were mild than when they were strong. Moreover, lexical effects were accompanied with a reduction in sensitivity to the acoustic cues. Sentential context only affected the listeners' response criterion. The results highlight the graded, interactive, and flexible nature of multicue segmentation, as well as functional differences between sentential and lexical contributions to this process. © *2007 Acoustical Society of America.* [DOI: 10.1121/1.2735105]

PACS number(s): 43.71.Sy [KWG] Pages: 554–567

I. INTRODUCTION

In an attempt to account for the humans' effortless recognition of connected speech, researchers have provided two strands of evidence for the perceptual and cognitive mechanisms involved in speech segmentation. The first one suggests that segmentation is carried out "bottom-up," based on the listeners' sensitivity to sublexical cues co-occurring with word boundaries, e.g., acoustic/allophonic variants (e.g., Church, 1987; Davis, Marslen-Wilson, and Gaskell, 2002; Mattys, 2004; Nakatani and Dukes, 1977; Umeda and Coker, 1974), metrical stress (e.g., Cutler and Butterfield, 1992; Cutler and Norris, 1988), and phonotactic regularities (e.g., McQueen, 1998; Vitevitch and Luce, 1999). The other strand considers segmentation a product of "top-down" expectations arising from word knowledge and higher-order inferences. Here, it is the identification of individual words through contextual information and lexical competition that allows word boundaries to be located (e.g., McClelland and Elman, 1986; Norris, 1994; see also Brent, 1999; Dahan and Brent, 1999).

We recently proposed that the communicative goal of speech acts in general implies that higher-order information outweighs sublexical cues when the two sources of information are in conflict (Mattys, White, and Melhorn, 2005). On this view, a sentential context such as "The baker looked at the drawing of a…" favors the segmentation of subsequent /plʌmpaɪ/ as "plum pie" over "plump eye," even when the phonetic realization of the phrase suggests otherwise. Similarly, listeners should prefer segmenting the phrase /skʌmpaɪ/ as "scum pie" even if its realization is more consistent with "skump eye," because the latter parse leaves a portion of the utterance lexically unaccounted for (skump).

However, defining segmentation cues as favorable *versus* unfavorable for a particular parse is a simplification of the listening conditions in which speech is usually experienced. In particular, the extent to which acoustic-phonetic cues such as aspiration, glottalization, lengthening, etc., are present at word junctures is shown to vary widely across speakers and conversational circumstances (e.g., Lindblom, 1990; Redi and Shattuck-Hufnagel, 2001). For example, Dilley, Shattuck-Hufnagel, and Ostendorf (1996), found rates of word-onset vocalic glottalization (e.g., /hɑrdʔækt/ for "hard act") to range from 13% to 44% among five professional radio announcers—leaving aside qualitative and quantitative differences in realizing the glottal stop (Redi and Shattuck-Hufnagel, 2001). Likewise, articulatory precision is shown to be reduced when contextual information is available [e.g., "nine" is less clearly articulated in "A stitch in time saves nine" than in "The number which you are about to hear is nine," (Lieberman, 1963; see also Hawkins and Warren, 1994)].

From a perceptual point of view, fine acoustic details have been shown to affect boundary processing (e.g., Christophe, Peperkamp, Pallier, Block, and Mehler, 2004; Davis *et al.*, 2002; Salverda, Dahan, and McQueen, 2003), especially when contextual information is uninformative (Hawkins and Smith, 2001; Mattys *et al.*, 2005). However, the converse interaction is unknown, i.e., whether the strength of word-boundary acoustic cues affects the extent to which listeners rely on sentential and lexical information for segmentation.

[a]Author to whom correspondence should be addressed. Electronic mail: Sven.Mattys@bris.ac.uk

 0001-4966/2007/122(1)/554/14/$23.00

A strict account of sentential/lexical dominance over sublexical cues would predict that listeners' reliance on higher-order information should be the same irrespective of the congruency and strength of the acoustic cues. A more graded account of cue combination, however, would have provisions for trade-offs between sentential/lexical expectations and acoustic cues. Trading relations have been amply documented for the perception of phonetic contrasts (e.g., /r/–/l/), where decreasing the strength of one cue (e.g., $F2/F3$ spectral configuration) can be offset by increasing the strength of another (e.g., $F1$ transition duration, e.g., Polka and Strange, 1985). Whether perceptual trade-offs or phenomena akin to them occur for the purpose of speech segmentation and across such contrasted sources of information as sentential/lexical and acoustic cues is a question for this study.

Another question concerns the nature of the interaction between higher-order information and sublexical cues. Does the presence of sentential/lexical information cause a reduction in sensitivity to acoustic word-boundary cues, or solely a strategic bias toward the sententially or lexically appropriate parsing? Research on the effects of sentential and lexical expectations on phoneme perception suggests that lexical knowledge affects the processing of the acoustic information itself whereas sentential context does not (Connine, 1987; Connine and Clifton, 1987; Samuel, 1981, 1997, 2001; but see Norris, McQueen, and Cutler, 2000). Although there are no a priori reasons to believe that different contingencies should apply to the perception of acoustic cues at word boundaries, the elusive nature of word boundaries—which are after all nothing more than junctures between perceptual objects—could make their processing quite distinct from that of word-internal phonemes. On the one hand, the fact that segmentation sometimes necessitates the integration of relatively large portions of an utterance suggests that acoustic perception should remain flexible while integration is ongoing to prevent the incorrect parses momentarily supported by higher-order information from weighing too heavily in the segmentation process. On the other hand, because the ultimate goal of spoken communication is to convey meaning rather than sublexical information *per se*, allowing higher-order information to alter the low-level analysis of conflicting acoustic cues might be an efficient way of guarding against segmentation errors.

The present study consisted of two experiments designed to assess the effect of sentential information (experiment 1) and lexical knowledge (experiment 2) on the perception of word boundaries, with the strength of the acoustic information manipulated. In experiment 1, listeners indicated which of two words (e.g., "pie" or "eye") they heard at the end of a sentence. The sentences ended with one of four short phrases piloted to have strong or mild acoustic cues favoring either one parse (e.g., "plum pie") or the other (e.g., "plump eye"). The portion of the sentence preceding the critical phrase was also piloted to provide a strong or mild semantic context for one parse or the other (see Table I for an example). Of interest was how sentential and acoustic sources of information varying in congruency and strength interacted to guide the listeners' segmentation responses. In experiment 2, we used the same task to investigate the role of lexical knowledge on word-boundary perception. Here, the beginning of the four test phrases used in experiment 1 was edited, via cross splicing, to introduce a lexical bias for one segmentation solution ("scum pie") or the other ("lump eye")—the alternative segmentation solution led to a stranded nonword (*"scump eye," *"lum pie"). In line with experiment 1, the question was whether lexically inferable segmentation affects the processing of acoustic cues for word boundaries, and whether this effect trades off with the strength of the acoustic cues. Moreover, in an attempt to find out whether sentential and lexical contexts result in a modification of the perception of the signal or in a simple response bias, the sensitivity index d' and bias index c, from signal detection theory (Green and Swets, 1996; Tanner and Swets, 1954), were calculated as well. Note that these experiments were not intended to speak to the dominance of sentential/lexical knowledge over acoustic cues *per se*, but rather to focus on acoustic processing and its modulation by sentential/lexical information. To provide a stringent test for this hypothesis, the task therefore emphasized bottom-up rather than top-down evidence.

TABLE I. Examples of stimuli in experiments 1 and 2.

Experiment 1: Context versus acoustics [a]

Context	
A[b] strong	The baker looked at the drawing of a…
A mild	The girl looked at the drawing of a…
B mild	The scientists looked at the drawing of a…
B strong	The surgeon looked at the drawing of a…
Acoustics	
A strong	…plum##pie[c]
A mild	…plum#pie
B mild	…plump#eye
B strong	…plump##eye

Experiment 2: Lexicality versus acoustics[d]

Acoustics	*A* lexicality (from "scum pie")	*B* lexicality (from "lump eye")	Neutral
A strong	**sc**um##pie	[e]***l**um##pie	plum##pie
A mild	**sc**um#pie	***l**um#pie	plum#pie
B mild	***sc**ump#eye	**l**ump#eye	plump#eye
B strong	***sc**ump##eye	**l**ump##eye	plump##eye

[a]Utterances were created by appending one of the acoustic phrases to the end of one of the context sentences in a 4×4 factorial design.
[b]In this stimulus set, *A* favors the "pie" alternative whereas *B* favors the "eye" alternative.
[c]## denotes strong acoustic word-boundary cues and # denotes mild ones.
[d]Utterances were created by splicing the underlined portion of the lexical phrases with the initial portion of the acoustic phrases (in bold) in a 2×4 factorial design. See Sec. IV A 2 for details on the cross-splicing procedure.
[e]Asterisks denote phrases with a lexically disfavored parsing.

II. STIMULUS PILOTING AND SELECTION

The goal of the pilot experiment was to create acoustic phrases and contextual sentences with calibrated and comparable segmentation biases when presented in isolation. A large number of renditions of near-homophonous phrases (e.g., /plʌmpaɪ/) were recorded and rated by listeners, of which we kept only four: one leading to a strong preference

for one parse (e.g., "plum pie"), one to a mild preference for that same parse, and the same for the alternative parse (e.g., "plump eye"). Similarly, we recorded a large number of sentences whose ending was more or less consistent with one version of the phrase or the other. Following the rating procedure used with the phrases, only those sentences leading to a strong versus mild preference for one phrase over the other were kept.

A. Acoustic phrases

We chose sixteen near-homophonous two-word-long pairs of phrases differing in the location of their word boundary, but of similar syntactic structure (e.g., "plum pie" versus "plump eye"). Some of the pairs were borrowed from other studies (e.g., Gimson, 1980; Turk and Sawusch, 1997; Turk and White, 1999). A male native speaker of standard British English was asked to produce ten renditions of each phrase, varying his pronunciation to cover a continuum ranging from one segmentation alternative (e.g., "plum pie") to the other (e.g., "plump eye"). As the speaker was asked to adopt an intonation contour he felt comfortable and keep it constant across all renditions, all phrases ended up receiving pitch accent on the second word. Because the speaker was instructed not to use silent pauses to mark word junctures, differences among renditions involved primarily acoustic-phonetic variations of the phonemes at word junctures. Although a detailed description of such variations could be provided on a stimulus-by-stimulus basis, it would not be particularly relevant here since we are more concerned with the quantitative and tractable strength of the bottom-up information, broadly construed, than with the qualitative nature of the cues. All stimuli were recorded in a sound-attenuated room and digitized (16 bit A/D) at 32 kHz. On output, the utterances were converted to analog form (16 bit D/A, 32 kHz) and delivered over good-quality headphones at an average sound level of 65 dB.

All renditions ($16\times10=160$) were pseudorandomized, with at least seven phrases between two versions of the same phrase, and played to 27 native English speakers who were asked to rate each rendition on an 11-point scale ranging from one segmentation alternative to the other. All participants were native British English undergraduate or graduate students at the University of Bristol. They received course credit or a small honorarium for taking part in the experiment. None reported a history of speech or hearing difficulties.

On each trial, the participants heard a phrase, then saw the two segmentation alternatives written on a computer monitor, separated by dots, for 2 s. Their task was to rate the rendition on that continuum by pressing one of eleven keys on a keyboard. The left versus right location of the two alternatives on the array of keys was randomly chosen for each trial. Participants had up to ten seconds to rate each phrase. The experiment began with a practice block of five phrases unused in the main block.

Ratings from 10 to 0 were converted to percentage values from 100 to 0 in 10% decrements. A value of 100% was given arbitrarily to the alternative with the longer final word (e.g., "plum pie," labeled alternative *A*) and a value of 0% to the alternative with the shorter final word (e.g., "plump eye," labeled alternative *B*). Ratings for each of the ten renditions of each phrase were averaged across the 27 participants. Of the 16 initial sets of phrases, eight were kept (see Appendix A). These were phrases for which we could find four renditions falling into categories defined as, respectively, strong *A* (range of 95%–80%, $M=86\%$), mild *A* (75%–55%, $M=68\%$), mild *B* (45%–25%, $M=32\%$), and strong *B* ($20\%-5\%, M=14\%$). The choice was also constrained by the ratings obtained on sentential contexts, as described in the next section. Note that the two strong categories of acoustic phrases were chosen purposely not to include any renditions yielding unanimous parsing (i.e., 100% or 0%). This ensured that phrases with idiosyncratically extreme cues or pauses were not considered in the study. The Celex lexical frequency of the final word (Baayen, Pipenbrock, and Gulikers, 1995) was comparable in the *A* (e.g., "pie") and *B* (e.g., "eye") conditions, 855 versus 740, $t(6)=1.53$, $p=0.88$ ("day" and "a" were not entered in the calculation because of the extremely high frequency of "a" and the difficulty in establishing the frequency of its realization as /eɪ/ as opposed to /ə ɪ/).

B. Sentential contexts

The selection of the sentences followed the rationale and procedure used for that of the acoustic phrases. For each of the 16 original near-homophonous phrases, ten sentences were created, which provided a semantic context varying in strength for one phrase alternative or the other, when the phrase was placed in sentence-final position. Most sentences within a set had a similar syntactic structure and, where possible, only differed by their first few words. All sentences were recorded by the speaker who recorded the acoustic phrases, using the same apparatus. The speaker was asked to produce each sentence with its semantically appropriate phrase at the end, e.g., "The baker looked at the drawing of a plum pie." The sentence-final phrase was then deleted, e.g., "The baker looked at the drawing of a." Following the procedure used with the acoustic phrases, all truncated sentences ($16\times10=160$) were pseudorandomized, with at least seven intervening sentences between sentences from the same set, and played to the same 27 participants as before. On each trial, the participants were asked to rate which of two phrase alternatives (e.g., "plum pie" versus "plump eye") was a more likely completion for the truncated sentence, using the 11-point scale described above. The rest of the rating procedure was the same as before.

Conversion to percentage values and a selection of eight among the 16 sentence sets followed the criteria used to select the acoustic phrases. The outcome of the selection, which was done concurrently with that of the acoustic phrases, was four sentences per set, differing in their completion likelihood: strong $A(M=87\%)$, mild $A(M=71\%)$, mild $B(M=31\%)$, and strong $B(M=14\%)$. Sentences and their ratings can be seen in Appendix A (see also Table I for an example).

III. EXPERIMENT 1

The goal of this experiment was to measure the influence of sentential context on the perception of word boundaries. The calibrated sentences and phrases were concatenated orthogonally. This resulted in 16 utterances (4×4) for each of the eight sets, with various degrees of consistency and inconsistency between context and acoustics. For each utterance, listeners were asked to decide which of two words they thought they heard in sentence-final position (e.g., "pie" or "eye"). This task was chosen instead of the more explicit task used in the pilot experiment (e.g., "plum pie" or "plump eye") in order not to highlight the critical manipulation and to allow a direct comparison with experiment 2.

A. Methods

1. Participants

Thirty native speakers of British English received course credit or a small honorarium for taking part in the experiment. None reported a history of speech or hearing difficulties. These participants were different from those used in the pilot experiment.

2. Materials

The four acoustic phrases of a set were appended at the end of the four sentences of the same set in an orthogonal fashion, creating 16 sentences per set ($\times 8$ sets $=128$ utterances). Although the concatenation of a sentence and a phrase resulted in a brief pause between the truncated sentence and the phrase, it did not interfere with the natural flow of the sentence and did not create acoustic artifacts. In order to encourage listeners to maintain a high level of attention throughout each sentence—hence giving them the opportunity to also process its semantic content, rather than focus exclusively on the acoustics of their ending—we added 256 filler sentences. All filler sentences were approximately the same length as the test sentences, but 128 of them were truncated in an early position (e.g., "A huge sun" truncated from "A huge sun filled the clear sky"), whereas 128 were truncated in a middle position (e.g., "It wasn't your right" truncated from "It wasn't your right of way"). Within these two filler categories, half the truncated sentences ended with a potentially ambiguous juncture (e.g., "Getting the stray tissue"), and semantic consistency between the sentence and the final word varied. The other half of the truncated sentences ended with a nonambiguous juncture (e.g., "It was discussed"). The filler sentences did not include any of the test materials. As in the test trials, two possible utterance-final words were presented at the offset of the truncated sentence (e.g., "sun" versus "son"; "tissue" versus "'issue"; "discussed" versus "disgust").

3. Procedure

The filler stimuli were recorded by the speaker who had recorded the test stimuli using the same apparatus. Test and filler utterances (384 in all) were pseudorandomized, with at least twenty sentences between two sentences from the same test set. On each trial, participants heard an utterance, followed by the presentation of two word alternatives displayed side by side on a computer monitor, 400 ms after the end of the utterance. Participants had up to 10 s to decide which of the two words they heard at the end of the utterance, using one of two keys on a computer keyboard. In an attempt to encourage the listeners to pay attention to the fine acoustic details of the utterances, they were told that the difference between the two alternatives was often very subtle and that they should respond based on what they really heard rather than what they thought should have been the last word of the utterance. Thus, the task was signal-oriented in nature. Participants were also told that the sentences may or may not be complete. The left versus right location of the two alternatives was randomly chosen for each trial. A 1 s interval followed the key press before the onset of the next utterance. The experiment began with a practice block of twelve trials unused in the main block.

B. Results

1. Segmentation accuracy

Responses averaged across the eight stimulus sets and 30 participants are displayed in Fig. 1(a). For the sake of simplicity, and because all conditions were symmetrical in design, the results were collapsed across the *A* and *B* acoustic conditions [Fig. 1(b)].[1] To do so, a response was counted as correct if it was consistent with the acoustic word boundary. As an example, accuracy in the "strong acoustics-mild congruent context" condition was the average between the percentage of *A* responses in the "strong *A* acoustics-mild *A* context" condition and the percentage of *B* responses in the "strong *B* acoustics-mild *B* context."

An analysis of variance was performed on the participants' accuracy data, with acoustic strength (strength versus mild), contextual congruency (congruent versus incongruent), and contextual strength (strong versus mild) as repeated measures. Both acoustic strength and contextual congruency significantly affected performance. Segmentation accuracy was higher when the acoustic cues were strong compared to mild, $F(1,29)=209.50$, $p<0.001$, and when the sentential context aligned with the acoustic cues than when it conflicted with them, $F(1,29)=21.92$, $p<0.001$. Thus, sentential context had a clear modulatory effect on the use of acoustic cues for word boundaries. However, an interaction between acoustic strength and contextual congruency, $F(1,29)=14.78$, $p=0.001$, indicated that this modulation was greater when the acoustic cues were mild than when they were strong (all pairwise comparisons reached $p \leq 0.005$).

The possibility that the interaction was due to a ceiling effect was examined in a median-split analysis. Participants were rank ordered on their average accuracy in the congruent, strong-acoustics condition (with contextual strength collapsed) and split into high-accuracy and low-accuracy subgroups ($N=15$ in each subgroup). We then reran the main analysis of variance with accuracy (high versus low) as an additional variable. The interaction between acoustic strength, contextual congruency, and accuracy was not significant, $F(1,28)<1$. Furthermore, an analysis of variance

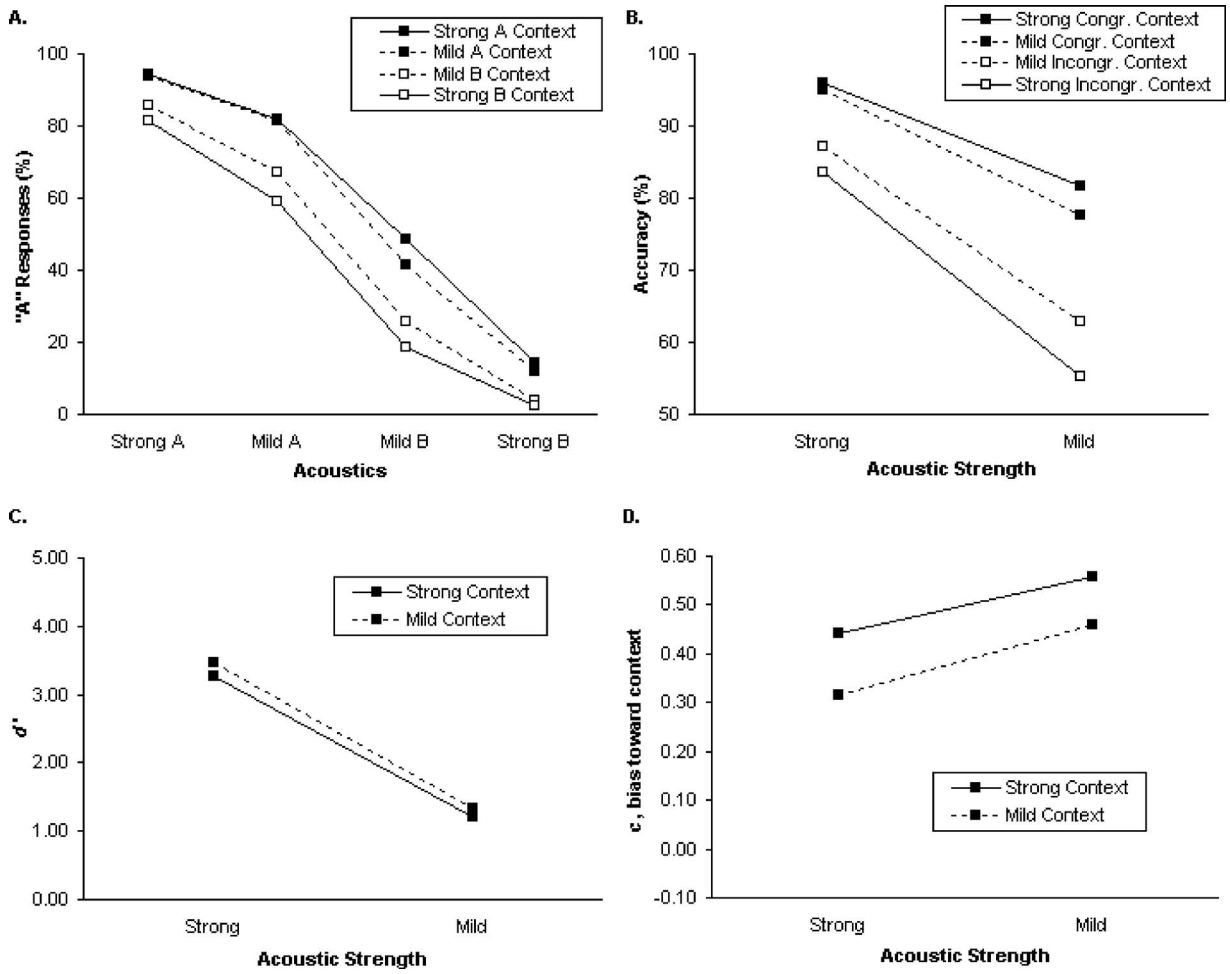


FIG. 1. Experiment 1. (a) Average percentage of sentence-final *A* responses (e.g., pie) as a function of the strength of the acoustic cues and the strength of the sentential context. (b) Same as (a) when symmetrical *A* and *B* acoustic conditions are collapsed. Accuracy is measured by reference to the intended acoustic juncture, (c) d', discriminability between *A* and *B* acoustic alternatives. (d) *c*, response criterion expressed as a bias toward the contextually appropriate segmentation.

focusing on the low-accuracy group showed the same patterns of results as those for the entire group: acoustic strength, $F(1,14)=67.30$, $p<0.001$; contextual congruency, $F(1,14)=9.79$, $p<0.01$; interaction, $F(1,14)=6.99$, $p<0.02$. Thus, there was no evidence that the reduced effect of context in the strong-acoustics condition was the result of a ceiling artifact.

The main effect of contextual strength was not directly interpretable from the main analysis of variance because the strong and mild levels had opposite polarity in the congruent and incongruent conditions. To test whether contextual strength had an effect on segmentation accuracy, strong and mild levels were aligned in the congruent and incongruent conditions by reversing the mild versus strong polarity of the incongruent conditions. An analysis of variance run on these modified labels showed an effect of contextual strength, $F(1,29)=10.91$, $p=0.003$, suggesting that segmentation accuracy was influenced by, and commensurate with the strength of the sentential context. This effect did not significantly interact with the strength of the acoustic cues, $F(1,29)=3.23$, $p=0.08$, or with contextual congruency, $F(1,29)=2.65$, $p=0.11$. None of the other effects or interactions reached the 0.05 significance level. Finally, when each mean value was compared to the 50% chance level, two-tail $t(29)$ yielded $p<0.001$ in all conditions except for the "mild acoustics-strong incongruent context," which did not reach 0.05 significance.

In sum, the sentential context had a modulatory role on the listeners' segmentation responses, with incongruence between context and acoustics costing an average 15% in accuracy. Additionally, this effect was more substantial when the acoustic cues were mild (21%) than when they were strong (10%), which indicates that trade-off relations operate between high-level and low-level cues, with contextual information more influential when the acoustic cues are less reliable.

Importantly, contextual strength, too, was found to influence segmentation accuracy, with greater modulation from a strong than a mild sentential context. In an attempt to find out whether this effect came about solely as a response bias for the appropriate context or whether it was accompanied with a change in sensitivity to the acoustic cues, we analyzed segmentation performance using the discriminability index d' and the bias index *c* from signal detection theory (Green and Swets, 1996; Tanner and Swets, 1954).

2. Juncture discriminability

Listeners could have used contextual information to simply shift their response criterion in the direction of the context without altering their underlying perception of the acoustic information. If so, we would expect discriminability between the *A* and *B* alternatives (e.g., plum#pie versus plump#eye) to have remained the same whether the phrase was preceded by a strong or a mild context. Differences in contextual strength would have arisen as simple proportional biases toward the context, at no cost for the acoustic analysis of the signal itself.

Alternatively, the presence of contextual information could have caused a genuine reduction in reliance on lower-level, acoustic information—i.e., more superficial phonetic analysis. In this case, we would expect discriminability between the *A* and *B* alternatives to be lower when the sentential context is strong than when it is mild. This would be predicted by a segmentation model in which cues trade weights at a deep level and in a hierarchically predictable manner.

To assess discriminability, we computed d' scores for each participant and for each of the four cells of the following design: acoustic strength (strong versus mild) by contextual strength (strong versus mild). Averages are shown in Fig. 1(c) d', which is often thought to be independent of response criteria (but see Norris, 1995, and Pylyshyn, 1999, for a discussion on this point), was calculated as z(hit rate) $-z$(false-alarm rate), with the *A* alternative chosen arbitrarily as the reference point. As an example, d' for the "strong acoustics-mild context" condition was the average of d' in the "strong acoustics-mild *A* context" and "strong acoustics-mild *B* context" conditions. The hit rate of the "strong acoustic-mild *A* context" was the percentage of *A* responses in the "strong *A* acoustics-mild *A* context" condition and the false-alarm rate was the percentage of *A* responses in the "strong *B* acoustics-mild *A* context" condition. A comparable compilation was done to calculate d' in the "strong acoustics-mild *B* context" condition.

An analysis of variance run on the d' values showed an effect of acoustic strength, $F(1,29)=157.91$, $p<0.001$, no effect of contextual strength, $F(1,29)=2.59$, $p<0.12$, and no interaction, $F(1,29)<1$. Thus, discriminability was greater when the acoustic cues were strong than when they were mild. Importantly, there was no evidence that the effect of contextual strength found in the accuracy analyses was mirrored at the perceptual level. It is therefore likely that the effect of contextual strength resulted from a general response bias rather than a modification of the treatment of the acoustic cues. In fact, an analysis of variance with the bias (criterion) statistic c, supposed to index response tendencies independent of discriminability, revealed just that. With c calculated for each participant as 0.5^{*} (z[hit rate] + z [false-alarm rate]), we found an effect of acoustic Strength, $F(1,29)=7.47$, $p=0.01$, an effect of contextual strength, $F(1,29)=14.14$, $p<0.001$, and no interaction, $F(1,29)<1$ [see Fig. 1(d)]. Listeners were more inclined to align their response criterion with the sentential context when the acoustic cues were mild and, independently, when the sentential context was strong.

C. Discussion

Experiment 1 shows that the contextual information provided by a sentence affects the way listeners use acoustic word-juncture cues: Contextual information inconsistent with the acoustic cues caused listeners to modify their segmentation choice in the direction of the context. This effect was all the more pronounced when the acoustic cues were weak, suggesting that sentential context influences segmentation as a function of the strength of other cues. The strength of the contextual information was a modulating factor as well, with greater contextual effects when contextual information was strong. These results provide evidence for a graded, rather than all-or-nothing use of segmentation cues.

While context affected the listeners' overt segmentation behavior, it did not alter the fundamental treatment of the acoustic signal. Thus, contextual information does not cause a loss in sensitivity to lower-level information; it simply shifts it toward contextually consistent segmentation. This finding is in agreement with prior studies which have shown that sentential context operates as a response bias on phoneme perception and categorization, and not as a source of perceptual distortion (Connine, 1987; Samuel, 1981). However, these authors also demonstrated that lexical information, in contrast, had a more radical effect on perception than did context. For example, Samuel (1981) found that the ability to discriminate between a phoneme replaced with noise and one superimposed with noise was significantly worse when the phoneme was part of a real word than when it was part of a nonsense word. He attributed this difference to a genuine constraining effect of lexical knowledge on phonological processing. Sentential context was not found to cause such a decrement. In experiment 2, we examined the effect of lexicality on the perception of strong and mild word-boundary cues, using an empirical approach comparable to that of experiment 1.

IV. EXPERIMENT 2

In this experiment, we asked whether lexical knowledge influences the listeners' reliance on acoustic word-boundary cues (strong and mild) and, if so, whether such influence manifests itself as a change in acoustic discriminability or as a response bias. For comparability with experiment 1, the same acoustic phrases were used and edited to introduce a lexical contrast. However, because lexicality was less amenable to continuous gradation than was sentential context, lexicality only had one level of contrast (congruent versus incongruent). We based our lexical contrast on the notion of lexically possible segmentation (Brent, 1999; Dahan and Brent, 1999; Mattys *et al.*, 2005). On this view, listeners tend to favor segmentation solutions that do not leave any portions of the input lexically unaccounted for. Recall that, in experiment 1, phrases were designed and piloted to produce acceptable segmentation alternatives (e.g., both "pie" and "eye" in /plʌmpaɪ/ lead to a lexical leftover, i.e., "plum" and "plump," respectively). In experiment 2, the beginning of the test phrases were edited to create a lexical bias in favor of one parse (e.g., /sk̲ʌmpaɪ/ as "scum pie" rather than *"sk̲ump eye,") or the other (e.g., /l̲ʌmpaɪ/ as "lump eye"

rather than *"lum pie"). In an attempt to control the semantic plausibility of the new phrases, these were first piloted in a rating experiment.

A. Methods

1. Semantic plausibility rating and phrase selection

For each of the eight test phrases of experiment 1, we generated a large number of lexically biasing variations. For example, for /plʌmpaɪ/, *A*-inducing (e.g., "pie"-inducing) variants were, e.g., "scum pie," "drum pie," "some pie," etc., that is, phrases in which the segmentation of "pie," but not "eye," left a lexical leftover. *B*-inducing (e.g., "eye"-inducing) phrases were generated in the same fashion, e.g., "lump eye," "pump eye," "trump eye," etc. All phrases were created by changing or removing the initial portion of the first word in the original phrase. Because of limitations in the number of phrases we could create following this criterion, the number of piloted phrases varied from one set to another. In all, 222 phrases were piloted. They were presented visually, one at a time, in the center of a computer monitor, to 26 native British English participants drawn from the same population as in experiment 1. Using an 11-point scale (0 to 10), participants were asked to rate each phrase for its semantic plausibility, defined as the likelihood of seeing or hearing the phrase in everyday life, with a high score reflecting high likelihood. Then, for each of the eight sets, we picked one "*A*"-inducing phrase and one "*B*"-inducing phrase with comparable ratings (selected phrases and ratings are shown in Appendix B).[2]

2. Participants and materials

The participants for the main phase of experiment 2 were 30 individuals different from those used in the pilot experiment and experiment 1, but they were selected from the same general population and based on the same criteria. The stimuli originated from the eight pairs of phrases selected in the above section. For each of the eight sets, eight test phrases were generated via acoustic splicing. These phrases resulted from the combination of two lexical alternatives (*A* lexicality, e.g., "scum pie" versus *B* lexicality, e.g., "lump eye") and the four original acoustic conditions (strong *A*, mild *A*, mild *B*, strong *B*). Splicing was done as follows. A speaker recorded the two phrases of each set (e.g., "scum pie" and "lump eye"), of which only the onset portion was kept (e.g., /sk/ and /l/). These onset portions were spliced to the beginning of the original acoustic phrases used in experiment 1, whose onset was first removed (e.g., um##pie, um#pie, ump#eye, ump##eye) (see Table I for an example). Constraints on acoustic splicing varied from one phrase to another. In some phrases, the replaced material aligned with a clear acoustic juncture, such as a stop consonant or a fricative, in which case the juncture was used as the splicing locus. In phrases involving greater coarticulation at the juncture, the splicing locus was a zero-crossing point in the stable part of the following vowel (e.g., the first /eɪ/ in "gray day" > "clay day"). In other phrases, material was simply excised (e.g., "beer drips" > "ear drips").

The test materials included not only the 64 lexically modified phrases described above but also the 32 original, lexically neutral phrases from experiment 1 (e.g., plum##pie, plum#pie, plump#eye, plump##eye). There were also 64 filler phrases. These consisted of 32 unused phrases from the pilot experiment with an ambiguous word-word segmentation structure (see Sec. II A) and 32 edited versions of these fillers yielding nonword-word segmentations.

3. Procedure

All new materials were recorded by the speaker used previously. Test and filler phrases (160 in all) were pseudorandomized, with at least ten intervening phrases between phrases from the same set. Participants were told that, on each trial, they would hear a short phrase and that their task was to decide which of two words (e.g., "pie" or "eye") they heard at the end of the phrase. They were not given any information on whether the phrases were meaningful or not. In order to encourage participants to pay attention to the fine acoustic details of the phrases, they were told that the difference between the two alternatives was often subtle—as in experiment 1. The two word alternatives were displayed side by side on a computer monitor, 400 ms after the end of the phrase. Participants had up to 10 s to give their answer, using one of two keys on a computer keyboard. The rest of the procedure was identical to that of experiment 1. The experiment began with a practice block of twelve phrases unused in the main block.

B. Results

1. Segmentation accuracy

Response averages and their collapsed values across *A* and *B* alternatives are displayed in Figs. 2(a) and 2(b), respectively.[3] An analysis of variance, factoring acoustic strength (strong versus mild) and lexical congruency (congruent, incongruent, neutral) showed an effect of acoustic strength, $F(1,29)=346.21$, $p<0.001$, lexical congruency, $F(1,58)=127.15$, $p<0.001$, and an interaction, $F(2,58)=21.69$, $p<0.001$. Segmentation accuracy was lower when the acoustic cues were mild compared to strong, and when lexical information was incongruent with the acoustic cues as opposed to congruent with them or neutral, $F(1,29)=142.36$, $p<0.001$, and $F(1,29)=169.02$, $p<0.001$, respectively. Although numerically small, the difference between the neutral and congruent conditions reached significance, $F(1,29)=6.37$, $p<0.05$ (strong acoustics: $F<1$; mild acoustics, $p<0.01$). The interaction between acoustic strength and lexical congruency suggests that the impact of lexicality on segmentation accuracy was more pronounced when the acoustic cues were mild than when they were strong. This interaction reached significance in all pairwise comparisons (congruent-incongruent, congruent-neutral, incongruent-neutral, $p\leq 0.005$).

As in experiment 1, we used a median-split analysis to examine the possibility that the interaction originated from a ceiling effect. Participants were split into high-accuracy and low-accuracy subgroups ($N=15$ in each subgroup) based on their accuracy in the congruent, strong-acoustics condition—

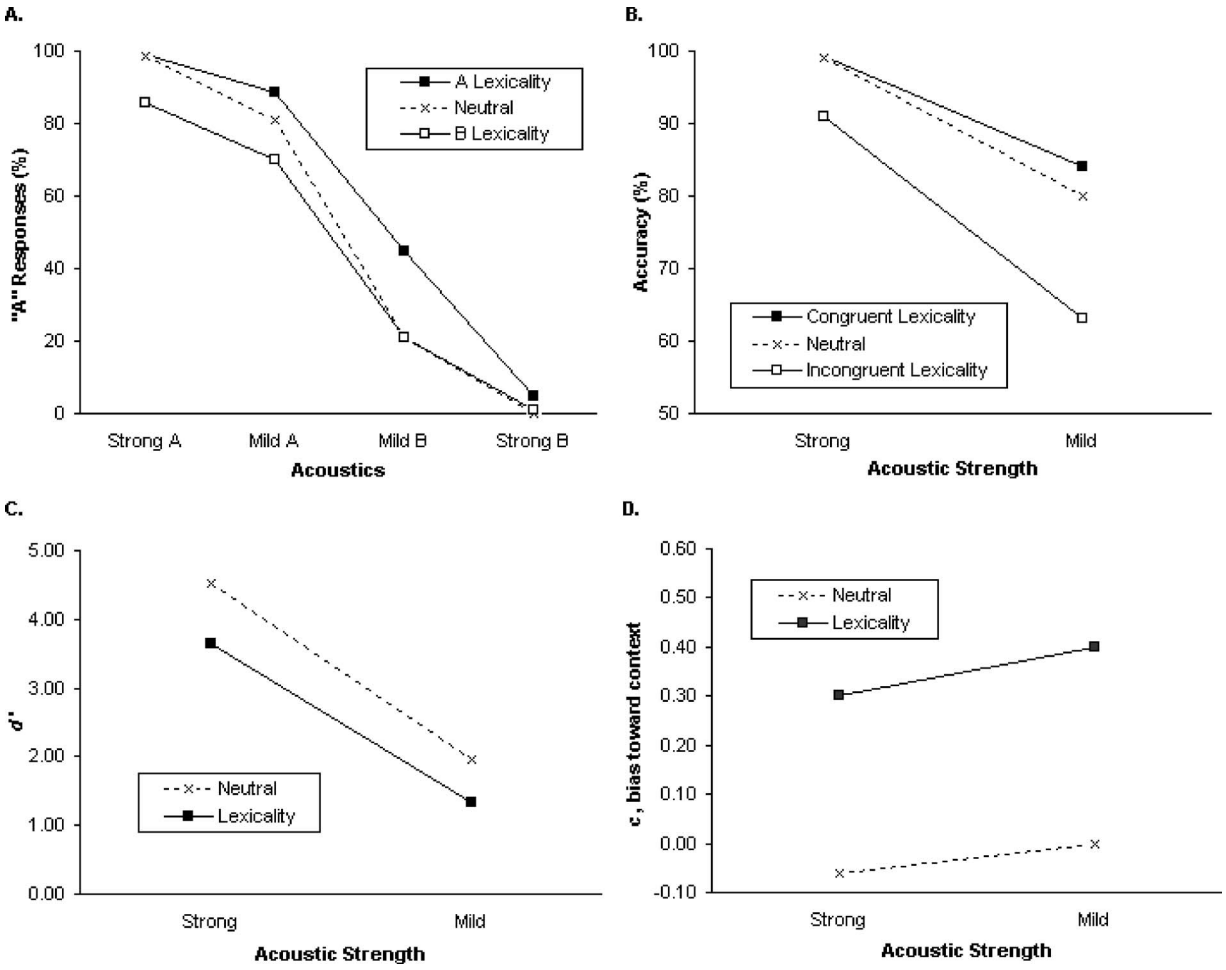


FIG. 2. Experiment 2. (a) Average percentage of *A* responses (e.g., pie) as a function of the strength of the acoustic cues and the lexical context. (b) Same as (a) when symmetrical *A* and *B* acoustic conditions are collapsed. Accuracy is measured by reference to the intended acoustic juncture. (c) d', discriminability between *A* and *B* acoustic alternatives, with lexical information present *versus* neutral. (d) *c*, response criterion expressed as a bias toward the lexically appropriate segmentation. In the neutral condition, *c* is the bias toward *A* responses.

with average accuracy across all conditions as a secondary sorting variable when needed. The analysis of variance, which included acoustic strength (strong versus mild), lexical congruency (congruent versus incongruent), and accuracy (high versus low), did not reveal any modulatory effect of accuracy on the acoustic strength by lexical congruency interaction, $F(1,28)<1$. An analysis of variance focusing on the low-accuracy group showed the same pattern of results as that for the entire group: acoustic strength, $F(1,14)=151.43$, $p<0.001$; lexical congruency, $F(1,14)=64.75$, $p<0.001$; interaction, $F(1,14)=9.66$, $p<0.01$. Thus, there was no evidence that the reduced effect of lexicality in the strong-acoustics condition was the result of a ceiling artifact.

A main effect of lexicality, calculated as the polarity-adjusted difference between the neutral condition and the congruent or incongruent condition, indicated that the presence of lexical information affected accuracy, $F(1,29)=142.36$, $p<0.001$. This effect was larger in the mild than in the strong acoustic condition, $F(1,29)=30.00$, $p<0.001$ (both pairwise comparisons reached $p<0.001$). Finally, when each of the six mean values was compared to the 50% chance level, two-tail $t(29)$ yielded $p<0.001$ in all conditions.

In sum, listeners performed the segmentation task less accurately when a possible lexical boundary misaligned with an acoustic juncture and somewhat more accurately when it aligned with it. Moreover, the penalty for lexical misalignment was larger when the acoustic cues were mild than when they were strong. This lexical/acoustics trade-off pattern is similar to that observed with sentential context in experiment 1, suggesting that less reliable acoustic cues provide a favorable ground for lexically/contextually driven segmentation.

On first analysis, the relatively small difference between the congruent and neutral conditions suggests that the contribution of lexical information to acoustic juncture detection is less of a confirmatory than a disconfirmatory nature. However, alignment benefits could have been hindered by a near ceiling performance in the neutral condition. The emergence of a significant difference when overall performance was lower, namely, the mild-acoustics condition, is consistent with this possibility. In either case, the presence of lexical information in the signal showed a clear modulatory effect on the listeners' segmentation responses. The nature of this lexical effect is analyzed further in the next section.

2. Juncture discriminability

As with contextual information, lexicality could affect segmentation responses by either shifting listeners' response criterion toward the lexically acceptable parsing, with no significant reduction in acoustic discriminability, or by attenuating the listeners' sensitivity to the acoustic cues. Because the lexical variable did not allow us to create strong and mild degrees of lexicality, lexical effects on discriminability were calculated by comparing d' in the lexically biased phrases with d' in the neutral phrases [Fig. 2(c)].

As an example, d' for the "strong acoustics-lexicality" condition was the average of d' in the "strong acoustics-*A* lexicality" and "strong acoustics-*B* lexicality" conditions, calculated individually for each participant. The hit rate of the "strong acoustics-*A* lexicality" was the percentage of *A* responses in the "strong *A* acoustics-*A* lexicality" condition, and the false-alarm rate was the percentage of *A* responses in the "strong *B* acoustics-*A* lexicality" condition. A comparable compilation was done to calculate d' in the "strong acoustics-*B* lexicality" condition.

The d' values were entered into an analysis of variance with acoustic strength (strong versus mild) and lexicality (lexicality versus neutral) as the main variables. The results showed an acoustic strength effect, $F(1,29)=489.26$, $p<0.001$, a lexicality effect, $F(1,29)=97.04$, $p<0.001$, and no interaction, $F(1,29)=2.70$, $p=0.11$. Thus, critically, discriminability was reduced when lexicality was present in the signal. An analysis of variance run on the response criterion c [Fig. 2(d)], indicated that listeners also aligned their response criterion with the lexical context: lexicality, $F(1,29)=34.42$, $p<0.001$; acoustic strength, $F(1,29)=2.23$, $p<0.15$; interaction, $F(1,29)<1$. In sum, when lexical information could be used to guide segmentation, it did so not only by biasing segmentation responses in the direction of the lexically acceptable alternative, but also by curtailing phonetic analysis.

C. Discussion

The results show that lexical knowledge, like sentential context, influences the listeners' reliance on acoustic word-juncture cues by causing them to align their segmentation choice with a lexically acceptable boundary. Like the sentential effect, the lexical effect was more pronounced when the acoustic cues were less reliable, suggesting a trade-off relation between lexical and acoustic cues. However, unlike sentential context, lexical knowledge affected the listeners' sensitivity to the acoustic cues. The ability to discriminate between the competing acoustic cues was significantly reduced when lexically driven segmentation was an option. The contrasted effects of sentential and lexical information on acoustic discriminability align well with prior research on phoneme perception, in which lexical, but not sentential information was shown to modify the processing of the sensory input (e.g., Connine, 1987; Samuel, 1981).

V. GENERAL DISCUSSION

This study set out to test the influence of sentential and lexical constraints on the segmentation of near-homophonous phrases bearing strong and mild acoustic word-boundary cues. We draw two conclusions from the results: (1) in multicue listening environments, listeners adopt compensatory segmentation strategies whereby reliance on sentential/lexical information is inversely related to the strength of the acoustic cues, and (2) lexically induced segmentation is accompanied with a reduction in phonetic acuity whereas sententially induced segmentation is not.

The graded effect of fine phonetic details on speech segmentation, although not entirely unexpected, had not been demonstrated before, as most experiments have so far restricted their analyses to dichotomous designs, in which highly favorable cues are pitted against highly unfavorable ones. Our results suggest that segmentation preferences are commensurate with the gradient of the available acoustic information. The way in which higher-order information combines with acoustic cues was found to be flexible too, with the former compensating for a lack of specificity in the latter. Thus, not unlike trade-off relations between phonetic features in the perception of phonetic contrasts (e.g., Pisoni and Luce, 1987; Repp, 1982), graded interactions between top-down and bottom-up information exist for the perception of word boundaries as well.

It should be noted, however, that the present study highlights segmentation patterns that are at odds with Mattys *et al.*'s (2005) claim that listeners generally adhere's to segmentation strategies that favor meaning over signal. To illustrate this, the notion of cue equivalence (e.g., Fitch, Halwes, Erickson, and Liberman, 1980) is helpful. In a situation of cue equivalence, listeners experience similar percepts in conditions where cues have mirror polarities. For example, in experiment 1, for sentential/lexical and acoustic cues to have equivalent strength, the percentage of *A* responses in the "strong *A* acoustics-strong *B* context" condition should be comparable to the percentage of *A* responses in the "strong *B* acoustics-strong *A* context" condition. However, the percentages were 81% and 14%, respectively. With mild cues, too, the percentages differed: 67% and 41%, respectively. Thus, the acoustic cues received greater weight than the contextual cues. Even when mild acoustic cues were played against a strong context, responses still favored an acoustic interpretation, at least numerically—all conditions in Fig. 1(b) received scores above 50%. Similar patterns were found in experiment 2.

An explanation for the discrepancy between these results and Mattys *et al.*'s (2005) proposal might be found in the task and instructions we used. In the present study, nothing explicitly encouraged participants to pay attention to the meaning of the sentences or phrases. Instead, the instructions directed attention to the fine acoustic details of the utterances. This was apparently sufficient to cause participants to give more credence to the acoustic than sentential/lexical cues. Obviously, signal-oriented tasks are not the most ecological proxy for everyday-listening strategies, and it is likely that a task involving semantic processing of some sort would show the usual dominance of higher-order information. Our signal-oriented task was meant to make effects of higher-order knowledge functionally unnecessary, and hence, to constitute a more stringent test for the graded interaction

hypothesis. The key finding, therefore, is that, even with an attentional bias for bottom-up evidence, clear modulatory effects of sentential and lexical contexts arise, and the size of these effects is a function of the strength of the acoustic cues. This can serve as a first step toward refining the hierarchical framework (Mattys *et al.*, 2005), in which cues were tested dichotomously and no provision was made for graded trading across tiers of information. The present results suggest that some room for graded compensation should be envisaged, and the hierarchy allowed to be modulated by the relative strength of the cues. Whether higher-order information can ever be overridden by lower-level cues in natural listening conditions is unclear at this stage. A methodological starting point to answer this question could be sought in studies showing that extreme reduction in intelligibility leads to indiscriminate, and often detrimental reliance on low-level information (e.g., Cutler and Butterfield, 1992; Liss, Spitzer, Caviness, Adler, and Edwards, 1998, 2000; Smith, Cutler, Butterfield, and Nimmo-Smith, 1989).

Although sentential and lexical information showed remarkably similar effects on segmentation accuracy, only lexical information affected the listeners' sensitivity to the acoustic cues. Contextual effects were confined to a decision stage, with no associated reduction in acoustic sensitivity. It could be argued that a smaller constraining range in the sentential condition (strong versus mild) than in the lexical condition (lexical versus neutral) might restrict comparability between the two sources of higher-order information. In a cross-experiment analysis run on segmentation accuracy, a significant interaction between experiment (context versus lexicality) and strength (strong context versus mild context or lexicality versus neutral), $F(1,58)=6.23$, $p=0.01$, did indeed show that the lexical effect was larger than the contextual-strength effect. However, when the magnitude of this effect was entered as a covariate in a cross experiment on the d' results, it did not interact with any of the effects previously found. In other words, the lack of a sentential effect on d' in experiment 1 and the presence of a lexical effect on d' in experiment 2 cannot be reduced to an initial imbalance in the magnitude of these effects on segmentation accuracy. Thus, we feel confident concluding that the contrasted d' patterns in the sentential and lexical conditions stemmed from distinct processes promoted by these two sources of information rather than from design artifacts. A similar functional distinction was noted by Samuel (1981) and Connine (1987) in the phoneme-perception domain. Likewise, Davis, Johnsrude, Hervais-Adelman, Taylor, and McGettigan (2005) found a more profound effect of lexical knowledge than sentential structure on the perceptual learning of distorted speech. In that study, perceptual learning of noise-vocoded speech was enhanced when the listeners were presented with lexical feedback, but not with "jabberwocky" sentences, that is, sentences with a normal syntactic structure but where the content words have been replaced with nonsense words. Moreover, syntactic prose, i.e., sentences whose meaning is eliminated by randomly replacing content words with unrelated words, was just as effective as normal prose. Thus, once again, lexical knowledge, not sentence-level information, was the driving force behind perceptual learning.

As far as segmenting speech into words, we propose that the reason why lexical knowledge has a more profound effect on acoustic processing than does sentential information has to do with the lexical viability—and associated communicative relevance—of the segmentation alternatives that the two sources of information generate. While our incongruent sentential context never yielded nonsense words, our incongruent lexical context did. Thus, it is conceivable that the recognition system relinquishes acoustic sensitivity if higher-order knowledge supports a single lexical solution, and is reluctant to do so if more than one segmentation solution is lexically viable, irrespective of contextual congruency.

The above analysis suggests that a change in acoustic sensitivity occurs subsequently to lexical competition—provided that such competition leads to a lexically unambiguous segmentation solution. This is not to say that the role of acoustic details to segmentation is negligible, as the accuracy data clearly highlight their contribution to the pattern of responses, but it would often be short lived. Acoustic details would initially increase or decrease the activation level of the candidates that align or misalign with them, as suggested in Gow and Gordon (1995) good-start model in which proper acoustic cues buttress lexical activation and, more generally, in Norris, McQueen, Cutler, and Butterfield's (1997) possible-word constraint conceptualization. As lexical competition is settling, sensitivity to phonetic details would decrease, unless alternative lexical parses make such details necessary for possible resegmentation. One way of testing this acoustic-sensitivity decay hypothesis would be to impose various deadlines on the listeners' responses and measure the modulation of the d' data across these deadlines in a sentential versus lexical context. Interestingly, the absence of an interaction between acoustic strength and higher-order information in the d' data of both experiments suggests that a modification of acoustic sensitivity is not more likely when the acoustic evidence is weak than when it is strong. The only critical factor, it seems, is whether the source of higher-order information is lexical or sentential, that is, in our experiments, whether it yields a single lexically possible solution (scum#pie) as opposed to multiple ones (plum#pie and plump#eye), respectively.

Current models of spoken-word recognition cannot directly account for the above results because either they do not specify the contribution of subsegmental information to word recognition or they are not equipped to deal with connected speech. Yet, a composite picture can be offered. The Shortlist model (Norris, 1994) is compatible with the general idea of segmentation resulting from the competition of multiple lexical candidates overlapping portions of the continuous input. However, Shortlist's entry layer is phonemic, not acoustic-phonetic. Thus, although Shortlist has probably no theoretical objection to coding the input subsegmentally, it currently cannot accommodate our effect of acoustic gradient on segmentation patterns. More importantly, Shortlist does not predict that the sensitivity to acoustic details should drop in a lexically constraining environment. For both Shortlist and Merge (Norris, McQeen, and Cutler, 2000), lexical effects should be confined to a decision stage, as was indeed the case with our sentential context. While TRACE (McClel-

land and Elman, 1986) also limits its description of the input to phonemic representations, it does have provision for lexical feedback to sublexical levels. Thus, the basic mechanism leading to a reduction in sensitivity to discrepant sublexical information is available in TRACE, but how this mechanism could operate at the acoustic-phonetic level is left unspecified. Moreover, TRACE has a rather uneconomical lexical-activation algorithm compared to that of Shortlist. One way in which our data can be modeled is by incorporating a more direct mapping between the continuous signal and lexical representations (e.g., Gaskell and Marslen-Wilson, 1997), thus allowing subsegmental details to bias lexical activation—though probably not completely determine it. Activation and selection themselves would be feedforward, as implemented in Shortlist, but the resulting reduction in sensitivity to sublexical details would imply top-down connections. Sentential constraints on lexical selection would not involve a modification of sublexical processing since lexical knowledge alone would favor more than one segmentation solution.

An issue for future research concerns the implications of the above hypothesis for episodic models of lexical encoding (Goldinger, 1996, 1998; Luce, McLennan, and Charles-Luce, 2003). Our results suggest that the extent to which subsegmental details are encoded in long-term memory could be a function of the success of lexically driven segmentation. A lexically unique solution would decrease sensitivity to the subsegmental details of the signal, and hence, reduce the encoding of episodic information, whereas lexically ambiguous solutions would cause subsegmental details to remain available, and hence, become part of lexical representations. If this is indeed the case, then models differing in the role they attribute to episodic information in lexical encoding may have to be examined in the context of connected speech and word segmentation. This could be done by adapting long-term priming and recognition-memory paradigms (e.g., Luce and Lyons, 1998) to stimuli whose allophonic specificity is more or less supported by sentential versus lexical information.

In conclusion, our experiments stress the need to study multicue speech segmentation with an eye on gradation, compensation, and flexibility, while highlighting functional distinctions between sentential and lexical contributions to this process.

ACKNOWLEDGMENTS

This research was supported by a Grant from the Biotechnology and Biological Sciences Research Council (BBSRC) to S. L. Mattys (Grant No. 7/S18783). We thank Lucy Series for collecting the data of experiment 2.

APPENDIX A: SELECTED ACOUSTIC PHRASES AND SENTENTIAL CONTEXTS AND THEIR RATINGS FROM THE PILOT EXPERIMENT

Ratings are expressed in percentages of *A* responses.

Phrases	Strong *A*	Mild *A*	Mild *B*	Strong *B*
1. plum pie (*A*) - plump eye (*B*)	84	74	23	13
2. a name (*A*) - an aim (*B*)	86	67	34	12
3. bake enforce (*A*) - bacon force (*B*)	90	72	25	10
4. gray tapes (*A*) - great apes (*B*)	85	72	29	14
5. beer drips (*A*) - beard rips (*B*)	88	75	31	18
6. gray day (*A*) - grade A (*B*)	83	66	30	14
7. more rice (*A*) - more ice (*B*)	89	66	35	11
8. her rear (*A*) - her ear (*B*)	83	56	45	12
Average	86%	68%	32%	13%

Sentences

1. plum pie (*A*) - plump eye (*B*)		
Strong *A*	The baker looked at the drawing of a	88
Mild *A*	The girl looked at the drawing of a	67
Mild *B*	The scientists looked at the drawing of a	25
Strong *B*	The surgeon looked at the drawing of a	16
2. a name (*A*) - an aim (*B*)		
Strong *A*	Before forging an identity, the first task is to take	89
Mild *A*	Before starting a new life, the first task is to take	68
Mild *B*	To start the game, you must take	44
Strong *B*	To be the best in his field was	7
3. bake enforce (*A*) - bacon force (*B*)		
Strong *A*	To make cakes and biscuits, use	83
Mild *A*	The first job of the day for the chefs was	67
Mild *B*	The first job of the day at the pork counter was	28

Phrases				
Strong *B*	The pig farming survey was dubbed the			13
4. gray tapes (*A*) - great apes (*B*)				
Strong *A*	His first assignment was to stack			93
Mild *A*	His first assignment was to collect			74
Mild *B*	His first assignment was to tag			26
Strong *B*	His first assignment was to befriend			14
5. beer drips (*A*) - beard rips (*B*)				
Strong *A*	The annoying sound can be heard coming from the bar as the			88
Mild *A*	It is possible to hear the beat as the			75
Mild *B*	It is possible to hear the crackle as the			35
Strong *B*	The painful sound can be heard coming from the man as the			15
6. gray day (*A*) - grade A (*B*)				
Strong *A*	There was little to cheer with this			86
Mild *A*	They were fairly satisfied despite the			75
Mild *B*	She was going to treat her friends after the			27
Strong *B*	Overall he was much happier after that			16
7. more rice (*A*) - more ice (*B*)				
Strong *A*	At the Chinese restaurant, there was a call for			87
Mild *A*	At the supermarket, there was a request for			69
Mild *B*	At the work's party, there was a call for			26
Strong *B*	At the cocktail party, there was a call for			11
8. her rear (*A*) - her ear (*B*)				
Strong *A*	She turned towards him when he put a hand on			85
Mild *A*	She turned towards him when he mentioned			71
Mild *B*	She turned towards him when he shouted to			26
Strong *B*	The ringing sound came from			10
		Average	Strong *A*	87%
			Mild *A*	71%
			Mild *B*	31%
			Strong *B*	14%

APPENDIX B: PHRASES IN EXPERIMENT 2

These were modified versions of the phrases in experiment 1, with one phrase lexically favoring the *A* segmentation alternative and the other phrase favoring the *B* segmentation alternative. The modifications, i.e., new spliced onsets, are underlined (some phrases without underlining were modified by removing material at their beginning). The average semantic plausibility of the phrases is also reported

(0= not likely, 1= very likely).

1. scum pie (*A*)	0.29	lump eye (*B*)	0.33
2. number name (*A*)	0.61	nation aim (*B*)	0.63
3. make enforce (*A*)	0.39	passion force (*B*)	0.48
4. today tapes (*A*)	0.55	create apes (*B*)	0.53
5. ear drips (*A*)	0.60	weird rips (*B*)	0.61
6. clay day (*A*)	0.31	aid *A*(*B*)	0.52
7. before rice (*A*)	0.72	your ice (*B*)	0.78
8. endure rear (*A*)	0.31	incur ear (*B*)	0.30
Average semantic Plausibility	0.47		0.52

[1]Although the strong and mild phrases were selected to be of comparable strength in the *A* and *B* acoustic conditions, we tested for any asymmetry in the listeners' responses. The percentage of correct responses collapsed across all four sentential contexts was entered in an analysis of variance with *AB* (*A* acoustics versus *B* acoustics) and strength (strong versus mild) as repeated-measure factors. As expected, the strength effect was highly significant, $F(1,29)=215.39$, $p<0.001$, showing higher accuracy with strong than with mild acoustic cues, but there was no indication that the listeners performed differently in the *A* versus *B* acoustic conditions, $F(1,29)<1$. However, an interaction between *AB* and strength, $F(1,29)=14.36$, $p<0.001$, revealed a bias for *B* responses (e.g., "eye") in the strong acoustic condition, $F(1,29)=6.44$, $p<0.05$, and a bias for *A* responses (e.g., "pie") in the mild acoustic condition, $F(1,29)=4.52$, $p<0.05$. The bias for *B* responses is difficult to account for, although it could result from an interaction with the sentential contexts undetected in the pilot experiment. The bias for *A* responses in the mild acoustic condition could reflect the fact that listeners might occasionally perceive the pivotal consonant (e.g., /p/) as geminated, and consequently hear plump#eye as "plump pie." This phenomenon would be less likely to occur with clearer (i.e., strong) acoustic cues. These slight biases have no major implications for the accuracy analyses, however, because the polarity of the acoustic conditions was always collapsed. The d' scores in the mild condition could have been somewhat underestimated due to the extra "pie" responses given to an illusory "third alternative" (e.g., "plump pie"). However, it is unlikely that such a third alternative—if validated—could have been the only cause of the highly significant d' decrement from the strong to the mild acoustic conditions. Note, too, that the acoustic effect was predicted from the pilot data which did not allow listeners to consider a third alternative. Finally, because the *c* data were calculated by collapsing context bias across both *A* and *B* contexts, any response bias incidentally favoring the *A* context was

canceled out by its opposite effect on the *B* context, and vice versa.

[2]The numerous constraints on stimulus selection made it difficult to find new onsets that rendered the lexical bias symmetrical within a pair of stimuli and comparable across pairs of stimuli. For example, it could be argued that the nonlexical leftover in *"maken force" is compatible with the conversational realization of "making," that the nonlexical phrase *"todate apes" is compatible with a three-word parsing ("to date apes"), and that "eared" in *"eared rips" is a low-frequency word. It could also be argued that the stimuli containing an optional word-final /r/ in British English at the critical juncture ("before ice" / "your ice" and "endure rear" / "incur ear") were not as lexically biasing as they could have been. If the above "nonlexical" parses had been substantially more lexically acceptable than intended, they would have made it hard for a lexical effect to emerge. The clear lexical effect in experiment 2 suggests that, if these alternatives were considered by the listeners, they were of little consequence. In fact, an analysis of variance restricted to the five sets above showed lexical congruency effects similar in magnitude and significance to those with the full set of stimuli. Thus, the present lexical effect on segmentation seems to be robust and not significantly affected by a slight artifactual reduction in lexical contrast.

[3]As in experiment 1, we tested for any *A* versus *B* asymmetry in the listeners' responses. The percentage of correct responses collapsed across the three lexical conditions was entered in an analysis of variance with *AB* (*A* acoustics versus *B* acoustics) and strength (strong versus mild) as repeated-measure factors. The results were comparable to those of experiment 1 (see footnote 1). The strength effect was significant, $F(1,29)=346.21$, $p<0.001$, and there was no significant difference between *A* and *B* acoustics, $F(1,29)<1$. As before, an interaction between *AB* and strength, $F(1,29)=16.33$, $p<0.001$, showed a bias for *B* responses (e.g., "eye") in the strong acoustic condition, $F(1,29)=8.43$, $p<0.01$, and a bias for *A* responses (e.g., "pie") in the mild acoustic condition, $F(1,29)=8.67$, $p<0.01$. Given the similarity between these biases and those in experiment 1, it appears that the task used in the pilot experiment (e.g., "plum pie" versus "plump eye") and that used in the main experiments (e.g., "pie" versus "eye") involved somewhat different judgment thresholds and stimulus idiosyncrasies. As indicated in footnote 1, however, the asymmetry between the *A* and *B* conditions should not have a major impact on the main pattern of results.

Baayen, R. H., Piepenbrock, R., and Gulikers, L. (**1995**). *The CELEX Lexical Database* (Release 2) [CD-ROM] (Philadelphia, PA: Linguistic Data Consortium, University of Pennsylvania [Distributor]).

Brent M. R. (**1999**). "An efficient, probabilistically sound algorithm for segmentation and word discovery," Machine Learning, **34**, pp. 71–106.

Christophe, A., Peperkamp, S., Pallier, C., Block, E., and Mehler, J. (**2004**). "Phonological phrase boundaries constrain lexical access: I. Adult data," J. Mem. Lang. **51**, 523–547.

Church, K. W. (**1987**). *Phonological Parsing in Speech Recognition* (Kluwer Academic Publishers, Dordrecht).

Connine C. M. (**1987**). "Contraints on interactive processes in auditory word recognition: The role of sentence context," Journal of Memory and Learning **26**, pp. 527–538.

Connine C. M. and Clifton, C. (**1987**). "Interactive use of lexical information in speech perception," Journal of Experimental Psychology: Human Perception and Performance, **13**, pp. 291–299.

Cutler, A., and Butterfield, S. (**1992**). "Rhythmic cues to speech segmentation: Evidence from juncture misperception," J. Mem. Lang. **31**, 218–236.

Cutler, A., and Norris, D. G. (**1988**). "The role of stressed syllables in segmentation for lexical access," J. Exp. Psychol. Hum. Percept. Perform. **14**, 113–121.

Dahan, D., and Brent, M. R. (**1999**). "On the discovery of novel wordlike units from utterances: An artificial-language study with implications for native-language acquisition," J. Exp. Psychol. Gen. **128**, 165–185.

Davis, M. H., Johnsrude, I. S., Hervais-Adelman, A., Taylor, K., and McGettigan, C. (**2005**). "Lexical information drives perceptual learning of distorted speech: Evidence from the comprehension of noise-vocoded sentences," J. Exp. Psychol. Gen. **134**, 222–241.

Davis, M. H., Marslen-Wilson, W. D., and Gaskell, M. G. (**2002**). "Leading up the lexical garden-path: Segmentation and ambiguity in spoken-word recognition," J. Exp. Psychol. Hum. Percept. Perform. **28**, 218–244.

Dilley, L., Shattuck-Hufnagel, S., and Ostendorf, M. (**1996**). "Glottalization of vowel-initial syllables as a function of prosodic structure," J. Phonetics **24**, 423–444.

Fitch, H. L., Halwes, T., Erickson, D. M., and Liberman, A. M. (**1980**). "Perceptual equivalence of two acoustic cues for stop-consonant manner," Percept. Psychophys. **27**, 343–350.

Gaskell, M. G., and Marslen-Wilson, W. D. (**1997**). "Integrating form and meaning: A distributed model of speech perception," Lang. Cognit. Processes **12**, 613–656.

Gimson, A. C. (**1980**). *An Introduction to the Pronunciation of English*, 3rd ed. (Edward Arnold, London).

Goldinger, S. D. (**1996**). "Words and voices: Episodic traces in spoken word identification and recognition memory," J. Exp. Psychol. Learn. Mem. Cogn. **22**, 1166–1183.

Goldinger, S. D. (**1998**). "Echoes of echoes? An episodic theory of lexical access," Psychol. Rev. **105**, 251–279.

Gow, D. W., and Gordon, P. C. (**1995**). "Lexical and prelexical influences on word segmentation: Evidence from priming." J. Exp. Psychol. Hum. Percept. Perform. **21**, 344–359.

Green, D. M., and Swets, J. A. (**1996**). *Signal Detection Theory and Psychophysics* (John Wiley and Sons, New York).

Hawkins, S., and Smith, R. (**2001**). "Polysp: A polysystemic, phonetically-rich approach to speech understanding," J. Ital. Ling. - Rivista di Linguistica **13**, 99–188.

Hawkins, S., and Warren, P. (**1994**). "Phonetic influences on the intelligibility of conversational speech," J. Phonetics **22**, 493–511.

Lieberman, P. (**1963**). "Some effects of semantic and grammatical context on the production and perception of speech," Iowa Dent. Bull. **6**, 172–187.

Lindblom, B. (**1990**). "Explaining phonetic variation: A sketch of the H&H theory," in *Speech Production and Speech Modelling*, edited by W. J. Hardcastle and A. Marchal (Kluwer, Amsterdam), pp. 403–439.

Liss, J. M., Spitzer, S., Caviness, J. N., Adler, C., and Edwards, B. (**1998**). "Syllabic strength and lexical boundary decisions in the perception of hypokinetic dysarthric speech," J. Acoust. Soc. Am. **104**, 2457–2466.

Liss, J. M., Spitzer, S., Caviness, J. N., Adler, C., and Edwards, B. (**2000**). "Lexical boundary error analysis in hypokinetic and ataxic dysarthria," J. Acoust. Soc. Am. **107**, 3415–3424.

Luce, P. A., and Lyons, E. A. (**1998**). "Specificity of memory representations for spoken words," Mem. Cognit. **26**, 708–715.

Luce, P. A., McLennan, C. T., and Charles-Luce, J. (**2003**). "Abstractness and specificity in spoken word recognition: Indexical and allophonic variability in long-term repetition priming," in *Rethinking Implicit Memory*, J. Bowers and C. Marsolek, edited by (Oxford University Press, New York), pp. 197–214.

Mattys, S. L. (**2004**). "Stress versus coarticulation: Toward an integrated approach to explicit speech segmentation," J. Exp. Psychol. Hum. Percept. Perform. **30**, 397–408.

Mattys, S. L., White, L., and Melhorn, J. F. (**2005**). "Integration of multiple speech segmentation cues: A hierarchical framework." J. Exp. Psychol. Gen. **134**, 477–500.

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

McQueen J. M. (**1998**). "Segmentation of continuous speech using phonotactics." Journal of Memory and Learning, **39**, pp. 21–46,

Nakatani L. H. and K. D.Dukes (**1977**). "Locus of segmental cues to word juncture." J. Acoust. Soc. Am. **62**, 714–719 .

Norris, D. (**1994**). "Shortlist: A connectionist model of continuous speech recognition," Cognition **52**, 189–234.

Norris, D. (**1995**). "Signal detection theory and modularity: On being sensitive to the power of bias models of semantic priming," J. Exp. Psychol. Hum. Percept. Perform. **21**, 935–939.

Norris, D., McQueen, J. M., and Cutler, A. (**2000**). "Merging information in speech recognition: Feedback is never necessary," Behav. Brain Sci. **23**, 299–370.

Norris, D., McQueen, J. M., Cutler, A., and Butterfield, S. (**1997**). "The possible-word constraint in the segmentation of continuous speech." Cogn. Psychol. **34**, 191–243.

Pisoni, D. B., and Luce, P. A. (**1987**). "Trading relations, acoustic cue integration, and context effects in speech perception," in *The Psychophysics of Speech Perception*, edited by M. E. H. Schouten, Martinus Nijhoff, Dordrecht, pp. 155–172.

Polka, L., and Strange, W. (**1985**). "Perceptual equivalence of acoustic cues that differentiate /r/ and /l/," J. Acoust. Soc. Am. **78**, 1187–1197.

Pylyshyn, Z. (**1999**). "Is vision continuous with cognition? The case for cognitive impenetrability of visual perception," Behav. Brain Sci. **22**, 341–365.

Redi, L., and Shattuck-Hufnagel, S. (**2001**). "Variation in realization of glottalization in normal speakers," J. Phonetics **29**, 407–429.

Repp, B. H. (**1982**). "Phonetic trading relations and context effects: New experimental evidence for a speech mode of perception." Psychol. Bull.

92, 81–110.
Salverda, A. P., Dahan, D., and McQueen, J. M. (**2003**). "The role of prosodic boundaries in the resolution of lexical embedding in speech comprehension," Cognition **90**, 51–89.
Samuel A. G.. (**1981**). "Phonemic restoration: Insights from a new methodology." Journal of Experimental Psychology: General, **110**, 474–494.
Samuel A. G. (**1997**). "Lexical activation produces potent phonemic percepts." Cognitive Psychology, **32**, 97–127.
Samuel A.G. (**2001**). "Knowing a word affects the fundamental perception of the sounds within it." Psychological Science, **12**, pp. 348–351.
Smith, M. R., Cutler, A., Butfetfield, S., and Nimmo-Smith, I. (**1989**). "The perception of rhythm and word boundaries in noise-masked speech." J. Speech Hear. Res. **32**, 912–920.
Tanner, W. P., and Swets, J. A. (**1954**). "A decision-making theory of human detection," Psychol. Rev. **61**, 401–409.
Turk, A. E., and Sawusch, J. R. (**1997**). "The domain of accentual lengthening in American English," J. Phonetics **25**, 25–41.
Turk, A. E., and White, L. (**1999**). "Structural influences on accentual lengthening in English," J. Phonetics **27**, 171–206.
Umeda, N., and Coker, C. H. (**1974**). "Allophonic variation in American English," J. Phonetics **2**, 1–5.
Vitevitch, M. S., and Luce, P. A. (**1999**). "Probabilistic phonotactics and neighborhood activation in spoken word recognition," J. Mem. Lang. **40**, 374–408.

Acoustical classification of woods for string instruments

Shigeru Yoshikawa
Department of Acoustic Design, Faculty of Design, Kyushu University, 4-9-1 Shiobaru, Minami-ku, Fukuoka, 815-8540 Japan

(Received 22 February 2007; revised 8 April 2007; accepted 2 May 2007)

Two basic types of wood are used to make stringed musical instruments: woods for soundboards (top plates) and those for frame boards (back and side plates). A new way to classify the acoustical properties of woods and clearly separate these two groups is proposed in this paper. The transmission parameter (product of propagation speed and Q value of the longitudinal wave along the wood grain) and the antivibration parameter (wood density divided by the propagation speed along the wood grain) are introduced in the proposed classification scheme. Two regression lines, drawn for traditional woods, show the distinctly different functions required by soundboards and frame boards. These regression lines can serve as a reference to select the best substitute woods when traditional woods are not available. Moreover, some peculiarities of Japanese string instruments, which are made clear by comparing woods used for them with woods used for Western and Chinese instruments, are briefly discussed. © *2007 Acoustical Society of America.*
[DOI: 10.1121/1.2743162]

PACS number(s): 43.75.Gh, 43.75.Mn, 43.75.De [NHF] Pages: 568–573

I. INTRODUCTION

Wood selection is the most important parameter in stringed instrument design. The body of a stringed instrument consists, in general, of a soundboard (top plate) and a frame board (back plate and/or side plate) as illustrated in Fig. 1. The lute or guitar family, shown in Fig. 1(a), is a typical example. Since the back plate is connected with the top plate by a sound post in the fiddle and violin family as shown in Fig. 1(b), the back plate can function as part of the soundboard as well as of the frame board. Modern pianos depicted in Fig. 1(c) have a very rigid frame structure, instead of a frame board, which firmly supports the soundboard as the prime sound radiator. The modern piano's single soundboard structure is essentially different from other string instruments, which have a common structure consisting of a resonant box and a sound hole (or sound holes). Note, however, that the fortepianos made by Bartolomeo Cristofori in the 1720s maintain such a box-hole structure (four big holes are hidden below the keyboard) as seen in the harpsichord and clavichord.[1,2]

Woods for string instruments are thus divided into two groups: (A) woods for soundboards and (B) woods for frame boards. Just as in human society we put "the right person in the right place," so in musical instrument design it is essential to put "the right wood in the right place," especially in string instruments. The objective of this paper is to propose a new classification diagram that clearly discriminates the acoustical characteristics of group A from those of group B. A scheme that can successfully classify the woods traditionally used for the highest-quality string instruments can also serve to guide the selection of both substitute woods and synthetic materials, with the next best, or possibly even better, quality. Such selection criteria have not been hitherto established. In the face of serious shortages of natural materials from endangered species, it is very important to develop alternatives.

In addition, the proposed classification scheme can be used to clarify and understand some of the differences between Western string instruments and East Asian ones.[3] Although both the style and manner of playing string instruments ultimately arises from differences in musical taste, it is found that the wood material also has a strong influence.

II. PARAMETERS TO CHARACTERIZE WOOD PROPERTIES

A. Wood vibration and its transmission

Fletcher and Rossing[4] added Chap. 22, "Material of Musical Instruments," to the second edition of their excellent textbook. Section 22.3, "Wood Material," provides concise and useful information on wood science, but it does not give a scheme to classify the two groups of woods. Neither Bucur's well-organized book[5] nor the anthology edited by Hutchins and Benade[6] addresses the problem of acoustical classification. It is hoped that the proposed classification scheme can fill this gap.

Two parameters that can be used to clearly discriminate between the two wood types are sought. In a good classification scheme the two wood types ought to be "orthogonal" because soundboards and frame boards have opposite vibrational and acoustical properties.

Various parameters have been proposed to describe the acoustical properties of traditional woods used for string instruments and of substitute woods.[7–12] Research has also been done on artificial composites.[13–15] Among the various parameters that have been proposed, the most plausible one is Schelleng's c/ρ, where c denotes the propagation speed of the longitudinal wave along the wood grain, and ρ denotes the wood density.[9] The relevance of c/ρ is confirmed in Ref. 4. Schelleng derived it by supposing that both the stiffness and the inertia of the plate should be the same for the wood substitute if its vibrational properties are to be the same.

 0001-4966/2007/122(1)/568/6/$23.00

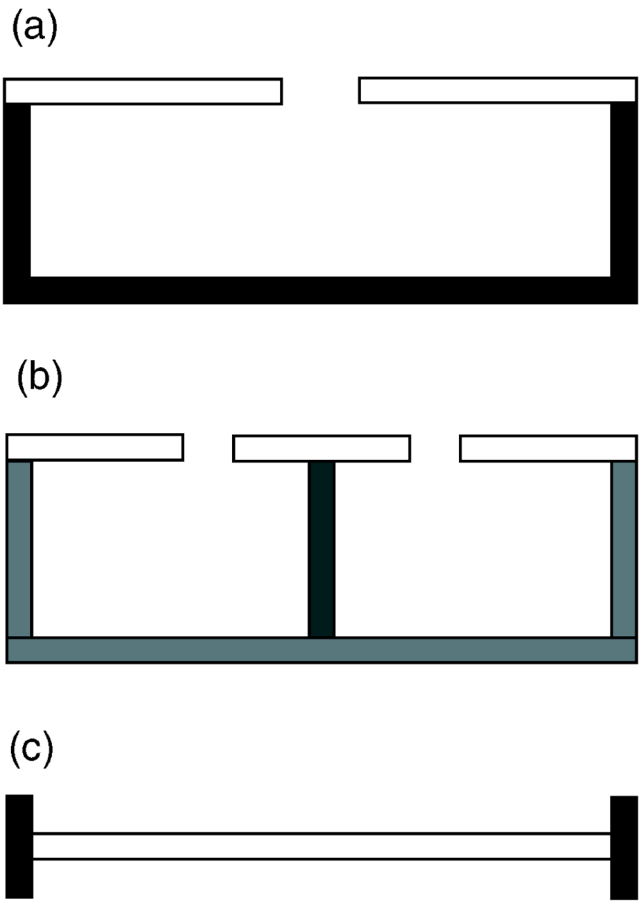


FIG. 1. (Color online) Cross-sectional schematics of typical structures of stringed instruments. (a) Box–sound-hole structure in the guitar, harpsichord, Cristofori's fortepiano, etc.. (b) Box–sound-hole–sound-post structure in the violin family (the sound post is depicted by a darker gray rectangle in the middle). (c) Soundboard—iron-frame structure in modern pianos. A white rectangle indicates top plate or soundboard; a black or gray rectangle indicates back (and side) plate or frame board.

Since the vibration of a wood plate produces sound radiation, c/ρ may be called the "vibration parameter" or "radiation ratio."[10,12]

Moreover, Schelleng[9] found a strong correlation between the resonant Q value and c/ρ. The Q value is the reciprocal of the loss factor, which is determined by the internal friction of wood. Since cellulosic microfibrils are highly crystalline in resonance wood, they have low damping, that is, high Q.[4,5] Although the Q is an important property of resonance woods for soundboards, it is not a useful parameter to classify soundboards and frame boards. The higher c/ρ the greater the vibration and radiation. Good radiators (soundboards) are good transmitters of vibration because the excitation is easily transmitted to the edge and corner. As a result, soft woods are used for soundboards, but hard woods are usually used for frame boards. Although hard woods are dense and vibrations are not easily excited, they are good transmitters of vibration and wave. For the back plate of the violin, and the body plate (and neck) of the Japanese shamisen (three-string instrument), good transmission is needed to make their sounds.

Thus, the characteristic acoustic transmission of woods is another important parameter for classifying woods for string instruments. If the attenuation (or damping) is relatively weak, the characteristic transmission is the reciprocal of the attenuation constant α of the longitudinal wave. The solution of the lossy wave equation gives $\alpha^{-1}=2Q/k=2cQ/\omega$ (k is the wave number; ω is the angular frequency).[16]

Barducci and Pasqualini[17] stressed the importance of the ratio c/Q and concluded that c/Q is independent of both direction and frequency. However, since c/Q has no physical acoustical significance, we consider the product cQ instead. Also, though the anisotropy is essential to the wood,[4–6,9–12] only the longitudinal wave that propagates along the wood grain is analyzed because it has the primary acoustical importance.

The acoustic conversion efficiency (ACE) proposed by Yankovskii[18] has also been used to characterize acoustic materials.[7,19,20] This ACE is the ratio of acoustic energy radiated from a beam to the vibration energy of the beam and is proportional to cQ/ρ. Thus ACE is simply Schelleng's vibration parameter (radiation ratio), c/ρ, multiplied by Q. Therefore, ACE has the same meaning as the radiation ratio,[7] and it is different from cQ or cQ/ω. It is thus proposed to use cQ to characterize the vibration transmission characteristic of wood. Obataya *et al.*[19] have also suggested that cQ (which they call the "relative acoustic conversion efficiency") be used to characterize woods. Their cQ was introduced to eliminate a strong dependence of ACE on ρ by assuming that ACE does not reflect the microstructure of the wood cell wall.

Since the Young's modulus E generally increases with density ρ, higher acoustic impedance $\rho c[=(\rho E)^{1/2}]$ might be required for frame boards.[21] On the other hand, excellent soundboards must have a high value of E/ρ.[5,7,8] Therefore the values of ρE and ρc alone do not predict the performance of soundboards and frame boards.

B. Transmission parameter and its measurement

Wood properties, such as c and Q, are usually measured by observing the first-mode bending vibration of strip-shaped sample plates with the free-free boundary condition.[19,22–24] Depending on the sample size, the frequency of the first mode is about 500 Hz. Wood parameters, such as Young's modulus and the attenuation, are almost frequency independent over the frequency range of 300 Hz to 1 kHz.[25] Thus it is appropriate to select data measured at around 500 Hz and to adopt cQ, instead of cQ/ω, as a measure of the characteristic transmission of wood because ω itself does not represent one of the wood properties. From now on we call cQ the "transmission parameter."

It must be noted that Q is measured by observing bending vibrations, while c is the longitudinal wave speed. The Q for longitudinal vibrations along the grain is required to exactly define the attenuation constant α of longitudinal waves along the grain. However, Q for vibrations along the grain is the same for longitudinal vibrations and bending vibrations if the frequency is the same and the mode frequency (and mode

TABLE I. Common names and botanical names of woods investigated in this paper.

Common name	Botanical name
Norway spruce	*Picea abies*
Sitka spruce	*Picea sitchensis*
Paulownia (Japanese kiri)	*Paulownia tomentosa*
Mulberry (Japanese kuwa)	*Morus alba*
Norway maple	*Acer platanoides*
Japanese maple (Japanese kaede)	*Acer sp.*
Amboyna wood (Japanese karin)	*Pterocarpus indicus*
Brazilian/Rio rosewood	*Dalbergia nigra*
White pine	*Pinus albicaulis*
Hemlock	*Tsuga sp.*
Redwood	*Sequoia sempervirens*
Western red cedar	*Thuja plicata*
Camphor wood (Japanese kusu)	*Cinnamomum camphora*
Zelkova (Japanese keyaki)	*Zelkova serrata*
Italian cypress	*Cupressus sempervirens*
Pear	*Pyrus communis*
American cherry	*Prunus serotina*
Black walnut	*Juglans nigra*
Andaman paduc	*Ptercarpus dalbergioides*
Balsa	*Ochroma pyramidale*

number) of vibration is not too high. This is because most bending deformations are due to the compression and dilatation in the longitudinal direction particularly in lower frequencies.

III. CLASSIFICATION DIAGRAM OF TRADITIONAL WOODS BEST SUITED FOR STRING INSTRUMENTS

The wood species investigated in this paper are summarized in Table I. Common names of woods are used hereafter. Traditional (meaning traditionally "best suited") woods for stringed instruments (the upper eight species in Table I) are first considered. Their wood constants are also summarized in Table II, where most numerical data on Western instruments are taken from Ref. 10, and those on Japanese instruments from Ref. 26. The measurement frequency is noted in these references. Unfortunately, the data at about 500 Hz are very limited.

TABLE II. Physical properties of traditional woods best suited for stringed instruments.

Wood name	f (Hz)	ρ (kg/m^3)	E (GPa)	c (m/s)	ρ/c (kgs/m^4)	Q	cQ (10^5 m/s)
Norway spruce[a]	532	560	16	5300	0.11	116	6.2
Sitka spruce[a]	484	470	12	5100	0.092	131	6.7
Sitka spruce[b]	617	408	10.0	4940	0.083	144	7.1
Paulownia[b]	569	260	7.3	5300	0.049	170	9.0
Mulberry	447	647	6.3	3130	0.21	70	2.2
Norway maple[a]	470	620	9.8	4000	0.16	85	3.4
Japanese maple[b]	447	695	11.8	4110	0.17	122	5.0
Amboyna wood[b]	519	873	20.0	4770	0.18	155	7.4
Brazilian/Rio Rosewood[a]	354	830	17	4400	0.19	185	8.1

[a]Reference 10.
[b]Reference 26.

TABLE III. Physical properties of substitute woods for stringed instruments.

Wood name	f (Hz)	ρ (kg/m^3)	E (GPa)	c (m/s)	ρ/c (kgs/m^4)	Q	cQ (10^5 m/s)
White pine[a]	514	380	10.0	5200	0.073	116	6.0
Hemlock[a]	533	440	8.4	4400	0.100	126	5.5
Redwood[a]	1067	380	9.5	5000	0.083	209	10.4
Western red cedar[a]	816	400	6.5	4000	0.100	174	7.0
Mulberry (M)	565	616	9.7	3960	0.16	121	4.8
Camphor wood[b]	497	550	9.0	4060	0.14	121	4.9
Zelkova[b]	439	720	12.6	4180	0.17	122	5.1
Italian cypress[c]	430	450	5.7	3560	0.13	97	3.5
Pear[a]	369	570	8.2	3800	0.15	67	2.5
American cherry[a]	795	700	12.0	4100	0.17	137	5.8
Black walnut[a]	522	680	20.0	5400	0.13	185	10.0
Andaman paduc[a]	474	710	12.0	4400	0.16	185	8.1
Balsa (high density)[b]	428	162	2.8	4170	0.039	140	5.8
Balsa (compressed)[b]	538	771	22.0	5150	0.15	163	8.4

[a]Reference 10.
[b]Reference 26.
[c]Reference 27.

Norway spruce and Sitka spruce are used for the violin top plate, the piano soundboard, the guitar top plate, etc. Paulownia (*kiri* in Japanese) is best for the Japanese 13-stringed long zither (*koto* or *soh*). Mulberry (*kuwa* in Japanese) is traditionally used for the whole body, including top and back plates, of the Japanese four-stringed lute (Satsuma *biwa*). Wood parameters ρ, E, and Q of mulberry were newly measured by Professor T. Ono of Gifu University using two samples (best quality and medium quality) provided by the Italian biwa maker, Doriano Sulis, living in Fukuoka, Japan. Measurements of E and Q were carried out by the free-free bending vibration method[19,22–24] mentioned in the previous section. The density ρ was measured after the sample was air dried (which usually leaves a moisture content of about 12%). The data for the best-quality and medium-quality samples are given in Tables II and III, respectively. Note that "mulberry (M)" is used in Table III to indicate the medium-quality sample.

On the other hand, the other four woods are used for frame boards. Both Norway and Japanese maple are used for the violin back plate. The amboyna wood (*karin* in Japanese) is best suited for the body and neck of the Japanese three-stringed lute, *shamisen*. Brazilian/Rio rosewood is traditionally best suited for the guitar back and side plates.

Note that Table II gives the reciprocal of Schelleng's vibration parameter c/ρ. Because ρ/c is roughly proportional to ρ we see that soft woods for soundboards are light while hard woods for frame boards are heavy. We may call ρ/c, which is a measure of the resistance to vibration, the "anti-vibration parameter." The higher the value, the greater the resistance to vibration. When transmission parameter cQ is plotted against ρ/c, a clear separation is seen between woods used for soundboards and woods used for frame boards (see Fig. 2). This shows the effectiveness of our newly proposed transmission parameter cQ.

In Fig. 2 the four points corresponding to soundboard woods (excluding mulberry) yield the regression line

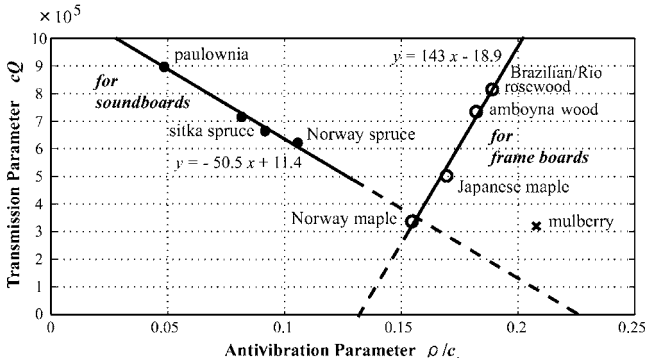


FIG. 2. Acoustical classification of traditional woods best suited for stringed instruments. The regression line for soundboards (full circles) is almost orthogonal to the regression line for frame boards (open circles). Note the position of mulberry, which is used for the top and back plates of the Japanese Satsuma biwa.

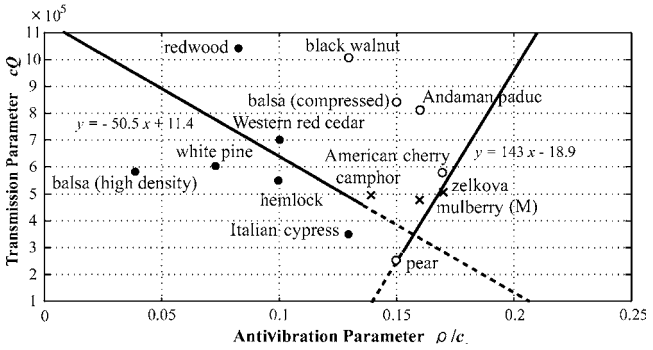


FIG. 3. A scatter diagram of substitute woods for stringed instruments. The regression lines given in Fig. 2 are drawn as the reference. The closer to the regression line that a proposed substitute material lies, the better its performance. Full circles indicate soundboard woods and open circles frame-board woods. Crosses show the positions of various substitutes used for mulberry, which is used for the best-quality Satsuma biwa.

$$y = -50.5x + 11.4, \tag{1}$$

and the four points corresponding to frame-board woods yield

$$y = 143x - 18.9, \tag{2}$$

where $x=\rho/c$ and $y=cQ/10^5$. These regression lines show a remarkable correlation with the different functions required by soundboards and frame boards. Soundboards require a strong positive correlation of transmission parameter cQ with vibration parameter c/ρ; frame boards require a strong positive correlation of cQ with antivibration parameter ρ/c. In other words, large values of c and Q are required by both good soundboards and good frame boards, while the ρ of soundboards should be much smaller than that of frame boards (cf. Table II). Although individual differences in wood samples must still be taken into account, it can be seen that the properties of traditional woods for string instruments given in Table II fall on one of two regression lines discussed above, and they fall naturally into one of two groups, either soundboard woods or frame-board woods.

Large values of c and Q along the grain are mainly due to the smaller fibril angles of the wood cell wall with respect to the grain direction.[19] Therefore, in a cell wall model of wood, small fibril angle seems to be the parameter that determines the performance of frame boards and soundboards. Thus the larger the transmission parameter cQ, the smaller the fibril angle.

Mulberry wood lies outside the traditional wood groups, and its position opposite Sitka spruce and Norway spruce with respect to line (2) for frame boards may seem puzzling. The idiosyncrasies of mulberry are discussed in Sec. V. It is interesting to note that Norway maple lies near the intersection of lines (1) and (2). Thus Norway maple has unique acoustical properties that make it suitable for both soundboards and frame boards. Such properties are very desirable for the back plate of the violin.

IV. CLASSIFICATION DIAGRAM FOR SUBSTITUTE WOODS

Substitute woods are woods that are not traditionally used for the best quality instruments, but alternative woods (including artificial composites) are often used in medium to high quality instruments. Table III lists the physical properties of typical substitute woods. Since Cristofori selected the Italian cypress and Italian poplar, respectively, for the soundboards and frame boards of his fortepianos, neither of these two woods has been used for modern pianos. Although, strictly speaking, these two woods are not substitute woods, it is, nevertheless, interesting to compare them with other woods. The Italian cypress data were courtesy of the French wood scientist Brémaud,[27] but data on the Italian poplar are, unfortunately, not available.

White pine and hemlock are applied to violin tops.[10] Redwood and Western red cedar are sometimes used for guitar tops,[10] although a majority of European guitar builders exclusively use Western red cedar as well as European spruce.[27] The data for mulberry (M) were obtained from new measurement of a sample used in medium-quality biwas. Camphor wood (*kusu* in Japanese) and zelkova (*keyaki* in Japanese) are very good for Japanese furniture and sometimes are used to make the biwa. The Canadian paulownia is used for medium-quality kotos, but its mechanical data were not available in the literature. The pear and American cherry are sometimes used for violin backs.[10] The black walnut and Andaman paduc may be used for guitar backs.[10] Balsa wood can be applied to the top and/or back plate of string instruments,[9] particularly the violin,[28] but its mechanical strength is questionable.[12]

From the data in Table III we construct the scatter diagram of Fig. 3, where the two regression lines given by Eqs. (1) and (2) are drawn for the reference. The closeness of data points to each regression line implies an excellent match to the "right wood." Hence the Western red cedar, white pine, and hemlock should be excellent substitutes for Sitka spruce and Norway spruce. Although the balsa with its high density has a very high value of the radiation ratio (very small antivibration parameter), it is not a good alternative for the top plate and soundboard. If its transmission parameter cQ (probably its Q value) could be increased (perhaps by the proper processing), balsa could be a promising substitute candidate for paulownia, which is best suited for the Japanese koto. On the other hand, redwood has too high a value

of cQ, and it seems less desirable as soundboard wood. However, this conclusion is based on data at high frequency—1067 Hz (see Table III). Italian cypress also has too high a value of ρ/c, making it unsuitable for the modern piano soundboard.

Concerning the frame-board substitutes, pear and American cherry may be excellent choices, although the pear data point is in the opposite direction of the regression line given by Eq. (1). On the other hand, black walnut, compressed balsa, and Andaman paduc are not such good alternatives.

The three data points for mulberry (M), camphor wood, and zelkova are very close to each other, but very far from the mulberry point of the best quality shown in Fig. 2. This result seems to endorse a clear difference between traditional wood and substitution wood. Nevertheless, the points of mulberry (M) and zelkova in Fig. 3 suggest that these woods might be excellent substitutes for frame-board woods.

As explained above, the classification diagram of Fig. 2 and the regression lines given by Eqs. (1) and (2) can be used to judge the suitability of substitute woods. In addition, they can guide the design of better quality artificial composite substitutes for stringed-instrument woods.

V. UNIQUE PROPERTIES OF JAPANESE STRING INSTRUMENTS

As shown in Fig. 2, mulberry, which is used for the best quality Japanese Satsuma biwa ("Satsuma" is the old name of the southernmost prefecture in Kyushu), lies far off the quality criteria for Western string instruments, given by the regression lines of Eqs. (1) and (2). Although this is partly because mulberry is used for both the top and the back plates, it nevertheless does seem extraordinary that mulberry has a very high value of the antivibration parameter and a very low value of the transmission parameter.

However, the poor vibrational properties of mulberry seem to match the playing style of the Satsuma biwa, in which the string is strongly struck with a large triangular wooden plectrum (*bachi* in Japanese) instead of ordinary plucking with a small pick as in the guitar. Striking the Satsuma biwa yields very characteristic impact tones. It should be noted that the peculiarity (low resonance nature) of mulberry makes this playing style possible because the top plate is simultaneously struck by a stroke of a large plectrum.

Also, since the mulberry body of the Satsuma biwa is a poor resonator, a mechanism has been invented to compensate. This mechanism is called the "sawari" ("touch"), which allows strings to vibrate against the neck or frets, creating a reverberating high-frequency emphasis.[3,29] A few variants of this sawari are seen as "jawari" (in Hindi) on the Indian sitar and tambura,[3,4] and as "bray pins" on medieval harps.[30]

Moreover, the peculiarity of the mulberry clearly discriminates the Satsuma biwa from the Chinese pipa (p′i-p′a). The Chinese pipa,[31] with a top plate of paulownia and a back shell of red sandal wood (close to amboyna wood) or maple, has many frets and is played with a small pick. This is quite similar to the Western guitar, although harmonic enhancement in the pipa is controlled by designing the top plate to have large resonances about one or two octaves higher than the fundamental frequencies of four strings (A_2, D_3, E_3, and A_3).[31]

The Japanese shamisen (meaning "three-stringed"), whose root is in the Chinese sanxian, has a long, unfretted wooden neck and a small body whose front and back are covered with white cat skin. The shamisen's neck and body are traditionally made of the amboyna wood, which is hard and does not readily vibrate but which transmits vibrations excellently (Fig. 2).

The shamisen also has the sawari, but its mechanism differs from that of the Satsuma biwa.[29] Most importantly, the shamisen's sawari is restricted to only the first (lowest) string. However, when the vibrations of the second and third strings are transmitted to the first string, the sawari effect is accompanied by appreciable sympathetic resonances if they are correctly tuned.[29] Receiving sawari on second and third strings serves as a reference to judge whether the tuning is correct or not. The amboyna wood is also the best material to facilitate the very smooth movement of fingers on the strings, and it very well transmits vibration once they have been excited.

Paulownia, used for the Japanese long zither (koto), is diametrically opposite to mulberry, used for the Satsuma biwa as indicated in Fig. 2. Thirteen strings, each about 1.5 m long, are stretched between two fixed bridges. In addition, a movable bridge is applied under each string for its tuning. The paulownia, with its very smooth surface, facilitates the movement of the bridge on the top plate when a chord change is required during the performance (this may also be the case for Korean gayageum and Chinese guzheng). Also, the koto strings are plucked with small plectra worn on three fingers (the thumb, index, and middle finger) of the right hand. Since this plucking is not so strong, and the koto body is very large, the material must support vibration well to maintain the sound. Thus, the high resonance of the paulownia seems to be a relevant requirement.

VI. CONCLUSIONS

A new scheme to classify the woods used in stringed instruments is proposed. Plotting ρ/c (the antivibration parameter) along the x axis, and cQ (the transmission parameter) along y axis, we obtain two regression lines that clearly discriminate soundboard woods from frame-board woods that are traditionally best suited for string instruments. These regression lines, defined by the traditional woods, constitute criteria to select substitute woods and synthesize artificial composites when making string instruments with the next best quality.

Since the propagation speed c of the longitudinal wave along the wood grain is determined by Young's modulus E and the density ρ, the wood classification diagram is based upon the fundamental physical quantities E, ρ, and Q. However, it should be remembered that the frequency used to measure these quantities is assumed to be almost the same for all wood samples. A frequency of around 500 Hz is practical for measurements.

Moreover, based on the classification diagram for traditional woods (Fig. 2), some peculiarities of Japanese string instruments such as the Satsuma biwa, shamisen, and koto emerge. In particular, mulberry (for the Satsuma biwa) and paulownia (for the koto) correspond to the opposite extremes on the regression line for soundboard woods. Such extreme wood properties are seldom seen in Western stringed instruments. Also, in comparing with the Chinese pipa, it may be understood that the unique playing style and the resulting tone of the Satsuma biwa depend strongly on mulberry's acoustical properties.

ACKNOWLEDGMENTS

The author would like to express his thanks to the prominent biwa maker, Doriano Sulis, for kindly providing mulberry samples. Professor Teruaki Ono of Gifu University measured the physical properties of the mulberry samples and provided helpful references during our discussions on wood science. Also, the author would like to thank Dr. Iris Brémaud of the University of Montpellier II, France, who provided many data on wood properties, and for stimulating discussions on European and Asian woods. She provided useful references and the correct botanical names of various woods. The author thanks Professor Eiichi Obataya of Tsukuba University, who patiently answered the author's questions on wood properties. The author also thanks Professor Thomas Rossing of Stanford University for his encouragement of considering Asian stringed instruments. Furthermore, the author thanks Professor James B. Cole of Tsukuba University for his careful and helpful editing of the English expressions. Finally, the author would like to show his acknowledgments to the anonymous reviewers and associate editor Neville Fletcher for their relevant comments and suggestions to improve the manuscript.

[1]S. Pollens, "The pianos of Bartolomeo Cristofori," J. Am. Musical Instrum. Soc. **10**, 32–68 (1984).
[2]S. Yoshikawa, "From the cembalo to the piano: Progress in early keyboard instruments," J. Acoust. Soc. Jpn. **57**, 704–711 (2001) (in Japanese).
[3]A. Odaka and S. Yoshikawa, "Acoustical characteristics of Chinese stringed instruments and their Asian relatives," J. Acoust. Soc. Am. **120**, 3118 (2006).
[4]N. H. Fletcher and T. D. Rossing, *The Physics of Musical Instruments*, 2nd ed. (Springer-Verlag, New York, 1998).
[5]V. Bucur, *Acoustics of Wood*, 2nd ed. (Springer-Verlag, Berlin, 2006).
[6]C. M. Hutchins and V. Benade (eds.), *Research Papers in Violin Acoustics 1975–1993* (Acoustical Society of America, Woodbury, NY, 1997), Vol. **2**, Chap. J. Wood.
[7]T. Ono, "Frequency responses of wood for musical instruments in relation to the vibrational properties," J. Acoust. Soc. Jpn. (E) **17**, 183–193 (1996).
[8]T. Ono and M. Norimoto, "On physical criteria for the selection of wood for soundboards of musical instruments," Rheol. Acta **23**, 652–656 (1984).
[9]J. C. Schelleng, "The violin as a circuit," J. Acoust. Soc. Am. **35**, 326–338 (1963).
[10]D. W. Haines, "On musical instrument wood," Catgut Acoust. Soc. Newsletters **31**, 23–32 (1979).
[11]I. Dunlop, "The acoustic properties of wood in relation to stringed musical instruments," Acoust. Aust. **17**, 37–40 (1989).
[12]C. Y. Barlow, "Materials selection for musical instruments," Proc. Instit. Acoust. **19**, 69–78 (1997).
[13]T. Ono, S. Miyakoshi, and U. Watanabe, "Acoustic characteristics of unidirectionally fiber-reinforced polyurethane foam composites for musical instrument soundboards," Acoust. Sci. and Technol. **23**, 135–142 (2002).
[14]D. W. Haines and N. Chang, "Application of graphite composites in musical instruments," Catgut Acoust. Soc. Newsletter **23**, 13–15 (1975).
[15]C. Besnainou, "From wood mechanical measurements to composite materials for musical instruments: New technology for instrument makers," MRS Bull. **20**, 34–36 (1995).
[16]E. Meyer and E.-G. Neumann, *Physical and Applied Acoustics*, translated by J. M. Taylor, Jr. (Academic, New York, 1972), pp. 14–15.
[17]I. Barducci and G. Pasqualini, "Measurement of the internal friction and the elastic constants of wood," Nuovo Cimento **5**, 416–466 (1948); translated by E. B. Abetti in a Benchmark Book Series, *Musical Acoustics, Part I Violin Family Components*, edited by C. M. Hutchins (Dowden, Hutchinson & Ross, Stroudsburg, Pennsylvania, 1975).
[18]B. A. Yankovskii, "Dissimilarity of the acoustic parameters of unseasoned and aged wood," Sov. Phys. Acoust. **13**, 125–127 (1967).
[19]E. Obataya, T. Ono, and M. Norimoto, "Vibrational properties of wood along the grain," J. Mater. Sci. **35**, 2993–3001 (2000).
[20]H. Aizawa, E. Obataya, T. Ono, and M. Norimoto, "Acoustic converting efficiency and anisotropic nature of wood," Wood Res. **85**, 81–83 (1998).
[21]H. Yano, Y. Furuta, and H. Nakagawa, "Materials for guitar back plates made from sustainable forest resources," J. Acoust. Soc. Am. **101**, 1112–1119 (1997).
[22]D. W. Haines, "On musical instrument wood—Part II Surface finishes, plywood, light and water exposure," Catgut Acoust. Soc. Newsletter **33**, 19–23 (1980).
[23]F. P. Kollmann and A. Cote, Jr., *Principles of Wood Science and Technology* (Springer-Verlag, New York, 1968), pp. 301–302.
[24]R. F. S. Hearmon, "The influence of shear and rotatory inertia on the free flexural vibration of wooden beams," Br. J. Appl. Phys. **9**, 381–388 (1958).
[25]E. Fukada, "The vibration properties of wood I," J. Phys. Soc. Jpn. **5**, 321–327 (1950).
[26]H. Aizawa, "Frequency dependence of vibration properties of wood in the longitudinal direction," master thesis, Faculty of Engineering, Kyoto University, 1998 (in Japanese).
[27]I. Brémaud, "Diversity of woods used or usable in musical instruments making: Experimental study of vibrational properties in axial direction of contrasted wood types mainly tropical—Relationships to features of microstructure and secondary chemical composition," Ph.D. thesis, Mechanics of Materials, University of Montpellier II, 2006 (in French).
[28]C. Waltham, personal communication on the balsa violin (February 2007).
[29]Y. Ando, *Acoustics of Musical Instruments*, 2nd ed. (Ongaku-no-tomo-sha, Tokyo, 1996), pp. 197–199, 202, 203, 208–210 (in Japanese).
[30]M. Campbell, personal communication (November 2006).
[31]S.-Y. Feng, "Some acoustical measurements on the Chinese musical instrument P′i-P′a," J. Acoust. Soc. Am. **75**, 599–602 (1984).

Density and sound speed of two gelatinous zooplankton: Ctenophore (*Mnemiopsis leidyi*) and lion's mane jellyfish (*Cyanea capillata*)

Joseph D. Warren[a)]
Marine Sciences Research Center, Stony Brook University, 239 Montauk Hwy, Southampton, New York 11968

Joy N. Smith
Department of Marine Science, Coastal Carolina University, Conway, South Carolina 29528

(Received 17 November 2006; revised 17 April 2007; accepted 21 April 2007)

The density and sound speed of two coastal, gelatinous zooplankton, *Mnemiopsis leidyi* (a ctenophore) and *Cyanea capillata* (lion's mane jellyfish), were measured. These parameters are important inputs to acoustic scattering models. Two different methods were used to measure the density of individual animals: one used a balance and graduated cylinder to determine the mass and displacement volume of the animal, the other varied the density of the solution the animal was immersed in. When the same animal was measured using both methods, density values were within 1% of each other. A travel-time difference method was used to measure the sound speed within the animals. The densities of both zooplankton slightly decreased as the animals increased in length, mass, and volume. The ratio of animal density and sound speed to the surrounding seawater (g and h, respectively) are reported for both animals. For *Mnemiopsis leidyi* ranging in length from 1 to 5 cm, the mean value (±standard deviation) of g and h were 1.009 (±0.004) and 1.007 (±0.001). For *Cyanea capillata* ranging in bell diameter from 2 to 11 cm, the mean value (±standard deviation) of g and single value of h were 1.009 (±0.004) and 1.0004. © *2007 Acoustical Society of America.* [DOI: 10.1121/1.2739433]

PACS number(s): 43.80.Cs, 43.30.Sf [KGF] Pages: 574–580

I. INTRODUCTION

Acoustic methods have been used for several decades to estimate the abundance and distribution of zooplankton and nekton in the marine environment (Foote and Stanton, 2000). These methods offer improved spatial coverage, vertical and horizontal resolution, and in some regions the ability to discriminate between different taxonomic groups. However, the acoustic methods provide indirect estimates of the biological scatterers in the water column, and the conversion of acoustic information to numerical abundance or biomass can be difficult. The process of transforming acoustic backscatter data into estimates of biologically relevant variables is known as the inverse problem and requires physics-based mathematical or empirical models that describe the acoustic scattering characteristics of a particular organism.

Ray-based models describing the scattering characteristics of several different types of zooplankton have undergone extensive development in the last few decades and are, in many cases, quite accurate when compared to empirical scattering data from individual animals (Stanton *et al.*, 1994; Stanton *et al.*, 1998). Although these models are not as accurate as full-diffraction solutions to the wave equation and require several initial assumptions, they can be expressed in relatively simple mathematical formulas. The inputs to these models include characteristics of the animals themselves such as: animal shape and size, orientation relative to the acoustic wave fronts, and body composition. The model output is typically the amount of backscattered acoustic energy that would be reflected by a scatterer. The shape and size of the animal can be assessed through net tow samples taken during an acoustic survey and the orientation of the animal within the water column can be evaluated through the use of video methods. In addition to these factors, two of the most important inputs into these models are the ratios of density and sound speed of the animal and its surrounding medium, seawater (Anderson, 1950; Johnson, 1977; Greenlaw and Johnson, 1982; Stanton and Chu, 2000; Chu *et al.*, 2000). The density ratio is g ($g=\rho_a/\rho_{sw}$) and the sound speed ratio is h ($h=c_a/c_{sw}$) where ρ is the density and c is the speed of sound and the subscript indicates the property is of the organism (a) or surrounding seawater (sw).

There have been several studies that have measured the material properties of fluid-like scatterers such as copepods and euphausiids (Greenlaw and Johnson, 1982; Foote, 1990); but few, if any, measurements have been made of gelatinous zooplankton. Two very common estuarine and coastal gelatinous zooplankton are ctenophores (*Mnemiopsis leidyi*) and lion's mane jellyfish (*Cyanea capillata*). Ctenophores are found throughout the temperate and tropical coastal oceans (Mianzan, 1999) and have become an important invasive species in the Black and Caspian Seas (Mutlu, 1999; Ivanov *et al.*, 2000) where they were one of the severe impacts on the local ecosystem (Kideys, 1994; Shiganova and Bulgakova, 2000). The lion's mane jellyfish and other related scy-

a)Author to whom correspondence should be addressed. Electronic mail: joe.warren@stonybrook.edu

 0001-4966/2007/122(1)/574/7/$23.00

phozoans are seasonally abundant in northern waters (Houghton *et al.*, 2006) and may have an important impact on the marine ecosystem (Broduer *et al.*, 2002). Increases in the distribution and abundance of gelatinous zooplankton have been linked to shifts in ecosystems due to climatic variation (Purcell, 2005; Lynam *et al.*, 2005; Mills, 2001) as well as removal of apex predators from the ecosystem due to fishing pressures (Purcell and Arai, 2001).

There have been several studies that have used acoustics to measure the scattering from ctenophores (Alvarez Colombo *et al.*, 2003) and scyphozoan jellyfish (Båmstedt *et al.*, 2003; Brierley *et al.*, 2004, 2005). Gelatinous animals are thought to be weak scatterers given their fluid-like body composition and ability to remain neutrally buoyant in the water column, and on a per unit biomass basis they are weaker scatterers than crustacean or shelled gastropod zooplankton (Stanton *et al.*, 1996). However, since the gelatinous animals are often much larger (both in size and biomass) than the other zooplankton in the water column, they have relatively large target strength values on a per animal basis (e.g. Table II in Alvarez Colombo *et al.*, 2003).

While there has been some progress in the development of a scattering model for gelatinous zooplankton (Monger *et al.*, 1998; Mutlu, 1996), several studies of zooplankton scattering (Ressler, 2002; Stanton *et al.*, 1994) have calculated gelatinous zooplankton contributions using a model developed for fluid-like crustacean zooplankton. In both cases, several model inputs (particularly g and h) have been estimates based on fits of scattering model predictions to measurements of target strength rather than direct measurements. As a first step in improving the ability to accurately model the scattering from gelatinous zooplankton, accurate measurements of the density and sound speed of the animals are needed.

II. METHODS

Ctenophores and lion's mane jellyfish were collected from Shinnecock and Peconic Bay on Long Island, New York during June–August 2006. Animals were collected using a small dip net and buckets. While some gelatinous animals cannot be handled without causing extensive damage to their bodies, the ctenophores and lion's mane jellyfish are relatively hardy animals and were not noticeably harmed by the collection process. In order to minimize the possibility of measuring damaged animals, the animals were placed in buckets or tanks containing water similar in temperature (24–28 °C) and salinity (25–26) to the water they were collected in. A temperature and salinity sensor (YSI 85, YSI Incorporated) was used to measure the properties of the water used in the experiment from which seawater density (ρ_{sw}) and sound speed (c_{sw}) were calculated. Seawater density and sound speed were calculated using the CSIRO MatLAB Seawater Library for each experiment and varied little between experiments ($1.016-1.0185$ g ml^{-1} and $1521-1532$ m s^{-1}). The animals were monitored for several hours or days and only animals that appeared to be healthy and mobile were used in the experiments. The animals were alive throughout the experiments.

A. Pipette method for density measurements

Measurements of an individual animal's density were made using two different methods. Individual animals were measured in order to ascertain the range of density values for each type of animal. The pipette method is a slight modification of the method used by Lowndes (1942). It consists of taring a graduated cylinder (either 50, 100, or 1000 ml depending on the size of the animal) containing a known volume of seawater (V_o) on an electronic balance. The animal was removed from the bucket and placed on paper towels where the excess water from the animal was removed. The animal was then placed into the graduated cylinder and the mass of the animal (m_a) was recorded. With the animal remaining in the graduated cylinder, a pipette was used to remove water from the cylinder until the total volume (animal and remaining water) was equal to the original volume in the cylinder. The water that was removed from the cylinder was then weighed on the balance (m_{rw}) and its volume calculated. The volume of extracted water is equal to the animal's volume (V_a). Weighing the extracted water in order to calculate the volume was a more accurate method for finding the volume of the animal than visually recording the water displaced on the graduated cylinder after the animal had been added. The resolution of the graduated cylinder was poor because a large cylinder was needed to hold the gelatinous animals. With these measurements, the density of the animal was calculated (Eq. (1))

$$\rho_a = \frac{m_a}{V_a} = \frac{m_a}{\left(\frac{m_{\mathrm{rw}}}{\rho_{\mathrm{sw}}}\right)}. \tag{1}$$

This method was performed a minimum of three times on each individual to determine the precision of the method. The mean value of the measurements that were within 15% of the mean value of all measurements was then used for the rest of the data analysis. In approximately 10% of the measurements, one of the three measurements was substantially (more than 15% different from the mean) different than the other two. The presence of these outliers decreased as more measurements were made suggesting that familiarity and experience with the measurement technique is essential for accurate measurements of the mass and volume of the animal. This method requires the use of an electronic balance but could be modified to work at sea if a linked dual-balance method was used (Childress and Mickel, 1980; Chu and Wiebe, 2005).

There are several potential sources of errors with this method including: inadequate removal of excess water from the gelatinous animals and damaging the animal in the process. Since animal density measurements were the goal of this experiment, most of the water trapped internally by the ctenophores was removed by squeezing the animal throughout the drying process. As a consequence, pieces of the ctenophore were occasionally removed by the paper towel causing the mass and volume to slightly decrease (by approximately 5%) between each trial and possibly affecting the density measurements. An additional concern was that air bubbles might be introduced on or in the animal during the

removal of excess water and placement of the animal in the graduated cylinder. However, air bubbles were not observed on or in the animal when it was put into the cylinder. While air bubbles are a great concern for target strength measurements of individual animals (as bubbles are quite strong scatterers), the presence of bubbles on the animals for these measurements would result in only small errors in density or sound speed.

B. Titration method for density measurements

Another method for measuring the animal's density was used that could be easily applied to measurements made at sea where ship motion often causes difficulty when using an electronic balance. This method is a modification of that used by others (Greenlaw, 1977; Greenlaw and Johnson, 1982) where the density of animals was measured by placing them in a series of solutions of varying density and noting the solutions where the animal became buoyant and nonbuoyant. While this method can be done at sea without the use of a balance, the resolution of this method is a function of the number of solutions available. In order to improve the resolution of the density measurements in this study, an animal was placed into a beaker containing a known volume of seawater (V_o) of the same density as the zooplankton were collected in. Two burets were mounted on a ring stand above the beaker. One contained a solution of saltwater with a higher density (ρ_{ss}) than the animal's normal environment and the other contained normal seawater (ρ_{sw}). The saltier water was added to the beaker until the animal began to rise and float at the surface (indicating that the solution's density was greater than that of the animal), then the normal seawater was added until the animal began to sink (at which point the solution's density was just slightly less than that of the animal). While the solutions were being added to the beaker, it was necessary to stir or mix the solution. By measuring the volumes of saltier water (V_{ss}) and normal seawater (V_{sw}) added to the beaker, the animal's density can be found (Eq. (2)).

$$\rho_a = \frac{(V_o + V_{sw})\rho_{sw} + V_{ss}\rho_{ss}}{V_o + V_{sw} + V_{ss}}. \tag{2}$$

The process of adding saltier and normal water can proceed for several iterations if an increased precision of animal density is desired, however only one addition of saltier and normal seawater was used for this study. Potential inaccuracies in this method would result from errors in measuring the volume of water added, poor mixing of the solution in the beaker leading to stratification and, most importantly, the possibility that the addition of saltier water will cause the animal to adjust to the new osmotic pressure. If these measurements are made rapidly, it is thought that this error will be minimized, however it is not known at what time scales these changes may occur. This method is more time consuming than the pipette method, so only one measurement using this method was made for each animal.

C. Sound speed measurements

Similar to previous measurements of sound speed in zooplankton, the travel-time difference method was used (Greenlaw and Johnson, 1982; Foote, 1990). This method uses the difference in travel time for an acoustic wave traveling from a transmitter to a receiver through a pathway that either contains a known volume of zooplankton or is empty. This ray-based method is applicable when the acoustic wavelength used is much smaller than the characteristic size of the scatterers, which is close to the length scales of this study where the acoustic wavelength (8 mm) is less than the typical animal dimensions (~3–10 cm). A t tube (PVC plumbing T joint of various sizes) was used with a 192 kHz transducer clamped to each end and the entire apparatus submerged in seawater. Measurements of the travel time for an acoustic wave from the transmitter to the receiver were made with the chamber empty and with animals inside the chamber. To maximize the accuracy of this method, the animals should fill the t tube as much as possible so that the acoustic pathway is most likely to contain animal tissue rather than seawater. Because of the size and number of animals placed in the tube, we are confident that animal tissue was distributed throughout the volume of the t tube.

The t-tube volume was measured (V_{tube}) as well as the volume of animals that were put inside the t tube (V_a). The transmitted wave form was a 5 μs long, 192 kHz sinusoid signal. A data acquisition program was used to transmit the wave form to the signal generator, then to a power amplifier, and finally to the transmitting transducer. The receiving transducer was connected to a digital oscilloscope which digitized the received wave form and the signal was recorded. The difference in travel time between the signals recorded when the chamber was empty and contained animals was measured (Δt) as well as the travel time from transmitter to receiver when the chamber was empty (t_d). Both measurements were used to calculate h, the speed of sound of the animal (c_a) relative to that of seawater (c_{sw}) (Eq. (3)).

$$h = \frac{c_a}{c_{sw}} = 1 + \frac{\Delta t}{\frac{V_{tube}}{V_a} t_d}. \tag{3}$$

This method and equation have been used in previous studies (Chu and Wiebe, 2005) and provides an accurate measure of the sound speed of fluid-like zooplankton; however, it is very difficult to get measurements for an individual animal since ideally the t-tube chamber should be completely filled with animals. Because of this constraint, measurements of sound speed were made for several ctenophores at a time and therefore only average measurements of sound speed for a group of ctenophores are reported.

It should be noted that the techniques used in this study produce values of density or sound speed based on the assumption that the animal has uniform material properties throughout its body. This is not the case for either animal and particularly for the lion's mane jellyfish simply holding the animal makes these differences readily apparent. Without separating the different organelles or structures from each other, these methods cannot provide information about how the material properties vary within a specific animal. Specific parts of other gelatinous zooplankton have been shown to

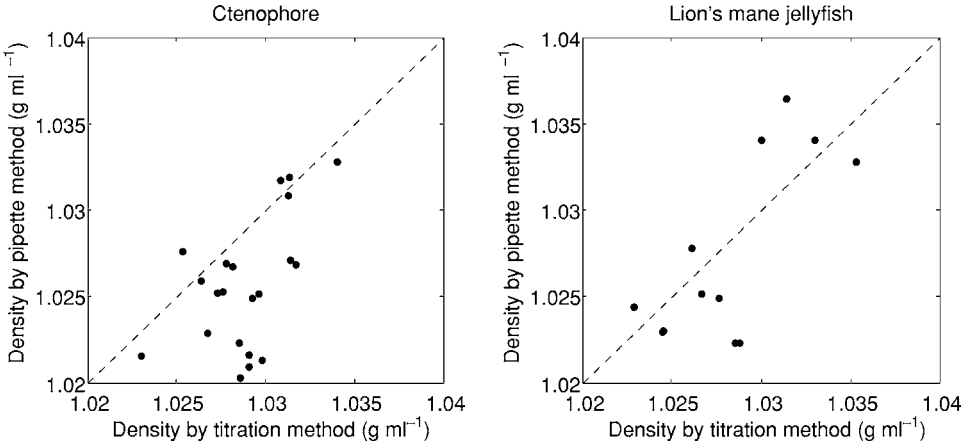


FIG. 1. Comparison of the pipette and titration method measures of animal density for individual ctenophores (left) and lion's mane jellyfish (right). While there is good agreement between the two methods, the titration method tends to produce a higher measure of ctenophore density than the pipette method which may be due to the animal adjusting to the increased density of the surrounding seawater.

have different scattering efficiencies (Stanton *et al.*, 1998; Warren *et al.*, 2001) and that is likely true for these animals as well.

III. RESULTS

The pipette method was used to measure the density of 21 individual ctenophores after becoming proficient with the procedure such that replicate measurements of the same animal were similar. These same animals also had their density measured using the titration method. One measurement per animal was made using the titration method. Density values ranged from 1.0203 to 1.0328 g ml^{-1} with the pipette method and from 1.0230 to 1.0340 g ml^{-1} for the titration method, although the mean values for each method were similar: 1.0257 (pipette) and 1.0289 (titration) g ml^{-1} (Fig. 1). A paired t test showed that the two measurement methods were statistically different ($n=21$, $t=4.5831$, $p<0.001$), however the means of the two distributions differ by only 0.003 g ml^{-1} which is small relative to the range of densities measured. The mean (and range) of g for the ctenophores was found to be: 1.0071 (1.0018–1.0140) for the pipette method and 1.0102 (1.0045–1.0153) for the titration method. Using the small sample size method, we calculated the 90% confidence interval for the mean value of g for both the pipette (1.0057–1.0085) and titration (1.0093–1.0111) methods using the critical value of a Student's t distribution ($n=21$).

Twelve lion's mane jellyfish were measured using both the pipette and titration methods. As with the ctenophores, replicate measurements using the pipette method were similar and only one titration measurement was made per animal. Jellyfish density ranged from 1.0223 to 1.0364 g ml^{-1} with the pipette method and from 1.0229 to 1.0353 g ml^{-1} for the titration method, although the mean values for each method were similar: 1.0275 (pipette) and 1.0283 (titration) g ml^{-1} (Fig. 1). A paired t test showed that the two measurement methods were statistically similar ($n=12$, $t=-0.7678$, $p=0.46$). The mean (and range) of g for the lion's mane jellyfish was found to be: 1.0088 (1.0037–1.0176) for the pipette method and 1.0096 (1.0043–1.0165) for the titration method. Using the small sample size method, we calculated the 90% confidence interval for the mean value of g for both the pipette (1.0062–1.0115) and titration (1.0077–1.0115) methods using the critical value of a Student's t distribution ($n=12$).

A comparison of the two density measurements for each animal shows that there is overall good agreement between the two techniques (Fig. 1) although the titration method generally results in densities that are larger than the densities from the pipette method. Two potential explanations for this discrepancy are that the pipette method will tend to underestimate the true density of the animal or that the titration method will tend to overestimate the true density of the animal. If there is excess water remaining on the animal before the mass and volume (of the animal and excess water) are measured, then the measurement will be less than the density of the animal since some fraction of the mass and volume being measured will have the density of seawater rather than of the animal. Conversely, the titration method which increases the density of the solution that the animal is in may result in a bias towards higher density measurements if the animal is able to adjust (via osmotic processes) to the new solution by increasing its own density.

Some species have been shown to have a relationship between the size of the animal and their material properties (e.g., Antarctic krill, Chu and Wiebe (2005)), so regressions between animal size, mass, and volume and animal density were calculated (Fig. 2, Table I). Because density, mass, and volume are all three-dimensional quantities, linear regressions between these variables were calculated. However, animal size (length) is a one-dimensional quantity and in order to compare quantities that are of different dimension, the data were fit to a power law relationship. The density values used in the analysis were from the pipette method. Density of individual ctenophores decreased as length of the animal increased. The range of lengths of animals measured is quite small, 1–5 cm, so this relation may not extrapolate to larger animals. A similar relationship between animal size and animal density was found for the lion's mane jellyfish. The range of bell diameters in this data set was from 2 to 11 cm. Similar relationships were found between density and either the mass or volume of the animal for both types of gelatinous zooplankton. This is not surprising considering that animal size, mass, and volume were strongly correlated with each other ($R^2>0.95$ for all combinations).

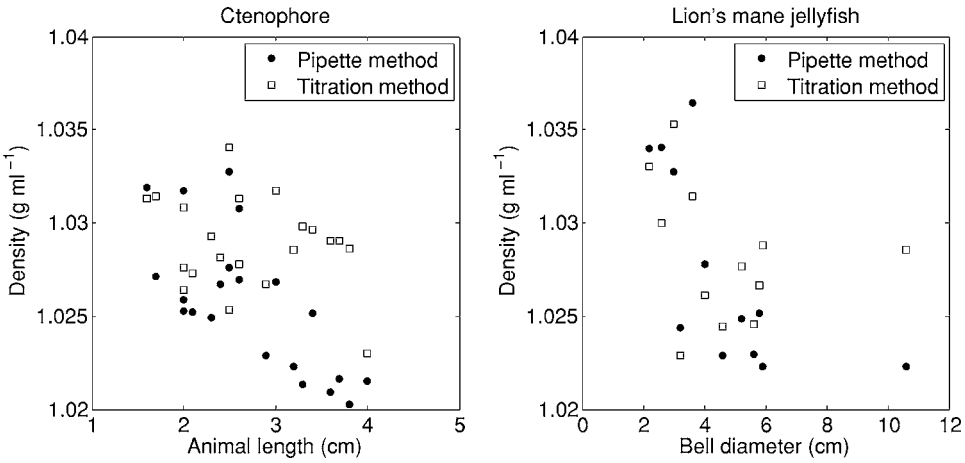


FIG. 2. Density of individual ctenophores (left) and lion's mane jellyfish (right) measured by the pipette (circles) and titration (squares) methods compared to animal size. Both methods show that as animal size increases animal density decreases. A similar negative relationship was found between animal mass or volume and animal density.

Measurements of h were made on three different groups of ctenophores containing 5, 7, and 10 animals, respectively. The volume of the empty t tube was 86 ml and the volume of the animals placed in the tube was 79, 70, and 60 ml, respectively. The value of h for each the three groups was 1.0066, 1.0079, and 1.0077 with an average value of h for all groups of 1.0074. Replicate measurements (with the chamber empty and full) were made for each trial. The mean value of the second set of measurements (1.0073) was nearly the same as the first trial, although the range of h values was larger (h =1.0106, 1.0056, and 1.0058). It is not known if the increased range of h values measured is due to the additional handling of the animals, although it is reassuring that the mean values between trials were very similar.

The h value for a single lion's mane jellyfish was also measured with this method, with h being 1.0004. The lion's mane jellyfish were larger than the ctenophores such that a larger t tube (with a volume of 241 ml) was used. Even with the larger sampling chamber, the lion's mane jellyfish were of such volume and shape that only one animal was able to fit in the t tube and occupied approximately one quarter of the t-tube volume. This animal was significantly damaged after removal from the sampling apparatus so only one h measurement was able to be made. No other lion's mane jellyfish that were collected were able to be placed in the sample chamber with the current experimental setup. The h value reported here should be viewed with appropriate skepticism as it is a single measurement from one animal and the volume of the animal was small compared to the t-tube volume.

IV. DISCUSSION

While it is not surprising that gelatinous zooplankton have g and h values that are much closer to unity than other fluid-like zooplankton (e.g., crustaceans), the variability in these values is surprising. There were relationships between g and the size of the animal, but relationships between h and animal size were unable to be tested in this experiment. It is unknown whether there are also seasonal or daily changes in these parameters, nor the effect that food availability, predation rate, or energetic expenditures by the animal may have on these parameters.

Previous studies of scattering by gelatinous animals (Mutlu, 2005; Ressler, 2002) have used an acoustic reflection coefficient (R=0.056) that was determined empirically by fitting backscatter target strength data to a theoretical scattering model (Monger *et al.*, 1998). The reflection coefficient can be related to g and h by $R=(gh-1)/(gh+1)$. Applying the mean values of g and h found in this study (g=1.009, h=1.007) results in a reflection coefficient that is seven times smaller (R=0.008). If the smaller reflection coefficient from this study is used instead of the Monger *et al.* (1998) value,

TABLE I. Regression coefficients, residuals, and sample size for comparisons of animal density (ρ) with animal size, mass, and volume. Three-dimensional quantities (density, mass, and volume) were linearly related, but because animal size is a one-dimensional quantity a power law relationship was used to compare size and density. For the linear regression equation $y=bx+a$, the dependent variable (y) corresponds to density and the independent variable (x) corresponds to either mass or volume. For the power law equation $y=cx^d$ the dependent variable (y) corresponds to density and the independent variable (x) corresponds to animal size. Density values are from the pipette method. NA means not applicable.

	Ctenophores (n=21)			Lion's mane jellyfish (n=12)		
	ρ, length	ρ, mass	ρ, volume	ρ, bell diameter	ρ, mass	ρ, volume
b	NA	−0.002	−0.002	NA	−0.0003	−0.0003
a	NA	1.031	1.031	NA	1.031	0.031
d	−0.0096	NA	NA	−0.0091	NA	NA
c	1.0351	NA	NA	1.0411	NA	NA
R^2	0.49	0.58	0.59	0.58	0.49	0.49
Standard error	2.50E-4	5.33E-4	5.31E-4	4.20E-4	1.10E-3	1.10E-3

then numerical densities of gelatinous animal would increase by at least an order of magnitude as the animal's backscattering cross-section (σ_{bs}) is a function of the reflection coefficient squared (Monger *et al.*, 1998).

Using the minimum measured values of g and h (1.0018 and 1.005) for a ctenophore produces a reflection coefficient of 0.0034 and if the maximum values of g and h (1.014, 1.0079) are used the reflection coefficient is 0.011. While these values are still much smaller than those found by Monger *et al.* (1998), the variability in these values would also affect numerical abundance estimates of gelatinous animals. These data also raise the possibility that modelers of acoustic scattering from zooplankton will need to use a distribution of values for g and h in their models rather than a single value as is commonly done. Distribution functions of zooplankton length and orientation are being used more frequently in acoustic modeling efforts (e.g., Demer and Conti, 2005; Lawson *et al.*, 2006) and distribution functions describing the range of measured g and h values should be used as well.

The values of g and h found in this study are much smaller (roughly half) of those found empirically by Monger *et al.* (1998). The animals used in this study are different species than those in the Monger *et al.* (1998) study although the size of the animals are similar. The ctenophores and lion's mane jellyfish were also found in water that was probably less dense (i.e., warmer and fresher) than Puget Sound (site of the Monger *et al.* (1998) study). It is unknown whether gelatinous animals adjust their material properties relative to their environment, although it is likely that they have this ability since the species in this study are found in a wide range of environments from shallow estuarine bays (seawater density ~ 1.008 g ml^{-1}) to off-shore oceanic conditions (seawater density ~ 1.024 g ml^{-1}). The values of g and h found in this study may only apply to animals that are in a similar estuarine environment and investigation of whether g and h are dependent on the density of the surrounding seawater is needed.

Acoustic techniques offer one assessment method for measuring the distribution and abundance of gelatinous zooplankton, however accurate scattering model inputs are important for accurate predictions of animal numerical abundance. Given that many studies have attributed relatively high target strengths to gelatinous scatterers (Table II in Alvarez Colombo *et al.*, 2003), it is important to have accurate scattering models and model inputs that support these empirical observations.

ACKNOWLEDGMENTS

The authors gratefully acknowledge the assistance of the Stony Brook University Southampton Marine Station, particularly Melanie Meade and Luke Ormand. Alexandre Nicholas, Jordan Mertes, and Krissy Forman assisted with measurements and sample collection. Bradley Peterson assisted with statistical calculations. Gareth Lawson and three anonymous reviewers provided insightful comments and useful feedback on this work. This work was supported in part by the National Science Foundation, Grant No. OPP-0633939. This is contribution 1334 of the Marine Sciences Research Center at Stony Brook University.

Alvarez Colombo, G., Mianzan, H., and Madirolas, A. (**2003**). "Acoustic characterization of gelatinous-plankton aggregations: Four case studies from the Argentine continental shelf," ICES J. Mar. Sci. **60**, 650–657.

Anderson, V. C. (**1950**). "Sound scattering from a fluid sphere," J. Acoust. Soc. Am. **22**, 426–431.

Båmstedt, U., Kaartvedt, S., and Youngbluth, M. (**2003**). "An evaluation of acoustic and video methods to estimate the abundance and vertical distribution of jellyfish," J. Plankton Res. **25**, 1307–1318.

Brierley, A. S., Axelsen, B. E., Boyer, D. C., Lynam, C. P., Didcock, C. A., Boyer, H. J., Sparks, C. A. J., Purcell, J. E., and Gibbons, M. J. (**2004**). "Single-target echo detections of jellyfish," ICES J. Mar. Sci. **61**, 383–393.

Brierley, A. S., Boyer, D. C., Axelsen, B. E., Lynam, C. P., Sparks, C. A. J., Boyer, H. J., and Gibbons, M. J. (**2005**). "Towards the acoustic estimation of jellyfish abundance," Mar. Ecol.: Prog. Ser. **295**, 105–111.

Broduer, R. D., Sugisaki, H., and Hunt, Jr., G. L. (**2002**). "Increases in jellyfish biomass in the Bering Sea: Implications for the ecosystem," Mar. Ecol.: Prog. Ser. **233**, 89–103.

Childress, J. J., and Mickel, T. J. (**1980**). "Instruments and methods: A motion compensated shipboard precision balance system," Deep-Sea Res. **27**, 965–970.

Chu, D., Wiebe, P., and Copley, N. (**2000**). "Inference of material properties of zooplankton from acoustic and resistivity measurements," ICES J. Mar. Sci. **57**, 1128–1142.

Chu, D., and Wiebe, P. H. (**2005**). "Measurements of sound-speed and density contrasts of zooplankton in Antarctic waters," ICES J. Mar. Sci. **62**, 818–831.

Demer, D. A., and Conti, S. G. (**2005**). "New target-strength model indicates more krill in the Southern Ocean," ICES J. Mar. Sci. **62**, 25–32.

Foote, K. G. (**1990**). "Speed of sound in *Euphausia superba*," J. Acoust. Soc. Am. **87**, 1405–1408.

Foote, K. G., and Stanton, T. K. (**2000**). "Acoustical methods," in *ICES Zooplankton Methodology Manual*, edited by eds. R. Harris, P. Wiebe, J. Lenz, H. Skjoldal, and M. Huntley (Academic, London), Chap. 6, pp. 223–258.

Greenlaw, C. F. (**1977**). "Backscattering spectra of preserved zooplankton," J. Acoust. Soc. Am. **62**, 44–52.

Greenlaw, C. F., and Johnson, R. K. (**1982**). "Physical and acoustical properties of zooplankton," J. Acoust. Soc. Am. **72**, 1706–1710.

Houghton, J. D. R., Doyle, T. K., Davenport, J., and Hays, G. C. (**2006**). "Developing a simple, rapid method for identifying and monitoring jellyfish aggregations from the air," Mar. Ecol.: Prog. Ser. **314**, 159–170.

Ivanov, V., Kamakin, A. M., Ushivtzev, V. B., Shiganova, T. A., Zhukova, O., Aladin, N., Wilson, S. I., Harbison, G. R., and Dumont, H. J. (**2000**). "Invasion of the Caspian Sea by the comb jellyfish *Mnemiopsis leidyi* (Ctenophora)," Biol. Invasions **2**, 255–258.

Johnson, R. K. (**1977**). "Sound scattering from a fluid sphere revisited," J. Acoust. Soc. Am. **61**, 375–377.

Kideys, A. E. (**1994**). "Recent dramatic changes in the Black Sea ecosystem: The reason for the sharp decline in Turkish anchovy fisheries," J. Mar. Syst. **5**, 171–181.

Lawson, G. L., Wiebe, P. H., Ashjian, C. J., Chu, D., and Stanton, T. K. (**2006**). "Improved parameterization of Antarctic krill target strength models," J. Acoust. Soc. Am. **119**, 232–242.

Lowndes, A. G. (**1942**). "The displacement method of weighing living aquatic organisms," J. Mar. Biol. Assoc. U.K. **25**, 555–574.

Lynam, C. P., Hay, S. J., and Brierley, A. S. (**2005**). "Jellyfish abundance and climatic variation: Contrasting responses in oceanographically distinct regions of the North Sea, and possible implications for fisheries," J. Mar. Biol. Assoc. U.K. **85**, 435–450.

Mianzan, H. W. (**1999**). "Ctenophora," in *South Atlantic Zooplankton*, edited by D. Boltovskoy (Backhuys, Leiden, The Netherlands), pp. 561–573.

Mills, C. E. (**2001**). "Jellyfish blooms: Are populations increasing globally in response to changing ocean conditions," Hydrobiologia **451**, 55–68.

Monger, B. C., Chinniah-Chandy, S., Meir, E., Billings, S., Greene, C. H., and Wiebe, P. H. (**1998**). "Sound scattering by the gelatinous zooplankters *Aequorea victoria* and *Pleurobrachia bachei*," Deep-Sea Res., Part II **45**, 1255–1271.

Mutlu, E., (**1996**). "Target strength of the common jellyfish (*Aurelia aurita*): A preliminary experimental study with a dual-beam acoustic system," ICES J. Mar. Sci. **53**, 309–311.

Mutlu, E., (**1999**). "Distribution and abundance of ctenophores and their

zooplankton food in the Black Sea. II. *Mnemiopsis leidyi*," Mar. Biol. (Berlin) **135**, 603–613.
Mutlu, E., (**2005**). "A comparison of the contribution of zooplankton and nekton taxa to the near-surface acoustic structure of three Turkish seas," Mar. Ecol.—Evol. Persp. **26**, 17–32.
Purcell, J. E., (**2005**). "Climate effects on formation of jellyfish and ctenophore blooms: A review," J. Mar. Biol. Assoc. U.K. **85**, 461–476.
Purcell, J. E., and Arai, M. N. (**2001**). "Interactions of pelagic cnidarians and ctenophores with fish: A review," Hydrobiologia **451**, 27–44.
Ressler, P. H. (**2002**). "Acoustic backscatter measurements with a 153 kHz ADCP in the northeastern Gulf of Mexico: Determination of dominant zooplankton and micronekton scatterers," Deep-Sea Res., Part I **49**, 2035–2051.
Shiganova, T. A., and Bulgakova, Y. V. (**2000**). "Effects of gelatinous plankton on Black Sea and Sea of Azov fish and their food resources," ICES J. Mar. Sci. **57**, 641–648.
Stanton, T. K., and Chu., D. (**2000**). "Review and recommendations for the modeling of acoustic scattering by fluid-like elongated zooplankton: Euphausiids and copepods," ICES J. Mar. Sci. **57**, 793–807.
Stanton, T. K., Chu, D., and Wiebe, P. H. (**1996**). "Acoustic scattering characteristics of several zooplankton groups," ICES J. Mar. Sci. **53**, 289–295.
Stanton, T. K., Chu, D., and Wiebe, P. H. (**1998**). "Sound scattering by several zooplankton groups. II Scattering models," J. Acoust. Soc. Am. **103**, 236–253.
Stanton, T. K., Wiebe, P. H., Chu, D., Benfield, M. C., Scanlon, L., Martin, L., and Eastwood, R. L. (**1994**). "On acoustic estimates of zooplankton biomass," ICES J. Mar. Sci. **51**, 505–512.
Warren, J. D., Stanton, T. K., Benfield, M. C., Wiebe, P. H., Chu, D., and Sutor, M. (**2001**). "*In situ* measurements of acoustic target strengths of gas-bearing siphonophores," ICES J. Mar. Sci. **58**, 740–749.

Anisotropy of the backscatter coefficient of formalin-fixed ovine myocardium

Min Yang, Todd M. Krueger, Mark R. Holland, and James G. Miller[a)]
Department of Physics, Washington University, One Brookings Drive, Campus Box 1105, St. Louis, Missouri 63130

(Received 10 January 2007; revised 16 April 2007; accepted 17 April 2007)

The objective of this study was to measure the backscatter coefficient of formalin-fixed myocardial tissue as a function of angle of insonification relative to the myocardial fiber direction. Backscatter measurements were performed on eight cylindrical formalin-fixed lamb myocardial specimens and compensated for attenuation and diffraction effects to determine the backscatter coefficient. The backscatter coefficient at 5 MHz was found to be maximum for insonification perpendicular to the predominant myofiber orientation and minimum for parallel insonification, with values of (17 ± 14) and $(1.2\pm0.7)\times10^{-4}$ cm^{-1} sr^{-1} (mean±standard deviation), respectively. © *2007 Acoustical Society of America.* [DOI: 10.1121/1.2737747]

PACS number(s): 43.80.Cs, 43.80.Ev, 43.80.Jz [CCC] Pages: 581–586

I. INTRODUCTION

Ultrasonic imaging has been widely implemented in the diagnosis and assessment of cardiac function. Furthermore, ultrasonic tissue characterization offers the potential to provide methods for enhancing diagnoses and delineating underlying structural and mechanical properties associated with specific cardiac pathologies. Many studies (Baldwin *et al.*, 2005a, b; Hall *et al.*, 1997; Hoffmeister *et al.*, 1996a; b, 1995; Marutyan *et al.*, 2006; Mottley and Miller, 1990; Verdonk *et al.*, 1996, 1992; Yang *et al.*, 2006) have demonstrated that the ultrasonic backscatter, attenuation, and velocity properties of the heart exhibit an angular dependence relative to local myofiber orientation, presumably as a result of the approximately uniaxially arranged myofiber structure (Humphery, 2001; Streeter and Hanna, 1973). The effects of these anisotropic acoustic properties are readily observed and have been quantified in clinical echocardiographic studies (Aygen and Popp, 1987; Holland *et al.*, 1998; Recchia *et al.*, 1995).

Echocardiographic images obtained in the clinical setting represent images based on measurement of the "apparent backscatter" that is uncompensated for the effects of overlying tissue attenuation and diffraction of the transducer. The quantity reflecting the intrinsic scattering properties of myocardial tissue is the backscatter coefficient. There have been a number of studies carried out to measure the backscatter coefficient in myocardium (Hall *et al.*, 1997; O'Donnell *et al.*, 1981; Wear *et al.*, 1989). The study by O'Donnell *et al.* (1981) utilized a modified compensation method for backscatter coefficient estimates in myocardium and limited their measurements to one angle of insonification relative to the myofiber direction. Wear *et al.* (1989) implemented an enhanced formalism to study the backscatter properties of myocardium with assumed attenuation properties at a specific insonification angle. Hall *et al.* (1997) later showed the anisotropy of the frequency dependence of the apparent backscatter without compensation for tissue attenuation and transducer diffraction. In the current study, we measured the backscatter of formalin-fixed myocardium and reduced the data to the backscatter coefficient using measured values of the corresponding attenuation and velocity for proper compensation. We report the backscatter coefficient of formalin-fixed myocardium at 5 MHz as a function of insonification angle relative to the predominant myofiber orientation for all angles of insonification.

II. BACKGROUND

Proper compensations, including compensation for the attenuation in the tissue, effects of diffraction, and effects of signal gating, are required to reduce the experimentally measured backscattered data to the absolute backscatter coefficient, as indicated in the following:

$$\text{backscatter coefficient} = \left(\frac{\text{apparent backscattered power}}{\text{power from the reference}}\right) \times \left(\begin{array}{c}\text{reflection and transmission}\\ \text{loss compensation}\end{array}\right) \times \left(\begin{array}{c}\text{attenuation}\\ \text{compensation}\end{array}\right) \times \left(\begin{array}{c}\text{diffraction effects}\\ \text{compensation}\end{array}\right). \quad (1)$$

A number of approaches have been reported to reduce the experimental data to the backscatter coefficient (Chen *et al.*, 1997; Madsen *et al.*, 1984; O'Donnell and Miller, 1981; Wear *et al.*, 1989). These approaches differ primarily in the compensation for diffraction effects arising from the finite surface area of the transducer and the volume of interest occupied by the scatters. A comparison of these approaches was described in detail by Marsh *et al.* (1998).

For attenuation compensation, as presented originally by Sigelmann and Reid (1973), one needs to consider both the effects of the overlying tissue before the start of the region of

[a)]Electronic mail: james.g.miller@wustl.edu

interest and the attenuation incurred within the gated region. The general expression for attenuation compensation is

$$\text{attenuation compensation} = e^{4\alpha x_0}\frac{2\alpha\nu\tau e^{\alpha\nu\tau}}{e^{\alpha\nu\tau}-e^{-\alpha\nu\tau}}, \quad (2)$$

where α is the attenuation coefficient of the specimen, x_0 is the distance from the front wall of the specimen to the start of the time gate, ν is the speed of sound in the specimen, and τ is the duration of the time gate applied to the backscattered signals.

One useful approach for compensating for diffraction effects is described by Chen *et al.* (1997). With a perfectly reflecting reference plate positioned in the focal plane, the diffraction compensation can be simplified to

$$\begin{matrix}\text{diffraction effects}\\ \text{compensation}\end{matrix} = \frac{r_0^2}{(\nu\tau/2)\pi a^2 E_\infty}e^{-(2/\pi)(G_p/\pi)^{-1/2}}, \quad (3)$$

where G_p represents the pressure gain factor and is defined as $G_p=ka^2/2r_0$, r_0 is the geometric focal distance of the transducer (i.e., the radius of curvature of a spherically focused transducer), ν is the speed of sound in the media, τ is the time gate duration, a is the radius of the transducer, E_∞ is a dimensionless constant, and k is the wave number.

III. METHODS

A. Myocardial specimens

Lamb hearts were acquired from a commercial slaughterhouse within an hour of the death of the animal. The hearts were thoroughly rinsed and fixed using a standard protocol for tissue fixation, a neutral buffered 10% formalin solution, for approximately two weeks before individual cylindrical specimens were excised from the left ventricular wall. Specimens were cored transmurally from epicardium to endocardium such that the axis of symmetry of the specimens was oriented orthogonal to the epicardial surface of the heart, resulting in cylindrical specimens approximately 14 mm in diameter and 20 mm in height. The individual specimens were glued to a plastic (Delrin™) mount of equivalent diameter and maintained in a 10% formalin solution before measurements. A total of eight myocardial specimens were examined for this study.

B. Data acquisition

Due to the prohibitively long scan times involved, all measurements were performed on fixed tissue at room temperature rather than at 37 °C to reduce any potential time dependent changes in the tissue during the course of the study. Figure 1 shows the experimental setup for measurements of backscatter data. The plastic mount was friction fit into a cylindrical plastic coupling, which was then attached to the metal shaft of the rotation axis. A Unidex 511 motion controller (Aerotech, Pittsburgh, PA) allowed the assembly to have three-dimensional translational movement as well as rotational freedom. A piezoelectric transducer (Panametrics V309, 5-MHz, 2-in. (5.1 cm) focal length, 1/2-in. (1.3 cm) diameter; Panametrics, Waltham, WA) was used to acquire the data. The tissue plug was positioned at the geometric

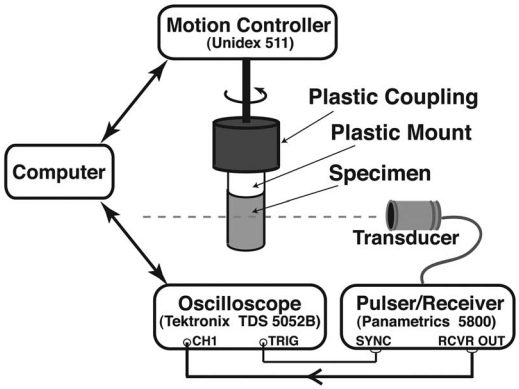


FIG. 1. Experimental setup for backscatter measurements.

focus of the transducer such that the focus lay roughly 6 mm into the tissue specimen, and the ultrasonic beam was centered on the midmyocardial region of each specimen. An ultrasonic pulse was generated with a Panametrics 5800 pulser/receiver operating in pitch/catch mode and fed to the transducer. The backscattered signal from the myocardial specimen was received by the same transducer, amplified by the Panametrics 5800, and subsequently digitized and stored with a Tektronix TDS 5052B digital oscilloscope (Tektronix, Beaverton, OR).

Backscatter signals were acquired every 6 degrees for a full rotation of the tissue specimen. At each angle of insonification, a total of 24 to 36 independent sites (depending upon the thickness of the individual tissue plug) were sampled in the midmyocardial region. This was accomplished by translating the plug in 1-mm steps and further averaging was performed by combining the backscatter measurements with data acquired at angles of ±2°. Each backscattered signal was digitized (8-bit) at 500 megasamples/s after 1024-trace averaging to improve the signal-to-noise ratio. A reference specular echo signal was acquired from a planar stainless-steel reflector positioned at the focus of the transducer. In addition, through-transmission data at each rotation angle were acquired to obtain velocity and attenuation information for proper compensation in the backscatter data reduction.

C. Data analysis

Typical backscatter signals from the myocardium at parallel and perpendicular insonification are shown in Fig. 2. As illustrated in Fig. 2, two regions of interest were isolated by applying time-domain gates to the backscattered signals at different depths. Separate gates were placed starting at 1.5 and 7 μs into the tissue, each with a duration of 4 μs. For a speed of sound of approximately 1.5 mm/μs, these starting times correspond to depths of 1.13 and 5.25 mm into the tissue, respectively, with a gate length of 3 mm (about 5.8% of the transducer's radius of curvature). A Hanning window was applied to the gated time-domain signals and the same window was applied to the reference reflection. A fast Fourier transformation (FFT) was applied to the windowed signals and the squared magnitude was calculated to produce the apparent backscattered power spectrum. A mean power

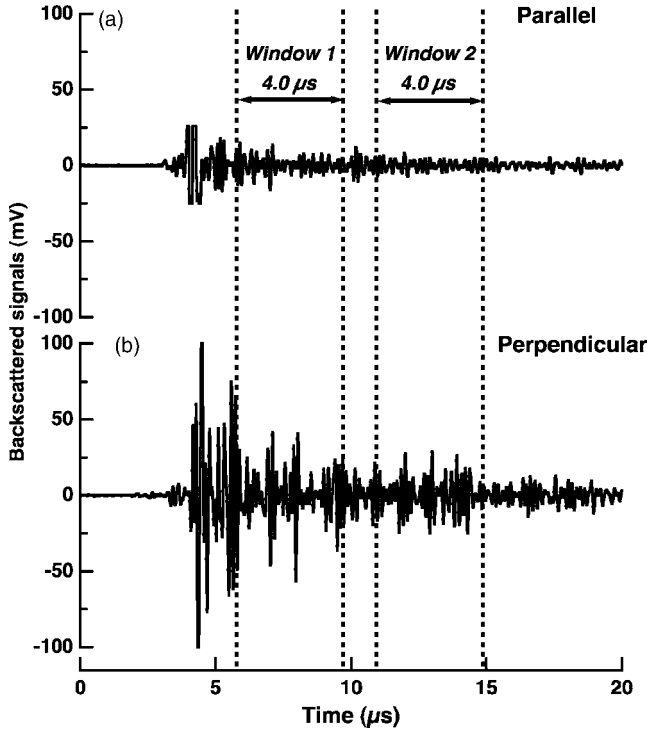


FIG. 2. Backscattered radio-frequency signals for (a) perpendicular insonification and (b) parallel insonification. The locations of both time-gate windows selected are shown.

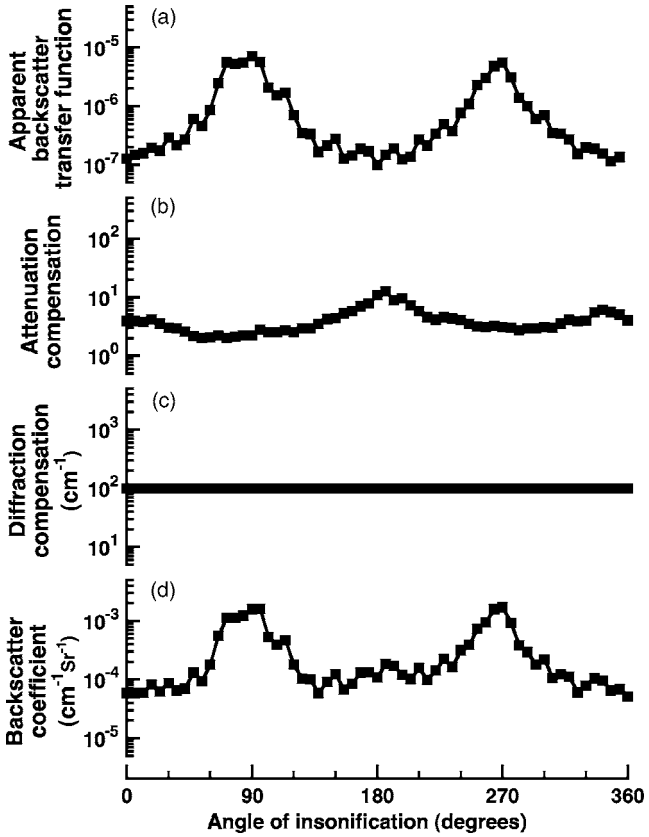


FIG. 3. The compensation components used to reduce the measured backscatter data to backscatter coefficient for one representative specimen. (a) Anisotropy of the apparent backscatter transfer function at 5 MHz. (b) Anisotropy of the attenuation compensation at 5 MHz. (c) Diffraction compensation at 5 MHz. (d) Anisotropy of the backscatter coefficient at 5 MHz.

spectrum representing each rotation angle was then obtained by averaging results from the spatially independent sampled sites.

To obtain measurements of the backscatter coefficient at 5 MHz at each angle of insonification, compensation for reflection and transmission loss, attenuation in the tissue, and the effects of diffraction were applied to the experimentally measured backscatter data using the methods described in Sec. II.

The measured mean apparent backscatter power spectrum for each angle of insonification was normalized by the reference power spectrum to remove the effects of the measurement system. This was accomplished by taking the magnitude-squared Fourier transform of a specular reflection from a stainless steel reference reflector positioned at the geometric focus of the transmitting transducer and windowed identically to our gated data, and dividing the measured power spectrum from the tissue by this reference power spectrum. This results in the apparent backscatter transfer function. The 5-MHz component of this normalized power spectral representation was used in subsequent data reduction. Because the speed of sound in tissue and speed of sound in water are quite similar (differing by less than 6%) and the density of myocardial tissue ($\sim$1.06 g/cm^3) is close to the density of water, resulting in a value of 99.6% for the intensity transmission coefficient, the transmission loss compensation factor was taken to be unity. The apparent backscatter transfer function at 5 MHz was compensated for the effects of attenuation from the overlying tissue as well as those within the gated region. This was accomplished by applying the attenuation compensation given by Eq. (2). The attenuation coefficient values at 5 MHz for this study ranged from 0.23 Np/cm (perpendicular) to 0.87 Np/cm (parallel) for all angles of insonification, which results in attenuation compensation factors varying from 1.28 to 2.49 for measurements at Window 1 (the shallower region of interest) and 1.91 to 11.28 for Window 2 (the deeper region of interest). In practice, because of the similarities of the speed of sound in tissue and in water, refraction caused by the sample's cylindrical shape was negligible; the diffraction pattern does not change significantly over the approximately 3-mm regions of interest. Estimates of diffraction effects from the focused transducer and the myocardial volume contributing to the observed backscatter were obtained by implementing the approach proposed by Chen *et al.* (1997) described by Eq. (3). For the transducer employed in this study with a focal length of 2 in. (5.1 cm) and a diameter of 1/2 in. (1.3 cm), and using the established value of E_∞=0.46 found in Chen *et al.*, the diffraction compensation factor at 5 MHz is 99.9 cm^{-1}.

Figure 3 illustrates the data reduction process for the backscatter coefficient obtained from a representative specimen at Window 2, the deeper region of interest. The relative contributions to apparent backscatter from the intrinsic attenuation and the scattering properties of myocardium are further discussed in the following.

IV. RESULTS

Measurements of the backscatter coefficient at 5 MHz were obtained for specific angles of insonification for each

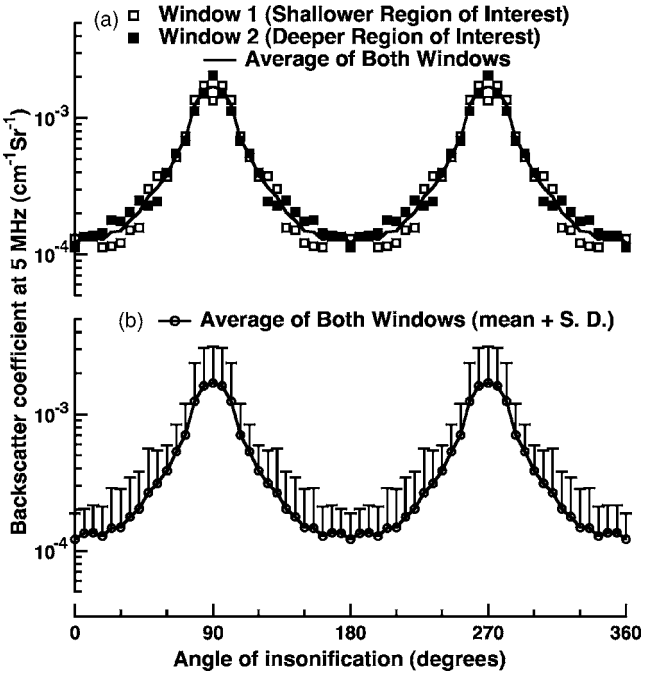


FIG. 4. (a) Anisotropy of the backscatter coefficient at 5 MHz for data gated by Window 1 (shallower region of interest; open square), Window 2 (deeper region of interest; closed square) and the average of the results from both windows. (b) Overall averaged anisotropy of the backscatter coefficient at 5 MHz (mean+standard deviation).

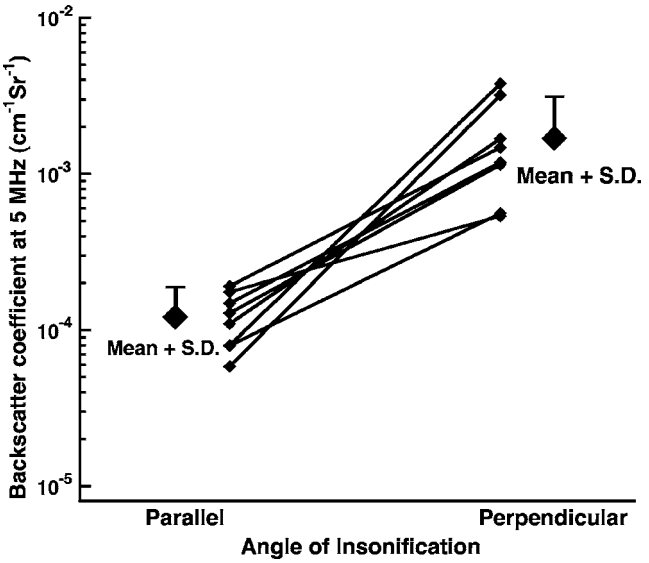


FIG. 5. Backscatter coefficient at 5 MHz from each individual specimen ($n=8$) for perpendicular and parallel insonification and the mean values (+ standard deviation).

specimen with the above-described data reduction method, then averaged across all specimens. To ensure that data from multiple samples were averaged at corresponding angles of insonification, the individual anisotropy curve from each specimen was aligned by cross correlation with a sinusoidal curve. A mean anisotropy curve was determined by averaging the aligned anisotropy curves from all eight specimens. Considering the approximately uniaxial symmetry, the angle of insonification relative to the predominant myofiber orientation ranges from 0 to 90° a total of four times over a complete rotation. Because the fiber orientation is equivalent for insonification angles at θ, $\pi+\theta$, $\pi-\theta$ and $2\pi-\theta$, we applied fourfold averaging to the measured data and displayed the results over 360° for clarity.

Previous studies have shown that a maximum in backscatter occurs for perpendicular insonification relative to the myofibers and a minimum level of backscatter occurs for parallel insonification (Hoffmeister *et al.*, 1995; Mottley and Miller, 1988). The measured backscatter coefficient at 5 MHz as a function of angle of insonification is shown in Fig. 4. The top panel illustrates the results from the two regions of interest at different depths. These data show good agreement with each other and hence suggest the appropriate compensation factors have been applied to the measured backscatter. The mean anisotropy curve from measurements of both regions of interest is indicated in the top panel and reproduced in the bottom panel with the corresponding uncertainties shown. As is evident in Fig. 4, the backscatter coefficient is minimum for parallel insonification relative to the myofibers and reaches a maximum for perpendicular insonification.

Figure 5 illustrates the measured backscatter coefficient from each individual specimen for perpendicular and parallel insonification along with corresponding mean values and standard deviations. In all cases, the backscatter coefficient is greater for perpendicular than for parallel insonification. The mean backscatter coefficient values were (1.2 ± 0.7) and $(17\pm14)\times10^{-4}$ cm^{-1} sr^{-1} (mean±standard deviation; $p=0.008$ for paired t-test, $p=0.001$ for unpaired t-test) for parallel and perpendicular insonification relative to the myofiber orientation, respectively. The uncertainties in the reported backscatter coefficient values represent measurement precision rather than accuracy, which is more challenging to determine (Madsen *et al.*, 1999; Wear *et al.*, 2005).

V. DISCUSSION

The backscatter coefficient, also known as the differential backscatter cross section per unit volume, describes the intrinsic backscatter property of the medium. In this study, we have quantified the anisotropy of backscatter coefficient at 5 MHz in formalin-fixed myocardium as a function of angle of insonification relative to the predominant myofiber orientation.

Figure 3 illustrates the data reduction process for a representative specimen analyzed for backscatter signal from Window 2 (deeper region of interest). Table I shows the measured backscatter and compensation values for perpendicular and parallel insonification and the anisotropy ratios for this specimen. These values illustrate the relative contributions of the anisotropy in intrinsic acoustic properties to the observed anisotropy in apparent backscatter (which is directly related to the apparent backscatter transfer function). As shown in Fig. 3 and Table I, the backscatter coefficient at perpendicular insonification is largest (resulting in the strongest backscattered signal) and it is smallest at parallel insonification (resulting in the weakest backscatter signal). The attenuation compensation is a minimum for perpendicular insonification and maximum for parallel insonification, which results in a

TABLE I. Contributions to the data reduction for the backscatter coefficient of a representative specimen for Window 2 (deeper region of interest). The comparison between perpendicular and parallel insonification is shown.

Angle of insonification relative to myofiber orientation	Apparent Backscatter transfer function at 5 MHz	Attenuation compensation at 5 MHz	Diffraction compensation as 5 MHz (1/cm)	Backscatter coefficient at 5 MHz (1/cm sr)
Perpendicular	6.4×10^{-6}	2.7	99.9	16×10^{-4}
Parallel	1.1×10^{-7}	7.4	99.9	0.8×10^{-4}
Ratio of perpendicular of parallel	58	0.36	1.0	20

smaller reduction in the apparent backscatter signal at perpendicular insonification than that at parallel insonification. The combination of these anisotropy effects in attenuation and intrinsic backscatter results in the observed magnitude of the anisotropy of apparent backscatter.

The anisotropy of longitudinal velocity and attenuation for formalin-fixed myocardium has been previously reported by our laboratory (Baldwin *et al.*, 2005a; b, 2005c). We have summarized those results together with our new measurements in Table II. It is interesting to note the degree of anisotropy present for each of these intrinsic acoustic parameters. As seen in Table II, velocity shows only modest anisotropy. The attenuation coefficient exhibits an almost twofold increase when the angle of insonification varies from perpendicular to the predominant myofiber orientation to parallel. On the other hand, the backscatter coefficient demonstrates a more than ten-to-one decrease when the angle of insonification changes from perpendicular to parallel.

Previous measurements of backscatter coefficient on fresh canine myocardium *in vitro* by O'Donnell *et al.* (1981) and *in vivo* by Wear *et al.* (1989) focused on the feasibility of separating normal and infarcted myocardium with the emphasis on the frequency dependence of the backscatter coefficient and apparent backscatter transfer function. Their studies revealed that the magnitude of backscatter coefficient was on the order of 1×10^{-3} cm^{-1} sr^{-1} for normal myocardium and was on the order of 1×10^{-2} cm^{-1} sr^{-1} for infarcted myocardium. One difference between the current study and previous studies is the implementation of a different diffraction correction as described earlier in Sec. II. Although angle dependence was not considered in the earlier studies, the insonification angle was approximately perpendicular based on the descriptions of the placement of the transducer on the heart wall. The backscatter coefficient for perpendicular insonification in the current study is of the same order of magnitude as the results for normal myocardium in those previous studies.

To the best of our knowledge, the results of this study represent the first reported measurements of the anisotropy of the backscatter coefficient in myocardium. These results may help in further delineating specific mechanisms associated with backscatter from myocardium.

TABLE II. Summary of acoustic properties of formalin-fixed myocardium (mean±standard deviation). The velocity (Baldwin *et al.*, 2005b) and attenuation (Baldwin *et al.*, 2005c) results were taken from previous measurements from our laboratory.

Angle of insonification relative to myofiber orientation	Velocity (mm/μs)	Attenuation coefficient at 5 MHz (Np/cm)	Backscatter coefficient at 5 MHz (1/cm sr)
Perpendicular	1.568±0.003	0.490±0.009	$(17\pm14)\times10^{-4}$
Parallel	1.581±0.004	1.01±0.02	$(1.2\pm0.7)\times10^{-4}$
Ratio of mean perpendicular to mean parallel	0.992	0.485	14

ACKNOWLEDGMENT

This work was supported, in part, by NIH R37 HL40302.

Aygen, M., and Popp, R. L. (**1987**). "Influence of the orientation of myocardial fibers on echocardiographic images," Am. J. Cardiol. **60**, 147–152.

Baldwin, S. L., Marutyan, K. R., Yang, M., Wallace, K. D., Holland, M. R., and Miller, J. G. (**2005a**). "Estimating myocardial attenuation properties from *m*-mode ultrasonic backscatter," Ultrasound Med. Biol. **31**, 477–484.

Baldwin, S. L., Yang, M., Marutyan, K. R., Wallace, K. D., Holland, M. R., and Miller, J. G. (**2005b**). "Measurement of the anisotropy of ultrasonic velocity in freshly excited and formalin-fixed myocardial tissue," J. Acoust. Soc. Am. **118**, 505–513.

Baldwin, S. L., Yang, M., Marutyan, K. R., Wallace, K. D., Holland, M. R., and Miller, J. G. (**2005c**). "Ultrasonic detection of the anisotropy of protein cross linking in myocardium," Proc.-IEEE Ultrason. Symp. **4**, 2263–2266.

Chen, X., Phillips, D., Schwarz, K. Q., Mottley, J. G., and Parker, K. J. (**1997**). "The measurement of backscatter coefficient from a broadband pulse-echo system: A new formulation," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **44**, 515–525.

Hall, C. S., Verdonk, E. D., Wickline, S. A., Perez, J. E., and Miller, J. G. (**1997**). "Anisotropy of the apparent frequency dependence of backscatter in formalin fixed human myocardium," J. Acoust. Soc. Am. **101**, 563–568.

Hoffmeister, B. K., Gehr, S. E., and Miller, J. G. (**1996a**). "Anisotropy of the transverse mode ultrasonic properties of fixed tendon and fixed myocardium," J. Acoust. Soc. Am. **99**, 3826–3836.

Hoffmeister, B. K., Handley, S. M., Wickline, S. A., and Miller, J. G. (**1996b**). "Ultrasonic determination of the anisotropy of young's modulus of fixed tendon and fixed myocardium," J. Acoust. Soc. Am. **100**, 3933–3940.

Hoffmeister, B. K., Wong, A. K., Verdonk, E. D., Wickline, S. A., and Miller, J. G. (**1995**). "Comparison of the anisotropy of apparent integrated ultrasonic backscatter from fixed human tendon and fixed human myocardium," J. Acoust. Soc. Am. **97**, 1307–1313.

Holland, M. R., Wilkenshoff, U. M., Finch-Johnston, A. E., Handley, S. M., Perez, J. E., and Miller, J. G. (**1998**). "Effects of myocardial fiber orientation in echocardiography: Quantitative measurements and computer simulation of the regional dependence of backscattered ultrasound in the parasternal short-axis view," J. Am. Soc. Echocardiogr **11**, 929–937.

Humphery, J. D. (**2001**). *Cardiovascular Solid Mechanics: Cells, Tissues, and Organs* (Springer, New York).

Madsen, E. L., Dong, F., Frank, G. R. *et al.* (**1999**). "Interlaboratory comparison of ultrasonic backscatter, attenuation, and speed measurements," J. Ultrasound Med. **18**, 615–631.

Madsen, E. L., Insana, M. F., and Zagzebski, J. A. (**1984**). "Method of data reduction for accurate determination of acoustic backscatter coefficients," J. Acoust. Soc. Am. **76**, 913–923.

Marsh, J. N., Hughes, M. S., Hall, C. S., Lewis, S. H., Trousil, R. L., Brandenburger, G. H., Levene, H., and Miller, J. G. (**1998**). "Frequency and concentration dependence of the backscatter coefficient of the ultrasound agent albunex®," J. Acoust. Soc. Am. **104**, 1654–1666.

Marutyan, K. R., Yang, M., Baldwin, S. L., Wallace, K. D., Holland, M. R., and Miller, J. G. (**2006**). "The frequency dependence of ultrasonic velocity and the anisotropy of dispersion in both freshly excised and formalin-fixed

myocardium," Ultrasound Med. Biol. **32**, 603–610.
Mottley, J. G., and Miller, J. G. (**1988**). "Anisotropy of the ultrasonic backscatter of myocardial tissue. I. Theory and measurements in vitro," J. Acoust. Soc. Am. **83**, 755–761.
Mottley, J. G., and Miller, J. G. (**1990**). "Anisotropy of the ultrasonic attenuation in soft tissues: Measurements in vitro," J. Acoust. Soc. Am. **88**, 1203–1210.
O'Donnell, M., and Miller, J. G. (**1981**). "Quantitative broadband ultrasonic backscatter: An approach to non-destructive evaluation in acoustically inhomogeneous materials," J. Appl. Phys. **52**, 1056–1065.
O'Donnell, M., Mimbs, J. W., and Miller, J. G. (**1981**). "Relationship between collagen and ultrasonic backscatter in myocardial tissue," J. Acoust. Soc. Am. **69**, 580–588.
Recchia, D., Hall, C., Shepard, R. K., Miller, J. G., and Wickline, S. A. (**1995**). "Mechanisms of the view-dependence of ultrasonic backscatter from normal myocardium," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **42**, 91–98.
Sigelmann, R. A., and Reid, J. M. (**1973**). "Analysis and measurement of ultrasound backscattering from an ensemble of scatterers excited by sine-wave bursts," J. Acoust. Soc. Am. **53**, 1351–1355.
Streeter, D. D., Jr., and Hanna, W. T. (**1973**). "Engineering mechanics for successive states in canine left ventricular myocardium. II. Fiber angle and sarcomere length," Circ. Res. **33**, 656–664.
Verdonk, E. D., Hoffmeister, B. K., Wickline, S. A., and Miller, J. G. (**1996**). "Anisotropy of the slope of ultrasonic attenuation in formalin fixed human myocardium," J. Acoust. Soc. Am. **99**, 3837–3843.
Verdonk, E. D., Wickline, S. A., and Miller, J. G. (**1992**). "Anisotropy of ultrasonic velocity and elastic properties in normal human myocardium," J. Acoust. Soc. Am. **92**, 3039–3050.
Wear, K. A., Milunski, M. R., Wickline, S. A., Perez, J. E., Sobel, B. E., and Miller, J. G. (**1989**). "Differentiation between acutely ischemic myocardium and zones of completed infraction in dogs on the basis of frequency-dependent backscatter," J. Acoust. Soc. Am. **85**, 2634–2641.
Wear, K. A., Stiles, T. A., Frank, G. R. *et al.* (**2005**). "Interlaboratory comparison of ultrasonic backscatter coefficient measurements from 2 to 9 mHz," J. Ultrasound Med. **24**, 1235–1250.
Yang, M., Baldwin, S. L., Marutyan, K. R., Wallace, K. D., Holland, M. R., and Miller, J. G. (**2006**). "Elastic stiffness coefficients (c11, c33, and c13) for freshly excised and formalin-fixed myocardium from ultrasonic velocity measurements," J. Acoust. Soc. Am. **119**, 1880–1887.

A tool for real-time acoustic species identification of delphinid whistles

Julie N. Oswald[a)]
Scripps Institution of Oceanography, University of California—San Diego, San Diego, California 92038

Shannon Rankin and Jay Barlow
NOAA-NMFS Southwest Fisheries Science Center, La Jolla Shores Drive, La Jolla, California 92037

Marc O. Lammers
Hawaii Institute of Marine Biology, Kailua, Hawaii 96834

(Received 10 October 2006; revised 1 May 2007; accepted 1 May 2007)

The ability to identify delphinid vocalizations to species in real-time would be an asset during shipboard surveys. An automated system, Real-time Odontocete Call Classification Algorithm (ROCCA), is being developed to allow real-time acoustic species identification in the field. This Matlab-based tool automatically extracts ten variables (beginning, end, minimum and maximum frequencies, duration, slope of the beginning and end sweep, number of inflection points, number of steps, and presence/absence of harmonics) from whistles selected from a real-time scrolling spectrograph (ISHMAEL). It uses classification and regression tree analysis (CART) and discriminant function analysis (DFA) to identify whistles to species. Schools are classified based on running tallies of individual whistle classifications. Overall, 46% of schools were correctly classified for seven species and one genus (*Tursiops truncatus, Stenella attenuata, S. longirostris, S. coeruleoalba, Steno bredanensis, Delphinus* species, *Pseudorca crassidens*, and *Globicephala macrorhynchus*), with correct classification as high as 80% for some species. If classification success can be increased, this tool will provide a method for identifying schools that are difficult to approach and observe, will allow species distribution data to be collected when visual efforts are compromised, and will reduce the time necessary for post-cruise data analysis. © *2007 Acoustical Society of America.* [DOI: 10.1121/1.2743157]

PACS number(s): 43.80.Ka [WWA] Pages: 587–595

I. INTRODUCTION

Acoustic techniques have been used to monitor a variety of species, ranging from birds (Mills, 1995; Chesmore, 2001), bats (Vaughan *et al.*, 1997; Parsons and Jones, 2000), and fallow deer (Reby *et al.*, 1997) to insects such as crickets and grasshoppers (Chesmore, 2001). Increasingly, acoustic techniques are being used to monitor marine mammal populations (Leaper *et al.*, 1992; Clark *et al.*, 1996; Stafford *et al.*, 2001; Gordon *et al.*, 2000; van Parijs *et al.*, 2002; Wang *et al.*, 2005). Many marine mammals produce distinctive sounds and therefore acoustic techniques can be used to detect not only their presence, but also species identity. Several species of large whales produce calls that are easily recognized (Thompson *et al.*, 1992; Goold and Jones, 1995; Thompson *et al.*, 1996), but the calls produced by many delphinid species are highly variable and overlap in frequency characteristics, making them more challenging to identify to species.

The calls produced by delphinids have generally been placed into three distinct categories: echolocation clicks, burst pulse sounds, and whistles. Echolocation clicks are short, broadband pulses with peak frequencies that vary from tens of kilohertz to well over 100 kHz (Norris and Evans, 1966; Au, 1980). Echolocation clicks generally occur in trains containing few to hundreds of clicks and are used for navigation and object detection and discrimination (Au, 1993). Burst pulse signals are broadband click trains with very short interclick intervals. These sounds take on a tonal quality to human ears because the clicks are repeated at such high rates that the rate itself, rather than the individual clicks, is audible (Watkins, 1967; Herzing, 2000). It is thought that these signals play a role in social interactions, although they may also function in echolocation tasks. Whistles are continuous, narrow band, frequency modulated signals that often have harmonic components. They range in duration from several tenths of a second to several seconds (Tyack and Clark, 2000). The fundamental frequency of most whistles ranges from 2 to 30 kHz (Lammers *et al.*, 2003; Oswald *et al.*, 2004). Whistles are believed to function as social signals (Janik and Slater, 1998; Herzing, 2000; Lammers *et al.*, 2003).

Delphinid species identification studies have commonly focused on whistle characteristics to develop classification algorithms (Steiner, 1981; Wang *et al.*, 1995; Matthews *et al.*, 1999; Rendell *et al.*, 1999; Oswald *et al.*, 2003). Correct classification scores obtained in these studies are generally significantly greater than expected by chance alone, suggesting that whistles contain information that could be used to identify delphinid species. Roch *et al.*, (2007) included both whistles and clicks in their species identification algorithms.

[a)]Electronic mail: oswald.jn@gmail.com

Their correct classification scores of 67% to 75% for four species suggest that clicks may also be useful for species identification, however, their method does not distinguish the relative contribution of whistles and clicks to the classifier. In most areas, such as the eastern tropical Pacific Ocean (ETP), whistle sounds propagate much farther than echolocation clicks, so whistles are likely to be more useful for long-distance species identification. For example, whistles were the only sounds detected from 60% of 1867 schools detected acoustically during five visual and acoustic surveys in the ETP (unpublished data). Therefore, species identification algorithms based on whistles are likely to be more useful than those that depend on clicks.

Species identification algorithms are developed by postprocessing of field recordings. While postprocessing provides valuable information, the ability to identify vocalizations to species in real time would be a great asset during shipboard marine mammal abundance surveys. Traditionally during these surveys, a team of visual observers searches for marine mammals and then directs the ship towards them for school size estimation and species identification (Wade and Gerrodette, 1993; Barlow, 1995). Recently, methods have been developed to tow a hydrophone array behind research vessels during standard visual cetacean surveys (Fristrup and Clark, 1997; Barlow and Taylor, 2005). While the addition of acoustic techniques has been shown to increase rates and distances of detection (Leaper *et al.*, 1992, Clark and Fristrup, 1997; Gordon *et al.*, 2000; Barlow and Taylor, 2005), real-time acoustic species identification would provide several further advantages. This capability would allow the acoustic team to aid visual observers with the identification of groups that are difficult to approach and observe due to factors such as animal behavior, inclement weather, and reduced visibility. In addition, because marine mammals spend much of their lives under water, schools are often detected acoustically but not visually (Barlow and Taylor, 2005). The ability to identify these detections would allow the use of acoustic methods when conditions do not allow for visual observations. Finally, real-time acoustic species identification can reduce the bottleneck of post-cruise data analysis.

Until now, no methods have been available for real time acoustic identification of delphinid whistles. In this paper we present a new software tool that is being developed for this task: Real-time Odontocete Call Classification Algorithm (ROCCA). ROCCA is a Matlab-based tool that extracts, measures, and classifies whistles to species in real-time.

II. METHODOLOGY

A. Data collection

Acoustic recordings were made during six shipboard marine mammal abundance surveys conducted by the Southwest Fisheries Science Center (NOAA, NMFS). Each 4-month survey occurred between the months of July and December. Four of the surveys took place in the ETP: *Stenella* Population and Abundance Monitoring (SPAM) 1998, and *Stenella* Abundance Research (STAR) 1999, 2000, and 2003. This study area extended from the United States/Mexico border southward to the territorial waters of Peru, and from the continental shores of the Americas west to the longitude of Hawaii (Fig. 1). The Hawaiian Islands Cetacean and Ecosystem Assessment Survey (HICEAS 2002) study area covered the waters within the 200 nmi Exclusive Economic Zone (EEZ) around the Hawaiian Island chain from the island of Hawaii in the southeast to Kure Atoll in the northwest (Fig. 1). The Pacific Islands Cetacean Ecosystem Assessment Survey (PICEAS 2005) took place in the US EEZ waters of Palmyra and Johnston Atoll and adjacent waters south of Hawaii (Fig. 1).

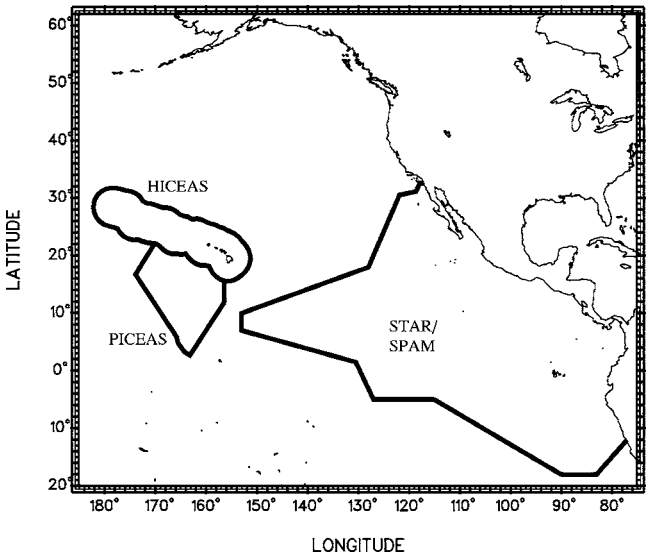


FIG. 1. Pacific Ocean study area boundaries for the HICEAS 2002, PICEAS 2005, SPAM 1998, and STAR 1999, 2000, and 2003 Southwest Fisheries Science Center (NOAA, NMFS) marine mammal abundance surveys.

During all surveys, a team of three experienced visual observers actively searched for marine mammals using two 25×150 binoculars and by naked eye (Kinzey *et al.*, 2001). When cetaceans were sighted, they were approached for species identification and school size estimation. Cetacean vocalizations were monitored and recorded using a towed hydrophone array and Type SSQ-57 sonobuoys. The array was towed 200 m behind the research vessel during daylight hours at a depth of approximately 4–6 m. Table I gives the frequency response characteristics of the arrays used during the surveys. During the SPAM 1998 survey, signals from the array were recorded onto digital audio tape (DAT) using Sony TCD-D7 and TCD-D8 DAT recorders (48 kHz sampling rate). During the STAR 2000 survey and all subsequent surveys, signals from the array were sent through a Mackie CR1604-VLZ mixer for equalization and were recorded directly to computer hard drive via an analog-to-digital conversion card (National Instruments BNC-2110 and DAQCard-6062E) using a 200-kHz sampling rate.

An acoustic technician monitored signals from two hydrophones in the array using a stereo headset and real-time scrolling spectrographic software (ISHMAEL, Mellinger, 2001). Whaltrak, a mapping program with a GPS interface, automatically logged time and position every 5 min while the array was being monitored. During the STAR 2000 survey and all subsequent surveys, acoustic detections were localized using ISHMAEL and Whaltrak. Bearing angles were

TABLE I. Frequency response and gain characteristics of hydrophone arrays used during Southwest Fisheries Science Center (NOAA, NMFS) marine mammal abundance surveys. Two arrays were used during the SPAM 1998 and STAR 2000 surveys. The array used during the PICEAS 2005 survey had four elements, three relatively narrow band and one relatively broadband.

Survey	Array frequency response	No. of hydrophone elements
SPAM 1998	500 Hz to 150 kHz±3 dB at −163 dB *re* 1 V/mPa	3
	32 Hz to 25 kHz±3 dB at −173 dB *re* 1 V/mPa	5
STAR 2000	2–45 kHz±4 dB at −132 dB *re* 1 V/mPa	5
	2–120 kHz±3 dB at −164 dB *re* 1 V/mPa	3
HICEAS 2002	500 Hz to 30 kHz±5 dB at −155 dB *re* 1 V/mPa	2
STAR 2003	500 Hz to 30 kHz±5 dB at −155 dB *re* 1 V/mPa	3
PICEAS 2005	1–40 kHz±5 dB at −150 dB *re* 1 V/mPa	3
	2–150 kHz±2 dB at −166 dB *re* 1 V/mPa	1

determined using phone-pair cross-correlation algorithms in ISHMAEL, and distance was determined by examining the convergence of bearing angles plotted on Whaltrak. Comparisons of the angle and distance to the acoustic detection with the location of the sighting allowed confirmation that the vocalizations detected were produced by the sighted dolphins.

A hydrophone array was not towed during the STAR 1999 survey. Instead, U.S. Navy sonobuoys (type SSQ-57) were deployed in close proximity to dolphin sightings. These sonobuoys had a flat frequency response from approximately 2 to 20 kHz and were deployed at a hydrophone depth setting of either 18 or 27 m. Sonobuoy signals were transmitted to a multichannel receiver aboard the research vessel and were recorded onto DAT using Sony TCD-D7 DAT recorders.

Single-species acoustic recordings were obtained from nine delphinid species during the six surveys: bottlenose dolphins (*Tursiops truncatus*), spotted dolphins (*Stenella attenuata*), spinner dolphins (*S. longirostris*), striped dolphins (*S. coeruleoalba*) rough-toothed dolphins (*Steno bredanensis*), short-beaked common (*Delphinus delphis*), long-beaked common dolphins (*D. capensis*), false killer whales (*Pseudorca crassidens*), and short-finned pilot whales (*Globicephala macrorhynchus*). A total of 2606 whistles from 176 schools were included in the analysis. Table II lists the number of whistles analyzed for each species.

B. Spectrographic analysis

Recordings of schools that had been visually identified to species and confirmed to contain only one species were chosen for analysis. Recordings were included only if the school was sighted at least 3 nmi from any other school in the area. This helped to ensure that the whistles analyzed were produced by the school being observed and not another nearby school. This was especially important during the SPAM 1998 and STAR 1999 surveys, when acoustic localization techniques were not available.

We randomly selected 50% of loud and clear whistles

TABLE II. Total number of whistles analyzed for each species, number of schools that those whistles were recorded from, and means and standard deviations (in parentheses) for continuous variables measured from the whistles. Frequency variables are given in kHz, duration is given in seconds, and number of inflection points (IP) and number of steps are count data.

Species	No. of whistles	No. of schools	Beginning frequency	Ending frequency	Minimum frequency	Maximum frequency	Duration	No. of IP	No. of steps
Bottlenose dolphin	306	15	11.61	10.24	7.92	17.07	1.11	2.85	2.17
			(5.11)	(4.78)	(2.49)	(4.55)	(0.69)	(2.67)	(3.61)
Spotted dolphin	399	26	9.92	14.92	8.41	17.99	0.75	1.29	3.06
			(3.94)	(5.66)	(2.39)	(4.69)	(0.38)	(1.45)	(3.84)
Striped dolphin	401	38	10.80	12.01	8.48	14.98	0.69	1.84	2.36
			(3.96)	(3.40)	(2.21)	(3.61)	(0.35)	(1.82)	(3.19)
Spinner dolphin	259	19	11.85	12.94	9.99	15.09	0.61	1.89	0.83
			(4.42)	(4.33)	(3.18)	(4.57)	(0.42)	(3.53)	(1.64)
Rough-toothed dolphin	192	14	7.41	8.33	6.46	9.53	0.64	2.56	1.51
			(3.15)	(2.95)	(2.33)	(2.97)	(0.36)	(3.00)	(1.84)
Short-beaked common dolphin	314	25	11.63	12.18	8.30	15.04	0.70	1.64	1.76
			(4.84)	(4.38)	(2.69)	(4.39)	(0.39)	(1.87)	(2.31)
Long-beaked common dolphin	174	11	10.87	14.46	8.48	16.21	0.62	1.59	1.74
			(4.89)	(5.12)	(2.70)	(4.94)	(0.34)	(3.29)	(2.19)
False killer whale	340	10	5.77	6.27	5.28	6.95	0.44	0.85	0.03
			(1.62)	(1.52)	(1.23)	(1.83)	(0.22)	(0.90)	(0.18)
Short-finned pilot whale	221	18	4.40	5.59	3.73	6.39	0.48	0.86	0.21
			(2.72)	(3.60)	(2.04)	(3.89)	(0.35)	(1.58)	(0.81)

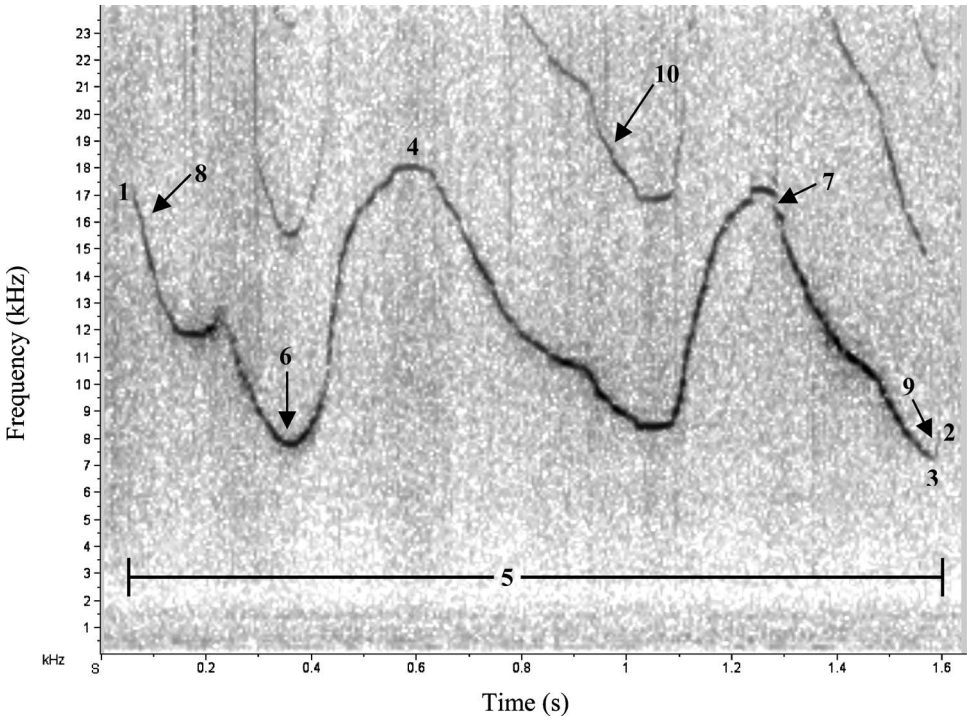


FIG. 2. Spectrogram of a bottlenose dolphin whistle (512-point FFT, Hanning window), showing the ten variables that were measured from each whistle. (1) Beginning frequency (kHz), (2) end frequency (kHz), (3) minimum frequency (kHz), (4) maximum frequency (kHz), (5) duration (s), (6) number of inflection points (defined as a change from positive to negative or negative to positive slope), (7) number of steps (defined as a 10% or greater increase or decrease in frequency over two contour points), (8) slope of the beginning sweep (positive, negative, or zero), (9) slope of the end sweep (positive, negative, or zero), and (10) presence/absence of harmonics (a binary variable).

from each recording session, up to a total of 35 whistles per recording session. This level of subsampling was chosen in order to obtain a sufficient sample size while avoiding oversampling of groups or individuals (which can lead to nonindependence of data). Overlapping whistles were included only if each individual whistle contour could be discerned without question. Whistles were considered to be "loud and clear" if they were at least 9 dB above ambient noise.

Ten variables were measured from the fundamental contour of each whistle: (1) beginning frequency (kHz), (2) end frequency (kHz), (3) minimum frequency (kHz), (4) maximum frequency (kHz), (5) duration (s), (6) number of inflection points (defined as a change from positive to negative or negative to positive slope), (7) number of steps (defined as a 10% or greater increase or decrease in frequency over two contour points), (8) slope of the beginning sweep (positive, negative, or zero), (9) slope of the end sweep (positive, negative, or zero), and (10) presence/absence of harmonics (a binary variable). These variables are shown in Fig. 2. Some whistles from the SPAM 1998 and STAR 1999 surveys were missing measurements for one or more variables because a portion of the whistle extended beyond the upper bandwidth limit of the recording equipment. These whistles were excluded from the analysis. Otherwise, all whistles from all years were combined by species for analysis. Table II gives descriptive statistics for the seven continuous whistle variables.

C. Classification algorithms

Following Oswald *et al.* (2003), classification algorithms were created using multivariate discriminant function analysis (DFA) and classification and regression tree analysis (CART). Discriminant function analysis classifies whistles to prespecified groups based on orthogonal linear functions derived from the ten variables listed above. Mahalanobis distances were calculated for each whistle being classified. The Mahalanobis distance is a measure of the distance in multivariate space of the whistle in question to the group centroid of each species in the analysis. A whistle was classified as the species that it was closest to in multivariate space. Prior to running DFA, continuous variables (frequency variables, duration, and number of steps and inflection points) were tested for normality and log or square-root transformed as necessary. Classification and regression tree analysis creates decision trees by separating data into groups known as nodes through a series of binary splits. Each split is based on the value of a single variable. Final classification is reached at terminal nodes. Terminal node probabilities reflect the certainty of the classification based on the purity of the node. Because CART is nonparametric, it was not necessary to transform variables for normality.

Two different methods of classification using DFA and CART were evaluated. In the first method, whistles were classified directly to species level. This will be referred to as the "direct" method for the remainder of this paper. The second method was hierarchical. Whistles were first classified to the broad categories of "blackfish" or "delphinid." The blackfish category consisted of two species: false killer whales and short-finned pilot whales. The delphinid category consisted of five species and one genus: bottlenose, spotted, spinner, striped, rough-toothed, and common dolphins. Common dolphin species (*Delphinus delphis* and *D. capensis*) were pooled in this analysis (see Sec. III). Once classified to category using the hierarchical approach, whistles within each category were then classified to species.

A jackknife method was used to calculate correct classification scores for DFA and CART. Six versions of the DFA and CART classification algorithms were created, each omitting all of the whistles from one cruise (one year of sampling). Whistles were classified using the algorithms that did

not include them. In this way, classification algorithms were created from data that were independent of the whistles being classified. This helped ensure that whistles were classified based on species-specific characteristics rather than group- or individual-specific characteristics. Fisher's exact test was used to test whether correct classification scores were significantly greater than expected by chance alone. Statistical significance was evaluated at $\alpha=0.05$ without corrections for multiple testing.

D. ROCCA

ROCCA was created using Matlab and interfaces with real-time scrolling spectrograph software, ISHMAEL (Mellinger, 2001). ISHMAEL is used to monitor signals detected by the hydrophone array. When a whistle of interest is detected, the user stops the scrolling spectrograph and selects the whistle. A Matlab routine called through ISHMAEL opens Matlab and saves the selection as a wav file. ROCCA is then run through Matlab. First, ROCCA automatically extracts the whistle contour from the wav file by stepping through the file one FFT window at a time. The fundamental frequency of the whistle contour is selected based on the peak frequency in each window. A routine within ROCCA ensures that random transient peaks in the spectrum are not mistaken for the fundamental peak frequency. For this study, a Hanning window was used. Window size was set at 1024 points and window overlap was set at 0.25.

When the whistle contour has been extracted, ROCCA automatically measures the ten variables described previously from the fundamental frequency contour of the whistle. The ten variables are then processed using DFA and CART classification algorithms and ROCCA outputs two predicted species, one based on each analysis. It takes approximately 20–45 s for ROCCA to extract a whistle contour, measure whistle variables, and provide an estimate of species identification. As multiple whistles from a single school of dolphins are processed, ROCCA keeps a running tally of species predictions. When all of the whistles from a school have been analyzed, ROCCA classifies the school as the species that the majority of whistles were predicted to be. When DFA and CART results do not agree, the algorithm that resulted in the greatest number of whistles classified as one species is chosen. For example, if DFA classifies 65% of the whistles in a school as bottlenose dolphins and CART classifies 58% of the same whistles as spotted dolphins, the school is classified as bottlenose dolphins. When the same numbers of whistles are classified as two or more species within DFA or CART, the species with the smallest mean Mahalanobis distance is chosen for DFA and the species with the largest mean terminal node probability is chosen for CART. In order to ensure that all whistles analyzed are from one discrete school, whistles are localized using ISHMAEL and Whaltrak before they are analyzed using ROCCA.

As whistles are analyzed, ROCCA creates three text files for each school. One contains the extracted whistle contours (time, frequency, and intensity of the peak frequency in each window). The second contains the ten whistle variables measured from each whistle in the school. The third contains DFA and CART predicted species, as well as Mahalanobis distances (DFA) and terminal node probabilities (CART) for each whistle in the school.

III. RESULTS

When ROCCA was run on all nine species, only 17.8% of short-beaked common dolphin whistles were correctly classified by DFA and 5.7% by CART. Similarly, 6.1% of long-beaked common dolphin whistles were correctly classified by DFA and 2.9% by CART. To explore the possibility that this result was caused by an inability to differentiate between the two common dolphin species, a version of ROCCA was created that included only short-beaked and long-beaked common dolphins. Overall correct classification scores in this analysis were not significantly greater than the 50% expected by chance (DFA: 49.7%, $p=1$, CART: 46.8%, $p=0.45$). Because the two common dolphin species could not be distinguished reliably from one another, they were pooled in subsequent analyses.

When the direct version of ROCCA was run on seven species and the pooled common dolphin species, DFA correctly classified 33.5% of whistles and CART correctly classified 33.6% of whistles. These correct classification scores are significantly greater than the 12.5% expected by chance ($p<0.0001$ for both DFA and CART). For individual species, DFA correct classification scores ranged from 14.7% (striped dolphins) to 63.8% (short-finned pilot whales). Correct classification scores were significantly greater than expected by chance for every species except striped dolphins ($p=0.41$). Correct classification scores for CART ranged from 18.5% (spinner dolphins) to 57.1% (false killer whales). All correct classification scores were significantly greater than chance with the exception of spinner dolphins ($p=0.07$). Based on the pooled tallies of individual whistle classifications, 43.8% of schools were correctly classified by DFA and CART combined. Correct classification scores for schools ranged from 31.6% (spinner dolphins) to 73.3% (bottlenose dolphins). Half were significantly greater than chance, with the exceptions being common dolphins ($p=0.24$), spinner dolphins ($p=0.23$), spotted dolphins ($p=0.10$), and striped dolphins ($p=0.06$).

The hierarchical version of ROCCA resulted in no significant difference in the overall correct classification of either individual whistles or schools compared to the direct version of ROCCA (individual whistles: DFA $p=0.23$; CART $p=0.31$; schools $p=0.75$). However, several significant differences were found for individual species when the hierarchical version of ROCCA was run. The correct classification of individual whistles increased significantly from 22.6% to 35.9% for false killer whales (DFA, $p<0.001$) and from 33.8% to 41.4% for spotted dolphins (CART, $p=0.03$) and decreased significantly from 26.2% to 14.7% for striped dolphins (CART, $p<0.001$). Correct classification scores for individual whistles were significantly greater than chance for every species with the exception of striped dolphins (CART, $p=0.41$).

The hierarchical version of ROCCA resulted in a significant difference in the correct classification of schools for

TABLE III. Classification results for schools classified based on multiple whistles using the hierarchical DFA and CART method. Percent of schools correctly classified for each species are in bold. Correct classification scores that are significantly greater (Fisher's exact test, $\alpha=0.05$) than the 12.5% expected by chance are marked by an asterisk, and p values are given in the last column. The number of schools included in the analysis for each species (n) is given in the second to last column. Overall, 46.0% of schools were classified to the correct species. This is significantly greater ($p<0.0001$) than expected by chance.

	% Classified as									
Actual species	Bottlenose dolphin	Spotted dolphin	Striped dolphin	Spinner dolphin	Rough-toothed dolphin	Common dolphin	False killer whale	Short-finned pilot whale	n	p
Bottlenose dolphin	**80.0***	0.0	0.0	6.7	6.7	6.7	0.0	0.0	15	<0.001
Spotted dolphin	23.1	**50.0***	7.7	7.7	0.0	7.7	3.8	0.0	26	0.006
Striped dolphin	26.3	23.7	**15.8**	7.9	0.0	26.3	0.0	0.0	38	0.76
Spinner dolphin	21.1	5.3	21.1	**26.3**	10.5	15.8	0.0	0.0	19	0.96
Rough-toothed dolphin	14.3	0.0	7.1	0.0	**64.3***	7.1	0.0	7.1	14	0.02
Common dolphin	22.2	13.9	5.6	0.0	2.8	**55.6***	0.0	0.0	36	<0.001
False killer whale	10.0	0.0	0.0	10.0	0.0	0.0	**80.0***	0.0	10	0.005
Short-finned pilot whale	5.6	0.0	0.0	5.6	0.0	5.6	38.9	**44.4**	18	0.06

only one species. Correct classification of common dolphin schools increased significantly from 27.8% to 55.6% ($p=0.03$). Schools of all species were correctly classified significantly more often than expected by chance, with the exception of short-finned pilot whales ($p=0.06$), striped dolphins ($p=0.76$), and spinner dolphins ($p=0.96$). The confusion matrix for this analysis is given in Table III as an example of classification errors. Overall, 46.0% of whistles were classified to the correct species. With the exception of false killer whales and short-finned pilot whales, whistles from most species were commonly misclassified as bottlenose dolphins. Short-finned pilot whales were most commonly misclassified as false killer whales. Striped dolphins were misclassified as bottlenose dolphins, common dolphins, and spotted dolphins more often than they were correctly classified.

To evaluate the effect of combining DFA and CART predictions and basing classification decisions on all whistles analyzed during an encounter rather than on individual whistles, correct classification scores were compared for three approaches: (1) classifying one whistle at a time, (2) classifying schools based on tallies of species predictions for DFA and CART individually, and (3) classifying schools based on a combination of DFA and CART predictions. These comparisons were made for both the direct version of ROCCA and the hierarchical version. Correct classification scores and p values for these comparisons are given in Tables IV and V. Basing classification decisions on schools rather than individual whistles for DFA and CART individually resulted in no significant differences in correct classification scores in the direct version of ROCCA. Correct classification of common dolphins increased significantly for both DFA and CART ($p=0.004$ and $p=0.006$, respectively) in the hierarchical version of ROCCA. When classification decisions were based on a combination of DFA and CART predictions, rather than on individual whistles, overall cor-

TABLE IV. Correct classification scores for individual whistles classified to species by DFA and CART individually and for schools classified based on multiple whistles using DFA and CART individually and DFA and CART combined. p values are given for comparisons of correct classification scores for individual whistles versus schools (DFA and CART individually) and for individual whistles versus schools (DFA and CART combined). Significant differences are marked by asterisks.

	Individual whistles		Schools						
Species	DFA % correct	CART % correct	DFA % correct	p Whistles versus schools (DFA)	CART % correct	p Whistles versus schools (CART)	DFA and CART % correct	p Whistles (DFA) versus schools (DFA and CART)	p Whistles (CART) versus schools (DFA and CART)
Bottlenose dolphin	60.8	35.9	86.7	0.06	40.0	0.8	73.3	0.4	0.005*
Spotted dolphin	29.6	33.8	38.5	0.4	34.6	1.0	34.6	0.7	1.0
Striped dolphin	14.7	26.2	13.2	1.0	34.2	0.3	34.2	0.005*	0.3
Spinner dolphin	24.7	18.5	26.3	1.0	21.1	0.8	31.6	0.6	0.2
Rough-toothed dolphin	46.9	34.4	64.3	0.3	57.1	0.2	71.4	0.1	0.008*
Common dolphin	28.3	21.7	41.7	0.1	22.2	1.0	27.8	1.0	0.4
False killer whale	22.6	57.1	0.0	0.1	80.0	0.2	70.0	0.002*	0.5
Short-finned pilot whale	63.8	50.7	72.2	0.6	44.4	0.6	61.1	0.8	0.5
Overall	33.5	33.6	39.8	0.1	36.4	0.5	43.8	0.007*	0.007*

TABLE V. Correct classification scores for individual whistles classified to species by hierarchical DFA and CART individually and for schools classified based on multiple whistles using hierarchical DFA and CART individually and DFA and CART combined. p values are given for comparisons of correct classification scores for individual whistles versus schools (DFA and CART individually) and for individual whistles versus schools (DFA and CART combined). Significant differences are marked by asterisks.

	Whistles		Schools						
Species	DFA % correct	CART % correct	DFA % correct	p Whistles versus schools (DFA)	CART % correct	p Whistles versus schools (CART)	DFA and CART % correct	p Whistles (DFA) versus schools (DFA and CART)	p Whistles (CART) versus schools (DFA and CART)
Bottlenose dolphin	60.8	37.9	86.7	0.06	53.3	0.3	80.0	0.2	0.002*
Spotted dolphin	29.6	25.4	34.6	0.7	50.0	0.4	50.0	0.04*	0.4
Striped dolphin	20.2	14.7	10.5	0.2	15.8	0.8	15.8	0.7	0.8
Spinner dolphin	24.3	19.7	15.8	0.6	21.1	0.8	26.3	0.8	0.6
Rough-toothed dolphin	42.7	41.7	57.1	0.4	57.1	0.3	64.3	0.2	0.2
Common dolphin	28.5	25.4	52.8	0.004*	47.2	0.006*	55.6	0.001*	<0.001*
False killer whale	35.9	60.6	40.0	0.8	90.0	0.1	80.0	0.007*	0.3
Short-finned pilot whale	56.1	50.2	55.6	1.0	44.4	0.8	44.4	0.5	0.8
Overall	35.1	35.0	39.8	0.2	40.8	0.09	46.0	0.004*	0.004*

rect classification increased significantly in both versions of ROCCA. Individual species correct classification scores increased significantly for bottlenose dolphins ($p=0.005$ when individual whistles classified by CART), rough-toothed dolphins ($p=0.008$ when individual whistles classified by CART), striped dolphins ($p=0.005$ when individual whistles classified by DFA), and false killer whales ($p=0.002$ when individual whistles classified by DFA) in the direct version of ROCCA, and for bottlenose dolphins ($p=0.002$ when individual whistles classified by CART), spotted dolphins ($p=0.04$ when individual whistles classified by DFA), false killer whales ($p=0.007$ when individual whistles classified by DFA), and common dolphins ($p=0.001$ when individual whistles classified by DFA, $p<0.001$ when individual whistles classified by CART) in the hierarchical version of ROCCA.

IV. DISCUSSION

Traditional visual monitoring techniques during shipboard marine mammal surveys are limited by animal behavior, environmental conditions, and logistical constraints. The addition of a passive acoustic component to these surveys provides a method for overcoming some of these limitations. Real-time acoustic species identification offers an additional tool for identifying schools that are difficult to approach and observe and allows species distribution data to be collected even when visual effort is compromised by factors such as poor visibility, inclement weather, and high sea states.

Real-time acoustic species identification is especially valuable during surveys dedicated to specific species. For example, the focus of the PICEAS 2005 survey was to determine the population status of false killer whales in an area of high fishery bycatch in the central tropical Pacific Ocean. Visual detection of these animals was extremely difficult due to animal behavior and high sea states. During the first month of this survey, five schools of false killer whales were encountered. Three of the five schools were detected and located by the acoustic team and identified as false killer whales using ROCCA. The acoustic identifications were confirmed visually. Time constraints demanded that the ship deviate from the survey trackline for acoustic detections of this focal species only, and therefore real-time acoustic identification was crucial. The combination of high correct classification scores for the species identity of false killer whale whistles, combined with high vocal rates and poor visual detection of this species, created a situation in which passive acoustics played an indispensable role. Without the capability for real-time species identification provided by ROCCA, the acoustic detections would not have been investigated and valuable data would have been lost. ROCCA was also used to estimate the fraction of schools of false killer whales missed by visual methods within 4.5 km of the transect line during the PICEAS 2005 survey (Barlow and Rankin, 2007). This provided a means of ground-truthing estimated line-transect parameters and showed that the fraction of detections missed visually (0.56) was consistent with the expected fraction missed (0.58).

In addition to providing assistance to the visual observers, ROCCA has the advantage of reducing the bottleneck of post-cruise analysis. ROCCA's automated whistle extraction, measurement, and data storage features reduce the time necessary for post-cruise analysis and make ROCCA valuable for other applications such as processing the voluminous amounts of data collected using seafloor mounted acoustic recorders.

While correct classification scores obtained using ROCCA are not at the level of near-certainty that would be optimal for shipboard surveys, results are promising as correct classification scores for the individual whistles of most species were significantly greater than expected by chance. Scores did not reach near certainty due to high within-species variability in whistle variables and a large degree of

overlap in the time and frequency variables of many species (Table II). Bottlenose dolphins and false killer whales had the highest correct classification scores, with 80% of schools of both species being correctly identified using the hierarchical version of ROCCA (Table III). Examination of descriptive statistics (Table II) shows that some whistle variables for these two species are distinctive. Bottlenose dolphin whistles have a longer mean duration and false killer whale whistles have lower mean frequencies compared to most other species in the analysis. However, while few whistles of species other than short-finned pilot whales and rough-toothed dolphins were misclassified as false killer whales, whistles from most species were commonly misclassified as bottlenose dolphins. This may be due to frequency and duration variables. Bottlenose dolphins had one of the lowest values for mean minimum frequency, one of the highest values for mean maximum frequency, and the longest mean duration. All of these values had high standard deviations, suggesting that, based on the variables measured, many different whistle types would fall into the bottlenose dolphin category. This implies that the variables measured were not sufficient for separating species in this analysis.

Correct classification scores were low for spinner and striped dolphins in all analyses. Striped dolphin classification errors were relatively evenly spread across all species except false killer whales, short-finned pilot whales, and rough-toothed dolphins. Similarly, spinner dolphin classification errors were generally evenly spread across all species other than short-finned pilot whales. These patterns are also likely due to frequency variables. The whistles of the small delphinid species (spotted, striped, spinner, short-beaked common and long-beaked common dolphins) had very similar frequency characteristics for the variables measured (Table II).

Because of the high degree of overlap in frequency characteristics among species, a method for increasing classification success may lie in the exploration of additional whistle variables such as the rate of change in frequency (slope), the locations of steps and inflection points within whistles, and relative intensities of different frequencies. Also, compound variables such as a combined value for slope and frequency range may be more effective for separating species. In addition, alternate classification algorithms such as artificial neural networks and hidden Markov models may be better suited to the task of identifying dolphin whistles. Work is currently under way to explore the effect of alternate whistle variables and classification algorithms on correct classification scores.

The species included in ROCCA are often found in single species schools, but have also been observed mixed with other species. ROCCA was created using recordings of single species schools and therefore does not currently have the capability to identify mixed species schools as such. Future plans for ROCCA include the analysis of recordings of mixed species schools in order to develop decision criteria for identifying schools as mixed versus single species.

ROCCA performed best when classification decisions were based on multiple whistles classified using the hierarchical method and when decisions were based on a combination of DFA and CART results. This approach resulted in an overall correct classification score of 46.0% and some very high correct classification scores of up to 80% for species such as bottlenose dolphins and false killer whales (Table III). Making classification decisions based on multiple whistles rather than individual whistles resulted in slight increases in correct classification scores, but these were not significant for DFA or CART (Tables IV and V). However, when classification decisions for schools were made based on a combination of DFA and CART results, correct classification increased significantly both overall and for several individual species. While the hierarchical version of ROCCA did not result in a significant increase in overall correct classification, it did increase slightly, and correct classification of common dolphin schools in particular increased significantly. Additionally, correct classification scores were significantly greater than chance for five of the eight species in the hierarchical version, compared to four of the eight species in the direct version of ROCCA. This approach shows some promise and may produce more significant results with different species categories or a greater number of levels within the hierarchy.

The results of this study point not only to the benefit of making classification decisions based on multiple whistles, but also to the benefit of using more than one classification algorithm. Different classification algorithms are sensitive to different characteristics of the data set, and the ability to combine the strengths of more than one algorithm can result in higher classification success. When the optimal set of whistle variables and classification algorithms is assembled, ROCCA will be valuable not only for real-time species identification during shipboard surveys, but also for analysis of vocalizations recorded using seafloor-mounted hydrophones. While ROCCA has been created for use in the eastern tropical and temperate Pacific Ocean, it can be modified for use in other regions.

ACKNOWLEDGMENTS

We would like to gratefully acknowledge the tireless efforts of the skilled acousticians, visual observers, and cruise leaders who dedicated months to the surveys. Without them this project could not have been completed. We thank the officers and crew of the NOAA ships *McArthur*, *McArthur II*, and *David Starr Jordan* and the NSF/UNOLS R/V *Endeavor* for their patience and cooperation. Tim Gerrodette and Lisa Balance (and J.B.) were the Chief Scientists for these surveys. Special thanks to Stephanie Grassia and Theresa Weber for their assistance with data analysis. We also thank Michael Oswald and two anonymous reviewers for providing insightful comments on drafts of this manuscript.

Au, W. W. L. (**1993**). *The Sonar of Dolphins* (Springer-Verlag, New York).
Au, W. W. L. (**1980**). "Echolocation signals of the Atlantic bottlenose dolphin (*Tursiops truncatus*) in open waters," in *Animal Sonar Systems*, edited by R. G. Busnel and J. F. Fish (Plenum, New York), pp. 251–282.
Barlow, J. (**1995**). "The abundance of cetaceans in California waters. Part I: Ship surveys in summer and fall of 1991," Fish. Bull. **93**, 1–14.
Barlow, J., and Rankin, S. (**2007**). "False killer whale abundance and density: preliminary estimates for the PICEAS study area south of Hawaii and

new estimates for the US EEZ around Hawaii," NOAA Administrative Report LJ-07-02.
Barlow, J., and Taylor, B. L. (**2005**). "Estimates of sperm whale abundance in the northeastern temperate Pacific from a combined acoustic and visual survey," Marine Mammal Sci. **21**, 429–445.
Chesmore, E. D. (**2001**). "Application of time domain signal coding and artificial neural networks to passive acoustical identification of animals," Appl. Acoust. **62**, 1359–1374.
Clark, C. W., Charif, R., Mitchell, S., and Colby, J. (**1996**). "Distribution and behavior of the bowhead whale, *Balaena mysticetus*, based on analysis of acoustic data collected during the 1993 spring migration off Point Barrow, Alaska," Rep. Int. Whal. Comm. **46**, 541–552.
Clark, C. W., and Fristrup, K. M. (**1997**). "Whales '95: A combined visual and acoustic survey of blue and fin whales off Southern California," Rep. Int. Whal. Comm. **47**, 583–600.
Fristrup, K. M., and Clark, C. W. (**1997**). "Combining visual and acoustic survey data to enhance density information," Rep. Int. Whal. Comm. **47**, 933–936.
Goold, J. C., and Jones, S. E. (**1995**). "Time and frequency domain characteristics of sperm whale clicks," J. Acoust. Soc. Am. **98**, 1279–1291.
Gordon, J. C. D., Matthews, J. N., Panigada, S., Gannier, A., Borsani, J. F., and Notarbartolo di Sciara, G. (**2000**). "Distribution and relative abundance of striped dolphins in the Ligurian Sea Cetacean Sanctuary: Results from an acoustic collaboration," J. Cetacean Res. Manage. **2**, 27–36.
Herzing, D. L. (**2000**). "Acoustics and social behavior of wild dolphins: implications for a sound society," in *Hearing by Whales and Dolphins*, edited by W. W. L. Au, A. N. Popper, and R. R. Fay (Springer-Verlag, New York), pp. 225–272.
Janik, V. M., and Slater, P. J. B. (**1998**). "Context-specific use suggests that bottlenose dolphin signature whistles are cohesion calls," Anim. Behav. **56**, 829–838.
Kinzey, D., Gerrodette, T., Perryman, W., Olson, P., and Rankin, S. (**2001**). "Marine mammal data collected during a survey in the eastern tropical Pacific ocean aboard the NOAA ships McArthur and David Starr Jordan, July 28–December 9, 2000," NOAA Technical Memorandum NOAA-TM-NMFS-SWFSC-303.
Lammers, M. O., Au, W. W. L., and Herzing, D. L. (**2003**). "The broadband social acoustic signaling behavior of spinner and spotted dolphins," J. Acoust. Soc. Am. **114**, 1629–1639.
Leaper, R., Chappell, O., and Gordon, J. (**1992**). "The development of practical techniques for surveying sperm whale populations acoustically," Rep. Int. Whal. Comm. **42**, 549–560.
Matthews, J. N., Rendell, L. E., Gordon, J. C. D., and MacDonald, D. W. (**1999**). "A review of frequency and time parameters of cetacean tonal calls," Bioacoustics **10**, 47–71.
Mellinger, D. K. (**2001**). "ISHMAEL 1.0 User's Guide," NOAA Technical Memorandum OAR PMEL-120. Available from NOAA/PMEL, 7600 Sand Point Way, NE, Seattle, WA 98115-6349.
Mills, H. (**1995**). "Automatic detection and classification of nocturnal migrant bird Calls," J. Acoust. Soc. Am. **97**, 3370.
Norris, K. S., and Evans, W. E. (**1966**). "Directionality of echolocation clicks in the rough-toothed porpoise, *Steno bredanensis* (Lesson)," in *Marine Bio-acoustics*, edited by W. N. Tavolga (Pergamon, New York), pp. 305–324.
Oswald, J. N., Barlow, J., and Norris, T. F. (**2003**). "Acoustic identification of nine delphinid 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 delphinid species," J. Acoust. Soc. Am. **116**, 3178–3185.
Parsons, S., and Jones, G. (**2000**). "Acoustic identification of twelve species of echolocating bat by discriminant function analysis and artificial neural networks," J. Exp. Biol. **203**, 2641–2656.
Reby, D., Lek, S., Dimopoulos, I., Joachim, J., Lauga, J., and Aulagnier, S. (**1997**). "Artificial neural networks as a classification method in the behavioral sciences," Behav. Processes **40**, 35–43.
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.
Roch, M. A., Soldevilla, M. S., Burtenshaw, J. C., Henderson, E. E., and Hildebrand, J. A. (**2007**). "Gaussian mixture model classification of odontocetes in the Southern California Bight and the Gulf of California," J. Acoust. Soc. Am. **121**, 1737–1748.
Stafford, K. M., Nieukirk, S. L., and Fox, G. C. (**2001**). "Geographic and seasonal variation of blue whale calls in the North Pacific," J. Cetacean Res. Manage. **3**, 65–76.
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.
Thompson, P., Findley, L. T., and Vidal, O. (**1992**). "20 Hz pulses and other vocalizations of fin whales, *Balaenoptera physalus*, in the Gulf of California, Mexico," J. Acoust. Soc. Am. **92**, 3051–3057.
Thompson, P., Findley, L. T., Vidal, O., and Cummings, W. (**1996**). "Underwater sounds of blue whales, *Balaenoptera musculus*, in the Gulf of California, Mexico," Marine Mammal Sci. **12**, 288–292.
Tyack, P. L., and Clark, C. W. (**2000**). "Communication and acoustic behavior of dolphins and whales," in *Hearing by Whales and Dolphins*, edited by W. W. L. Au, A. N. Popper, and R. R. Fay (Springer-Verlag, New York), pp. 156–224.
van Parijs, S. M., Smith, J., and Corkeron, P. J. (**2002**). "Using calls to estimate the abundance of inshore dolphins: A case study with Pacific humpback dolphins *Sousa chinensis*," J. Appl. Ecol. **39**, 853–864.
Vaughan, N., Jones, G., and Harris, S. (**1997**). "Habitat use by bats (*Chiroptera*) assessed by means of a broad-band acoustic method," J. Appl. Ecol. **34**, 716–730.
Wade, P. R., and Gerrodette, T. (**1993**). "Estimates of cetacean abundance and distribution in the eastern tropical Pacific," Rep. Int. Whal. Comm. **43**, 477–493.
Wang, D., Wursig, B., and Evans, W. (**1995**). "Comparisons of whistles among seven odontocete species," in *Sensory Systems of Aquatic Mammals*, edited by R. A. Kastelein, J. A. Thomas, and P. E. Nachtigall (De Spil, Woerden), pp. 299–323.
Wang, K., Wang, D., Akamatsu, T., Songhai, L., and Xiao, J. (**2005**). "A passive acoustic monitoring method applied to observation and group size estimation of finless porpoise," J. Acoust. Soc. Am. **118**, 1180–1185.
Watkins, W. A. (**1967**). "The harmonic interval: fact or artifact in spectral analysis of pulse trains," in *Marine Bio-acoustics*, edited by W. N. Tavolga (Pergamon, New York), pp. 15–42.

Age-related differences in the acoustic characteristics of male leopard seals, *Hydrurga leptonyx*

Tracey L. Rogers[a)]
Australian Marine Mammal Research Centre, Zoological Parks Board of NSW and Veterinary Faculty, University of Sydney, Australia, P.O. Box 20, Mosman, New South Wales, 2088, Australia

(Received 3 August 2006; revised 3 April 2007; accepted 13 April 2007)

During the breeding season, the underwater vocalizations and calling rates of adult male leopard seals are highly stereotyped. In contrast, sub-adult males have more variable acoustic behavior. Although adult males produce only five stereotyped broadcast calls as part of their long-range underwater breeding displays the sub-adults have a greater repertoire including the adult-like broadcast calls, as well as variants of these. Whether this extended repertoire has a social function is unknown due to the paucity of behavioral data for this species. The broadcast calls of the sub-adults are less stereotyped in their acoustic characteristics and they have a more variable calling rate. These age-related differences have major implications for geographic variation studies, where the acoustic behavior of different populations are compared, as well as for acoustic surveying studies, where numbers of calls are used to indicate numbers of individuals present. Sampling regimes which unknowingly include recordings from sub-adult animals will artificially exaggerate differences between populations and numbers of calling animals. The acoustic behavior of sub-adult and adult male leopard seals were significantly different and although this study does not show evidence that these differences reflect vocal learning in the male leopard seal it does suggest that contextual learning may be present.
[DOI: 10.1121/1.2736976]

PACS number(s): 43.80.Ka [WWA] Pages: 596–605

I. INTRODUCTION

Vocal learning, or production learning, refers to instances where vocalizations themselves are modified in form as a result of experience with those of other individuals, whereas learning that affects usage and comprehension of a sound is described as contextual learning and is not considered as a type of vocal learning (Janik and Slater, 1997, 2000).

Contextual learning, in relation to vocal communication, is relatively common among mammals and has been reviewed in Janik and Slater (1997). It occurs when a preexisting signal comes to be associated with a new context as a result of experience with the signals of other individuals (Janik and Slater, 1997, 2000). Janik and Slater (2000) describe two types of contextual learning: usage learning, where an existing signal is produced in a new context, or comprehension learning, where a receiver comes to extract a novel meaning from a signal.

By comparison, as reviewed by Janik and Slater (1997), production learning has been convincingly shown more rarely in mammals: in the greater horseshoe bat (Jones and Ransome, 1993); bottlenose dolphin (Caldwell and Caldwell, 1972; Reiss and McCowan, 1993); killer whale (Ford, 1991); humpback (Payne and Payne, 1985; Payne *et al.*, 1983), and bowhead (Würsig and Clark, 1993) whales. The phocid seals are one of the few mammals to have shown vocal (production) learning where seals learned new sounds by imitating human speech. This was exhibited by two male harbor seals that were able to mimic speech sounds (Ralls *et al.*, 1985). Strong dialects between adjacent colonies of Weddell seals (Morrice *et al.*, 1994) suggest that vocal learning may be a likely explanation for geographic variation in this species (Janik and Slater, 1997), so southern phocid seals may be a valuable mammalian group to examine for the existence of vocal development.

Leopard seal vocal behavior. As the leopard seal is a highly vociferous phocid we investigate here whether vocal learning plays a part in the development of the adult vocal repertoire. Leopard seals have been shown to have geographic variation in their vocal behavior (Thomas and Golladay, 1995) and their acoustic behavior is believed to be linked with their breeding behavior as it coincides with elevated reproductive hormones (Rogers *et al.*, 1996). Females independently give birth to their pups and wean them on the ice floes of the Antarctic pack ice. Males do not remain with the females, only mother-pup groups are observed on ice floes (Siniff and Stone, 1985). Females advertise their sexual receptivity by vocalizing underwater from the beginning of their oestrus until mating (Rogers *et al.*, 1996). Mating occurs from December to early January, after the pups have weaned (Siniff and Stone, 1985).

Male leopard seals become highly vocal prior to and during the breeding season, from early November through January (Stirling and Siniff, 1979). At this time the leopard seals are solitary and widely dispersed (Rogers and Bryden, 1997) and so must communicate over great distances. Lone male leopard seals vocalize for many hours each day. These

[a)]Electronic mail: trogers@zoo.nsw.gov.au

calls are believed to function for mate attraction and/or territorial signaling and are referred to as broadcast calls (Rogers *et al.*, 1996).

Sub-adult male leopard seals, from 1 to 4 years of age, produce broadcast calls underwater for long periods each day during the breeding season. The function of these vocalizations is unknown as the sub-adult animals will not be part of the breeding population until they reach sexual maturity at age $4\frac{1}{2}$, so their vocal displays are not likely to serve for mate attraction. Similarly, the sub-adult vocalizations are not likely to serve a territorial function as these animals are found in close proximity to one another, and become less spatially tolerant as they mature (Rogers and Bryden, 1997). Perhaps the sub-adult male seals are vocalizing to practice their broadcast calls so that they become more adult-like in form or are learning the appropriate context in which to produce these sounds.

This project aimed to identify whether the acoustic behavior of sub-adult and adult male leopard seals differ significantly. Three acoustic features were examined: (I) vocal repertoire, (II) acoustic characteristics of broadcast calls, and (III) calling patterns of vocalizations.

II. METHODS

A. Sound recordings

Between November and early January in 1992/93 and 1993/94 underwater sound recordings were made along the edge of the fast ice between 68° 25′ S, 77° 10′ E and 68° 35′ S, 77° 50′ E, Prydz Bay, Eastern Antarctica. The leopard seals were confined to ice floes at the outer edge of the fast ice. Recordings were made between 1600 and 0300 h to coincide with the period when leopard seals are most vocal underwater (Thomas and DeMaster, 1982). All recordings were made using a Sony WMD6C audio-cassette recorder, a Brüel and Kjær 8103 hydrophone, and a custom preamplifier. This system had a frequency response from 40 to 15 000 Hz±3 dB. The hydrophone was placed between 5.5 and 6 m below the water surface and 4 m below the undersurface of the fast ice.

B. Identification of vocalizing seals

To individually identify seals, both when they were hauled out on the ice and when they were in the water, their fur was marked with a nontoxic oil-based paint in individually distinctive patterns. Paint was applied by approaching the seal to within 5 m and throwing a zip lock bag filled with paint. A typical response by the seals to being painted was to rear their head up and move toward the researcher before resettling after a couple of minutes. There were no long-term adverse effects to the seals from this procedure. Animal Ethics Committee and other relevant permits were conducted under Antarctic Science Advisory Committee (ASAC) Program No. 1144. Seals were identified as being males when lying on their side by the presence of a penis slit. The age of individuals in the study was not known as animals were not caught and so could not be measured. However the sub-adult seals were likely to have been between 1 and 3 years as none of the animals used was small enough to be a newly weaned juvenile. Thus, the minimum age of a sub-adult was determined to be a seal weaned from the previous year. To maximize the likelihood of making recordings from known individuals the research team traveled to the edge of the fast ice in the early afternoon when seals were likely to be hauled out on the ice (Thomas and DeMaster, 1982; Rogers and Bryden, 1997). This allowed underwater recordings to be made prior to and after the seal's entry into the water. A vocalizing seal could be identified if: (1) no other leopard seals were vocalizing nearby, (2) vocalizing commenced within 5 min of the marked seal entering the water, and (3) if vocalizing was heard when the marked seal submerged and stopped when it reappeared at the surface. It was difficult to record underwater sounds from a large number of marked individuals as most seals moved out to sea, away from the recording site, before vocalizing commenced or did not vocalize at all. Some seals, however, did remain in the vicinity while vocalizing and recordings were made from five different known adult males and five sub-adult males. Ninety minutes of sounds were recorded from the sub-adult seals: SA1, SA2, SA3, SA4, and SA5 and adult seals: A1, A2, A3, A4, and 45 min from A5.

C. Calls in the vocal repertoire

The leopard seal broadcast calls in this study include those described by Stirling and Siniff (1979) and Rogers *et al.* (1995); namely, the high double trill (H), low double trill (L), medium single trill (M), low descending trill (D), and hoot with a single trill (O). While adult leopard seal calls are highly stereotyped (Rogers and Cato, 2002) the sub-adult seals produced some calls that resembled adult broadcast calls and other calls that are similar but not the same as the adult calls. This second group of calls show a variable degree of similarity to adult calls from extremely different to only slightly different. The key (Table I) outlines the criteria which were used to distinguish whether a sub-adult seal's call was a broadcast call or a variant type call. When a call had features that varied from any one of the criteria it was classified as a variant. Variant calls were grouped into four categories according to their frequency range: high frequency variants (HV) lie in the frequency range 2800–3200 Hz, medium frequency variants (MV) range 1500–2000 Hz, low descending variants (DV) range 400–1000 Hz, and low frequency variants (LV) range 100–400 Hz. The variants of the low double trill and the hoot with a single trill are indistinguishable from each other, as they are in a similar frequency band and have similar components; therefore they were grouped into the low frequency variant category.

To identify whether subadult seals produced more variant calls than adult seals, the number of each type of broadcast call and each variant call type were counted from a random sample of 900 calls produced by all five adult (450) and five sub-adult (450) seals (Fig. 1). Log-linear analysis was used to identify differences in the production rate of broadcast and variant calls, to identify differences in call production among individual males and to identify whether there is a difference in the production of variants of any

TABLE I. Criteria used to distinguish whether a subadult's call was a broadcast call or a variant type call.

Call types	Criteria
High double trill	The call is composed entirely of short-duration pulses produced in series; there should be no narrow-band component
	The series is broken into two parts
	Each part should be of similar length, although the second part may be shorter in duration but no shorter than 50% of the length of the first part
	The maximum frequency is found at the beginning of the call
	The frequency initially drops at the beginning of the call but then remains constant through the call
	There is no change in the repetition rate throughout the call
Medium single trill	The call is composed entirely of short-duration pulses produced in series; there should be no narrow-band component
	There is only one part to the call
	Carrier frequency remains relatively constant throughout the call
	There is an increase in the repetition rate within the first third of the call but then remains constant through the call
Low descending trill	The call is composed entirely of short-duration pulses produced in series; there should be no narrow-band component
	There is only one part to the call
	The frequency gradually decreases in the fundamental frequency through the call
	There is an increase in the repetition rate within the first third of the call but then remains constant through the call
	The vocalization has a gradual onset, but rapid decay
Low double trill	The call is composed of both a narrow-band component and short-duration pulses produced in series
	The call is broken into two parts
	Each part should be of similar length, although the second part may be shorter in duration but no shorter than 50% of the length of the first part
	The first part of the call has an initial narrow-band component which then becomes pulsed but there is no gap between the narrow-band component and the series of pulses
	The second part is composed of short duration pulses produced in series
	The frequency remains constant through the call
	There is no change in the repetition rate throughout the call
Hoot with low single trill	The call is composed of both a narrow-band component and short-duration pulses produced in series
	The call is broken into two parts
	The first part of the call is a narrow-band component
	The second part is composed of short-duration pulses produced in series
	The frequency remains constant through the call
	There is no change in the repetition rate throughout the second part of the call

particular type of broadcast call. To identify whether variants were produced of particular broadcast calls, the proportion of variants of any broadcast call type (e.g., HV/H+HV) were calculated for each sub-adult seal.

D. Acoustic characteristics of the broadcast calls

To compare the acoustic characteristics between adult and sub-adult male leopard seals 250 high double trill and

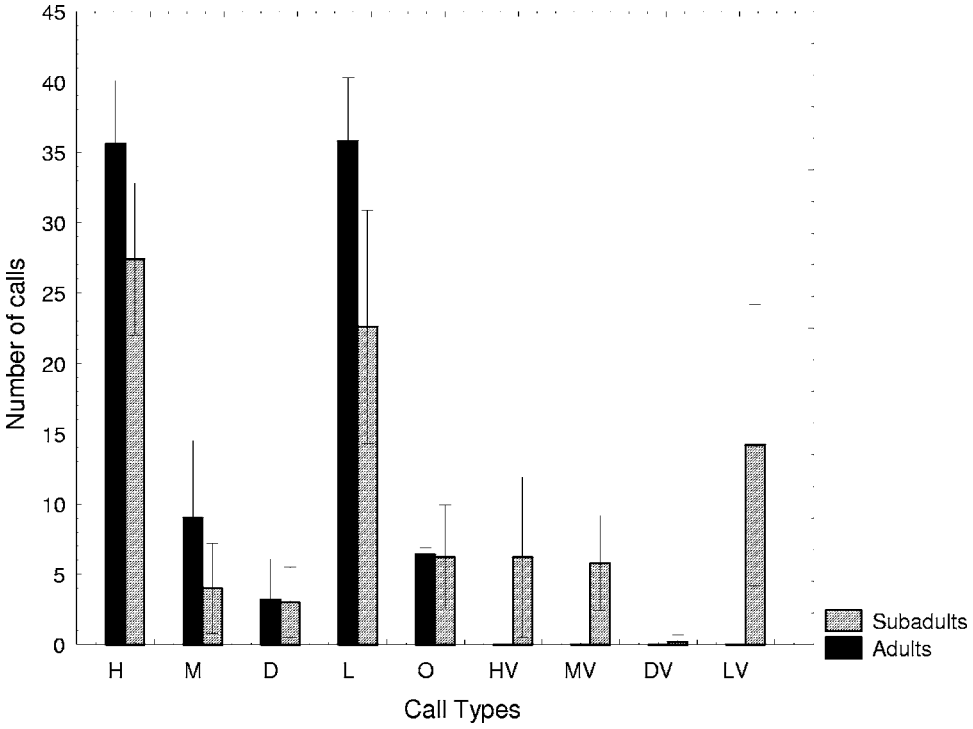


FIG. 1. Comparison of frequency of occurrence (mean±s.d.) of broadcast calls (H, M, D, L, O) and variant calls (HV, MV, DV, LV) occurring in randomly chosen leopard seal vocalizations; 450 from adult seals and 450 from subadult seals.

250 low double trill vocalizations were chosen randomly from recordings from each of the five adult and five subadult seals (Table II). As there were fewer representatives of the remaining broadcast calls, data sets for these calls are smaller. One hundred and thirty eight representatives of the medium single trill were compared, 56 representatives of the low descending trill, and 78 representatives of the hoot with a single trill were compared (Table II).

Vocalizations were analyzed using a spectrum analyser (Kay Elemetrics 5500 Sona-Graph). Frequency characteristics were measured from sonagrams with 29 Hz frequency resolution (filter bandwidth) and only the values of the fundamental were used. The maximum and minimum frequencies were defined as the upper and lower values of the dominant band on the sonagrams. The duration and pulse repetition rates were measured from sonagrams with 117 Hz frequency resolution, giving a time resolution of 8.9 ms.

Different acoustic characteristics were measured for different call types. The medium and low descending trills are composed of a single series of short-duration pulses whereas the high and low double trills and the hoot with a single trill are broken into two parts (Figs. 2–4). The low double trill and the hoot with a single trill both have a narrow-band component. The acoustic characteristics measured from sonagrams for each call type are summarized in Table III. Multivariate analysis of variance (MANOVA) was used to compare the acoustic characteristics of adult and sub-adult male leopard seal vocalizations using the following characteristics

TABLE II. Number of broadcast calls analyzed per seal: high double trills (H), medium single trills (M), low descending trills (D), low double trills (L), and hoots with a single trill (O); and the number of vocalizing bouts and periods analysed per seal.

		Call type				
	Seal	H	M	D	L	O
Adults	A1	25	0	0	25	9
	A2	25	25	5	25	10
	A3	25	13	6	25	6
	A4	25	12	8	25	7
	A5	25	19	9	25	7
Subtotal		125	69	28	125	39
Sub-adults	SA1	25	31	8	25	5
	SA2	25	0	3	25	9
	SA3	25	13	5	25	11
	SA4	25	17	6	25	9
	SA5	25	8	6	25	5
Subtotal		125	69	28	125	39
Totals		**250**	**138**	**56**	**250**	**78**

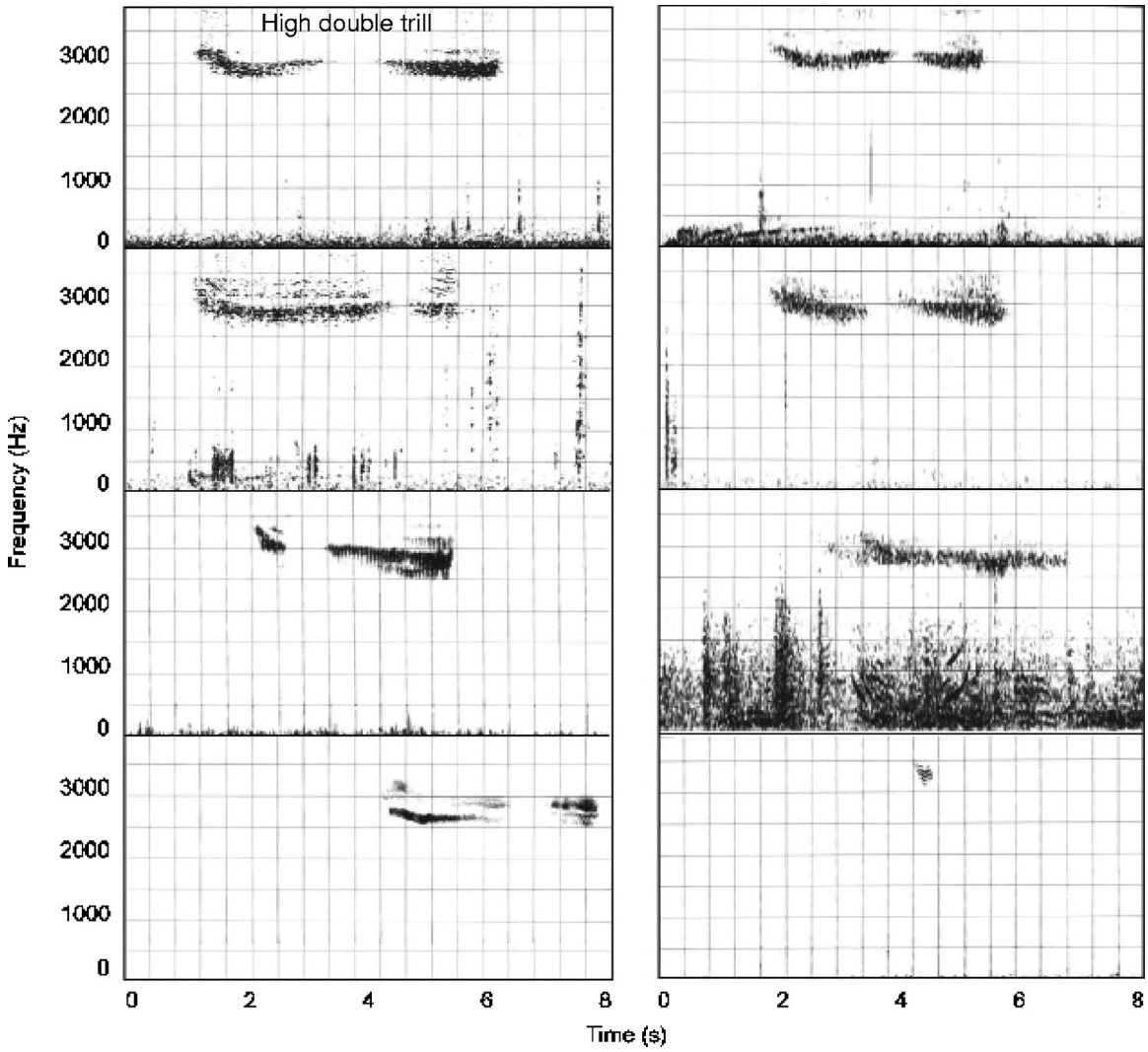


FIG. 2. Sonagrams of a leopard seal high double trill (H) and examples of the seven types of variants of this call (HV).

for each call type: Max, BW, Dur2, and Dur for the high double trill; FNB, Frq1, Frq2, Dur1, and Dur2 for the low double trill; Min, Max, and Dur for the medium single trill; Min, Max, and Dur for the low descending trill; FNB, Min, Max, Dur1, and Dur2 for the hoot with a single trill. ANOVA was used to give comparisons for each of the acoustic characteristics between adult and sub-adults for every call type and it was used to compare the difference in frequencies of part 1 (Frq1) and part 2 (Frq2) of the sub-adult and then adult seals calls.

E. Calling rate

One hundred and eight vocalizing bouts and 108 intervocalizing periods were chosen randomly, 54 representatives from adults and 54 from sub-adult seals. The following characteristics were measured from vocalizing bouts and intervocalizing periods: number of units in a vocalizing bout, duration of vocalizing bouts, and duration of intervocalizing period. MANOVA was used to compare the calling rates to see whether sub-adults produced more calls in each vocalizing bout and whether their vocalizing bout was longer in duration than adults.

Using 30 min recordings from each of the adult and sub-adult seals the following characteristics were noted: frequency of occurrence of each broadcast call; ratio of high double trills to low double trills; number of pauses between consecutive vocalizations; number of intervocalizing periods; and total number of vocalizations. ANOVA was used to give comparisons between adult and sub-adults for duration and number of units in vocalizing and intervocalizing periods produced by adult and sub-adult male leopard seals. ANOVA was also used to give comparisons for calling rates in 30 min recordings for each of the call types between adult and sub-adults for every call type.

III. RESULTS

A. Calls in the repertoire

Adult male seals produced only the stereotyped broadcast calls while sub-adult male seals produced not only broadcast calls but also less stereotyped versions of these which I have called variants (Fig. 1). Examples of variant types are shown for the high double trill (HV) in Fig. 2, for the medium single trill (MV) in Fig. 3, and for the low double trill and hoot with single trill (LV) in Fig. 4. There

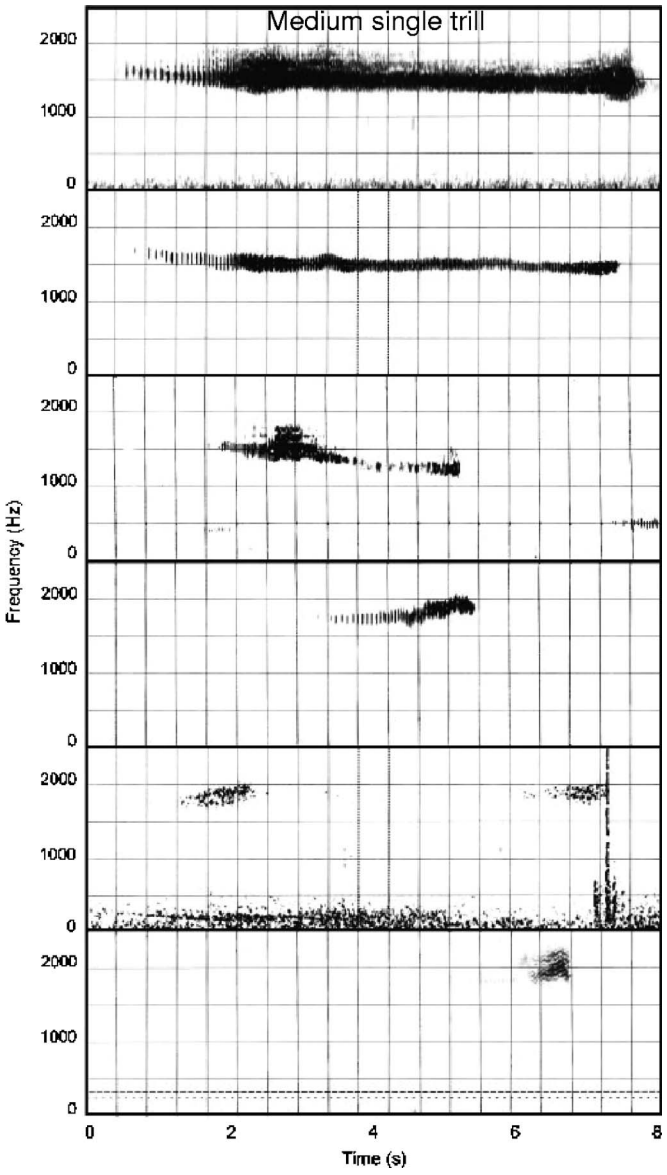


FIG. 3. Sonagrams of a leopard seal medium single trill (M) and examples of the five types of variants of this call (MV).

was just one example of a variant of the low descending trill (DV) so this is not shown here. Although sub-adult males produced variant calls (30% of all calls, 134 out of 450), by far the more commonly (Log-linear analysis: $X^2=76.94$, $d.f.=1$, $p<0.000001$) produced calls (70% of all calls, 316 out of 450) were the stereotyped broadcast versions. Sub-adult seals produced LV's, the low frequency variants, significantly (ANOVA: $F_{3,16}=8.78$, $p=0.001$) more often than other variant types (Fig. 1) however the LV category combines variants of both the low double trill and the hoot with the single trill. There was no significant difference (Log-linear analysis: $X^2=0.013$, $d.f.=4$, $p=1.00$) among the five sub-adult seals in the frequency with which they produced broadcast or variant calls.

B. Acoustic characteristics of broadcast calls

High double trills (H): produced by adult seals had significantly different (MANOVA: Wilks' lambda$_{3,245}=0.74$, Rao's $R=21.22$, $p<0.001$) frequency and temporal characteristics compared to those produced by sub-adult males. The high double trills of adult seals have significantly (Table III) lower maximum frequencies and bandwidth frequency, but longer duration. There is no significant difference in the duration of the first part of the call or the minimum frequencies.

Medium single trills (M): adult seal calls are not significantly different (MANOVA: Wilks' lambda$_{(3,133)}=0.989$, Rao's $R=0.373$, $p<0.827$) to the calls produced by sub-adult seals (Table III).

Low descending trills (D): produced by adult seals had a significantly lower minimum frequency compared to the calls of the sub-adult seals (MANOVA: Wilks' lambda$_{3,52}=0.72$, Rao's $R=6.87$, $p<0.001$). There is no significant difference in the maximum frequency, bandwidth, or the total duration (Table III).

Low double trills (L): most temporal and frequency characteristics measured were significantly different (MANOVA: Wilks' lambda$_{5,243}=0.20$, Rao's $R=160.42$, $p<0.001$) in sub-adult and adult low double trills. The low double trill calls of adult seals have significantly lower frequencies of the narrow-band component, higher frequencies of Part 1 and Part 2, longer total duration as well as longer Parts 1 and 2. There is no significant difference in the duration of the narrow-band component (Table III). The frequencies of Part 1 were significantly higher (304±25 Hz) than those of Part 2 (296±26 Hz) for the sub-adults (ANOVA: $F_{1,248}=6.54$, $p=0.01$) although there was no difference between the frequencies of Part 1 and Part 2 for adult seals (Table III).

Hoot with a single trill (O): produced by adult seals had significantly different frequency and temporal characteristics compared to those produced by sub-adult seals (MANOVA: Wilks' lambda$_{4,72}=0.72$, Rao's $R=5.61$, $p<0.001$). Hoot with a single trill calls of adult seals have significantly narrower bandwidths and longer durations of both Parts 1 and 2. There is no significant difference between the maximum or minimum frequencies, the frequency of the narrow-band components, or the total duration (Table III).

C. Calling rate

Sub-adult seals produce significantly (MANOVA: Wilks' lambda$_{3,104}=0.66$, Rao's $R=17.62$, $p<0.001$) more calls in each vocalizing bout (number of units in a vocalizing bout) and their vocalizing bouts are longer in duration than adult seals. There is no difference in the duration of the intervocalizing periods between adults and sub-adults (Table IV). Sub-adult seals produce significantly more high double trills, pauses, and total number of calls than the adults however there was no difference in the rates of any of the other call types or the number of bouts (Table V).

IV. DISCUSSION

Adult male leopard seals in the Prydz Bay region, Eastern Antarctica, produce five stereotyped broadcast calls. The sub-adult male seals, however, produce both 5 broadcast calls and the 4 variant categories, for a total of 9 general call types and if the different types of variants subjectively iden-

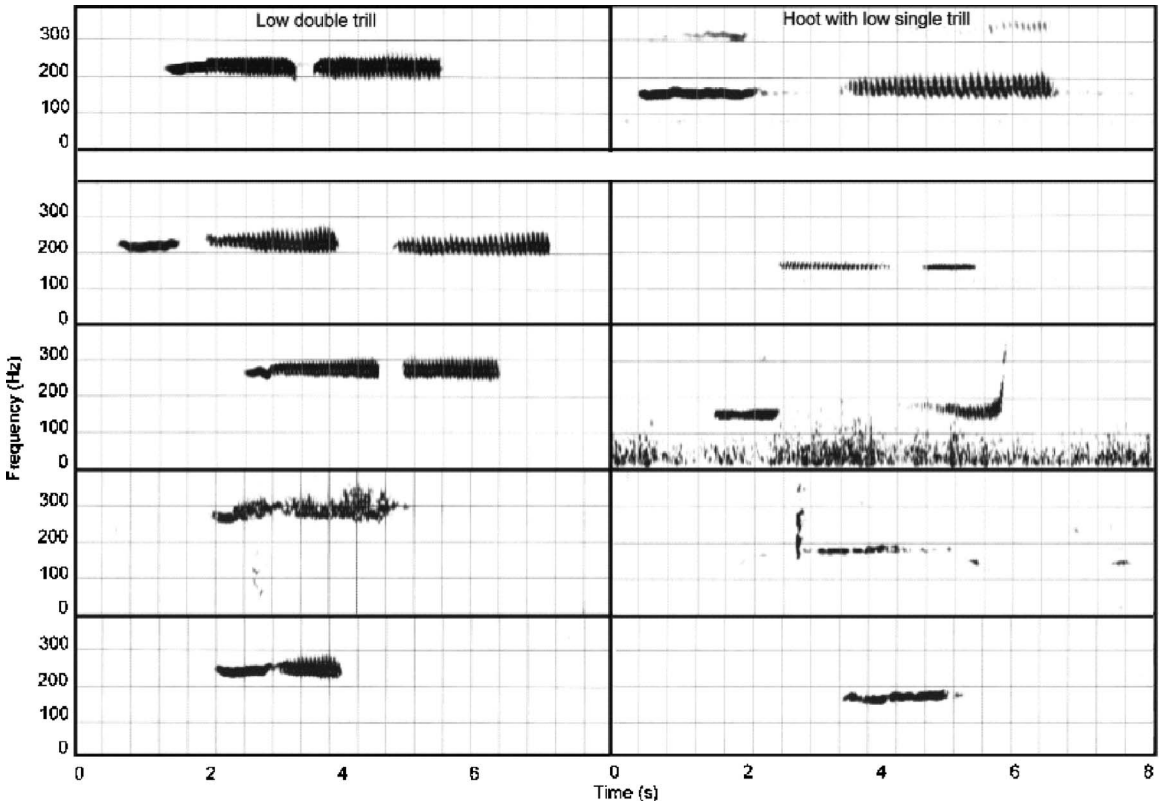


FIG. 4. Sonagrams of a leopard seal low double trill (L) and hoot with a single trill (O) and examples of the eight types of variants of these low frequency calls (LVs).

tified within each variant category were included the sub-adults repertoire for this study would extend to up to 26 different calls. Irrespective of the level of categorization the repertoire of the sub-adults are much greater than those of the adults. The repertoires of juvenile animals of many species including birds (Nelson and Marler, 1993; Catchpole and Slater, 1995; Podos *et al.*, 1999), wolves (Coscia *et al.*, 1991), and the elephant seal (Shipley *et al.*, 1986) contain a greater number of call types compared to those of adults. This has been referred to as an overproduction stage (Nelson and Marler, 1993) where young animals produce a wide range of vocalizations memorized at an early age or imitated from nearby adults (Nottebohm, 1972; Ralls *et al.*, 1985; Catchpole and Slater, 1995; Podos *et al.*, 1999). Overproduction allows young animals to experiment with their repertoire (Marler and Peters, 1981; De Wolfe *et al.*, 1989). Sub-adult animals in this study were of similar age class, approximately 1–3 years, and there was a small sample size. The described repertoire size of sub-adult male seals may be increased further in studies which include younger animals and/or more sub-adults.

Although sub-adults produced a range of sound types, they all resembled the classic hoot and trill-type sounds produced by adults and with the exception of the low double trill and the hoot with single trill they were easily categorized into variants of each particular adult broadcast call. All forms of the adult[1] broadcast calls are in the repertoire of the sub-adult seals so it is presumed that the younger seals do not need to practice these calls in order to produce them. Perhaps the sub-adults are learning the appropriate context(s) in which to produce the more adult-like sounds. This would suggest that the repertoire of the adult seals is developed through contextual rather than production learning. The younger animals develop their adult-like repertoires by selectively learning to produce some of the calls out of the original broader repertoire. The young leopard seals would be exposed to adult male leopard seal calls when they first enter the water postweaning. Leopard seal pups whelp from late October (Southwell *et al.*, 2003) coinciding with the period when adult males are just becoming vocal underwater. Whether this reflects usage learning, where an existing signal is produced in a new context as a result of experience with the usage of signals by other individuals, or comprehension learning, where a receiver comes to extract a novel meaning from a signal as a result of experience with the usage of signals by other individuals (Janik and Slater, 2000), is undetermined because we do not know the context in which the calls are used or their meaning.

The greatest age-related differences were between the acoustic characteristics of the low and high double trills compared to the other calls, the medium single trill and the hoot with a single trill. This may be because there is greater interindividual stereotypy in the high and the low double trills (Rogers and Cato, 2002) and perhaps less so in the other calls. Again, there was no simple age-related distinction made between the temporal or frequency characteristics of the low descending trill, however there was an impression that the calls of the adults dropped with individually distinctive slopes while those of the sub-adults tended to be more similar to one another. Perhaps the younger seals are trying

TABLE III. ANOVA comparison between acoustic characteristics measured from the five adult and five sub-adult male leopard seal calls, data from variant calls are not included.[a]

Variable	Adults mean±s.d.	Sub-adults mean±s.d.	F statistic	*d.f.*	*p* value *significant
High double trill (H)					
Min (Hz)	2684±72	2666±199	0.834	1,248	0.361
Max (Hz)	3429±182	3644±439	25.661	1,248	0.001*
BW (Hz)	745±204	978±326	45.670	1,248	0.001*
Dur 1 (s)	2.25±0.88	2.11±0.79	1.887	1,248	0.171
Dur 2 (s)	2.29±0.60	1.66±0.56	66.041	1,248	0.001*
Dur (s)	4.51±1.31	4.01±1.30	9.237	1,248	0.002*
Medium single trill (M)					
Min (Hz)	1552±164	1567±181	0.248	1,136	0.619
Max (Hz)	1979±209	2001±203	0.405	1,136	0.526
BW (Hz)	421±84	434±85	0.839	1,136	0.361
Dur (s)	4.29±1.82	4.26±1.81	0.113	1,136	0.915
Low descending trill (D)					
Min (Hz)	311±10	325±23	8.410	1,54	0.005*
Max (Hz)	629±52	640±18	1.132	1,54	0.292
BW (Hz)	317±54	315±29	0.061	1,54	0.805
Dur (s)	6.18±1.16	5.90±0.53	1.349	1,54	0.250
Low double trill (L)					
FNB (Hz)	312±19	324±29	14.231	1,248	0.001*
Frq 1 (Hz)	328±21	304±25	64.588	1,248	0.001*
Frq 2 (Hz)	328±21	296±26	114.154	1,248	0.001*
Dur NB (s)	0.48±0.14	0.46±0.14	1.197	1,248	0.275
Dur 1 (s)	1.90±0.49	1.72±0.48	8.760	1,248	0.003*
Dur 2 (s)	1.89±0.42	1.54±0.39	46.797	1,248	0.001*
Dur (s)	4.05±0.85	3.48±0.72	32.399	1,248	0.001*
Hoot with a single trill (O)					
FNB (Hz)	226±25	216±27	2.680	1,76	0.106
Min (Hz)	195±21	187±27	2.611	1,76	0.110
Max (Hz)	274±29	276±24	0.060	1, 76	0.807
BW (Hz)	79±17	89±18	6.746	1, 76	0.011*
Dur 1 (s)	1.61±0.44	1.40±0.37	5.262	1, 76	0.024*
Dur 2 (s)	3.05±0.78	2.72±0.61	4.400	1, 76	0.039*
Dur (s)	6.29±1.73	5.74±1.73	1.672	1, 76	0.120

Min: minimum frequency; Max: maximum frequency; BW: bandwidth: difference between the maximum and minimum frequencies; FNB: frequency of the narrow-band component; Frq 1: maximum frequency of Part 1; Frq 2: maximum frequency of Part 2; Dur NB: duration of the narrow-band component; Dur 1: duration of Part 1; Dur 2: duration of Part 2; Dur: total duration.

to initially produce a more stylized call whereas the adults were producing more individually distinctive low descending trills.

TABLE IV. ANOVA comparison between duration and number of units in vocalizing and inter-vocalizing periods produced by adult and sub-adult male leopard seals (an asterisk denotes significant at 0.05 level).

Variable	Adults mean±s.d.	Sub-adults mean±s.d.	*F* statistic	*d.f.*	*p* value
Number of units in a vocalizing bout	9.7±1.9	14.8±5.0	48.066	1,106	0.001*
Duration of vocalizing bouts	138.4±43.8 s	184.5±84.7 s	12.610	1,106	0.001*
Duration of inter vocalizing period	79.4±28.5 s	91.7±43.8 s	3.003	1, 106	0.086

Sub-adult seals called more often, their vocalizing bouts were longer and the time spent resting between bouts, the intervocalizing periods, were no longer than those of adults. Although sub-adult seals called more often they did not produce more of each type of call but selectively produced more high double trills. Similarly, juvenile elephant seal bulls produce the same basic calls but lack the patterns characteristic of adult bulls (Shipley *et al.*, 1986).

As adult male leopard seals have highly distinctive and stereotyped vocalizations, acoustic surveys could be a powerful tool for estimating density for this species. However as sub-adult male leopard seals have more calls in their repertoire and they produce some calls more often caution must be taken not to include these animals in sampling. If sampling regimes unknowingly include recordings from sub-adult seals, the repertoire for that region and/or the number of animals calling may be artificially exaggerated, which suggests grave problems for acoustic surveying. A similar situation may exist for studies looking at geographic varia-

TABLE V. ANOVA comparison between calling rates in 30 min recordings from adult and sub-adult male leopard seals (an asterisk denotes significant at 0.05 level).

Variable	Adults mean±s.d.	Sub-adults mean±s.d.	*F* statistic	*d.f.*	*p* value
Number of high double trill (H)s	31.4±6.5	43.6±7.7	7.253	1,8	0.027*
Number of medium single trills (M)	10.6±4.3	12.6±9.8	0.176	1,8	0.686
Number of low descending trills (D)	4.0±1.4	2.0±1.9	3.636	1,8	0.093
Number of low double trills (L)	31.0±2.5	44.6±13.4	4.951	1,8	0.057
Number of hoots with single trill (O)	7.4±2.1	9.6±2.5	2.283	1,8	0.169
Ratio of high to low double trills	1.0±0.1	1.0±0.2	0.034	1,8	0.857
Number of pauses between consecutive vocalizations	74.8±13.1	106.0±22.4	7.208	1,8	0.028*
Number of intervocalizing periods	8.8±2.6	7.2±1.1	1.620	1,8	0.239
Total number of vocalizations	84.4±12.6	113.8±23.6	6.049	1,8	0.039*

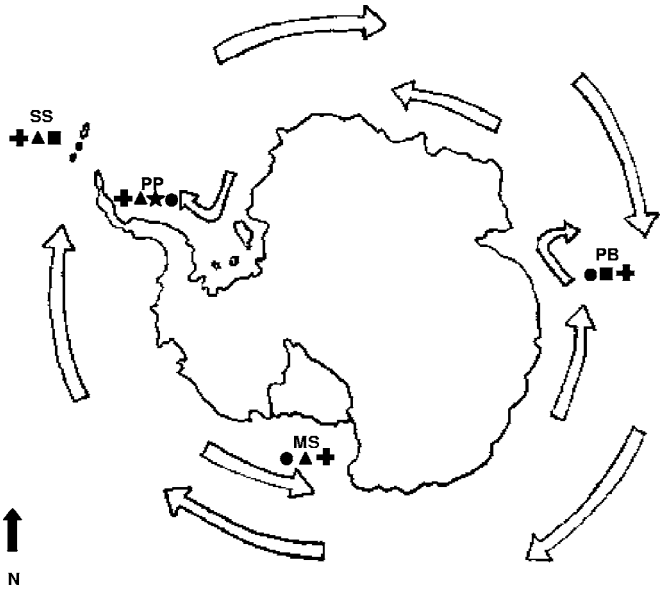


FIG. 5. Leopard seal call repertoires recorded form different Antarctic areas; South Shetland Islands (SS) (Stirling and Siniff, 1979), Palmer Peninsula (PP), McMurdo Sound (MS) (Thomas and Golladay, 1995) and Prydz Bay (PB) (Rogers *et al.*, 1995). Westward drift of the ice is the ice movement parallel to the bathymetry that hugs near the Antarctic continent where the eastward drift of the ice is further out from the Antarctic continent. Call 1 =low double trill and 2=high double trill (+), call 3=medium single trill and 5=low descending trill (●), Call 4=medium double trill (▲), Calls 6, 7, 8, 9, (★), and Call 10=hoot with a single low trill (■).

tion. Thomas and Golladay (1995) found geographic variation when comparing the acoustic repertoire of leopard seals from McMurdo Sound with those of Palmer Peninsula. They found that populations from McMurdo Sound produced five call types where those from Palmer Peninsula produced nine (Thomas and Golladay, 1995). An earlier study on a population found around the South Shetland Islands showed that leopard seals in that region produced four call types (Stirling and Siniff, 1979) while another study of leopard seals in Prydz Bay, Eastern Antarctica, showed that males produced five call types (Rogers *et al.*, 1995). The calls of male leopard seals in the Prydz Bay area showed a high degree of stereotypy with little intraindividual and/or interindividual variation (Rogers and Cato, 2002) so that the five calls of the adult seals were representative of the repertoire of that population. In the other studies extensive recordings were made throughout the summer and it was believed that the vocalization repertoires of leopard seals had been described fully for each region. In combining all four Antarctic regions studied ten different call types have been reported for leopard seals (Fig. 5) and there are patterns in the distribution of the calls.

Two calls, the high and low double trills, are likely to have a circumpolar presence as they are shared with all four regions. Thomas and Golladay (1995) give the prefix P to denote that a call was from Palmer Peninsula and an M if it was from McMurdo Sound, and where the low double trill is similar in form to the P1 and M1 the high double trill is similar to P2 and M2. The high and low double trills are not only the most stereotyped of the calls but they are also the most frequently heard sound types in both the South Shetland Islands (Stirling and Siniff, 1979) and in Prydz Bay (Rogers and Cato, 2002).

There are four calls, P6, P7, P8, and P9 that are specific to only one region, Palmer Peninsula, whose population has the greatest repertoire, nine calls as opposed to four or five from other regions. The radical frequency changes within calls P7, P8, and P9 are typical of calls produced by subadult male seals in Prydz Bay. As the age classes of the individuals producing the calls were unknown perhaps they were produced by younger animals. Whereas the call P6 has the stereotyped characteristics of an adult call so it is likely to be an adult call specific to the Palmer Peninsula region. If this were the case the Palmer Peninsula repertoire would be restricted to six rather than nine call types, however this can be confirmed only with further study.

The remaining four calls are found in some areas but not in others and the sharing of calls in between regions, indicated in Fig 5, suggests that some animals have moved between populations. Thomas and Golladays' (1995) call type 3, the medium single trill, and call type 5, the low descending trill, are found at Palmer Peninsula, McMurdo Sound, and in Prydz Bay. This distribution pattern (Fig. 5) may reflect that seals have moved with the westward drift of the ice which is an ice movement parallel to the bathymetry that hugs near the Antarctic continent (Heil and Alison, 1999). Call type 4, the medium double trill, was found at Palmer Peninsula, McMurdo Sound, and the South Shetland Islands. This distribution pattern suggests that seals may have moved in the opposite direction with the eastward drift of the ice which is north of the zonal shear zone further away from the Antarctic continent (Heil and Alison, 1999). The hoot with a single trill is found unusually in only the South Shetland Islands and in Prydz Bay which are on opposing positions of the Antarctic continent but it is not found at either of the two other sites which sit between them (Fig 5).

The leopard seal is a highly mobile species and young animals are capable of traveling long distances (Rounsevell and Pemberton, 1994) but if large proportions of the leopard seal population were to move between areas, similar call repertoires would be expected from all regions so the existence of geographic variation suggests that there is likely to be some segregation of populations. Recent satellite tracking work has shown that while sub-adult leopard seals are highly mobile the adult seals are relatively sedentary, spending long periods of time remaining within defined areas (Rogers *et al.*, 2005). Leopard seals haul-out to sleep, rest, and/or pup on the ice floes of the pack ice, a highly dynamic region where ice floes move with the currents and eddies. To remain within an area the adult leopard seals swim back after having drifted while hauled out (Rogers *et al.*, 2005). The currents, and their associated eddy systems, as well as the Antarctic Peninsula itself appear to be barriers to progressive mixing of populations.

Acoustic analysis can be a powerful tool for surveying and evaluating geographic variation between populations. However, for the leopard seal caution must be taken with sampling regimes. As sub-adult male leopard seals have more calls in their repertoires due to overproduction, their broadcast calls are less stereotyped in acoustic characteristics and specific calls have more variable calling rates, caution must be taken in including sub-adult animals in sampling for geographic variation studies. If sampling regimes unknowingly include recordings from sub-adult seals, the repertoire for that region may be artificially exaggerated, and the acoustic characteristics of calls and calling rates more variable.

ACKNOWLEDGMENTS

I thank the Australian Antarctic Division for providing logistic support in the Antarctic; the personnel at Davis station particularly: Paul Egan, Sel Peacock, Adrian Payt, and Sue Robinson. Thanks must also go to Tara Cheesman and Kai Schmidt for assistance with the acoustic analysis and Carolyn Hogg, Michaela Ciaglia, and Toby Scrivener for the figures. This study was supported financially by the Sea World Research and Rescue Foundation, Scott Foundation, and the Antarctic Scientific Advisory Committee.

Caldwell, D. K., and Caldwell, M. C. (**1972**). "Senses and Communication," in *Mammals of the Sea: Biology and Medicine*, edited by S. H. Ridgway (Thomas, Springfield), pp. 466–501.

Catchpole, C. K., and Slater, P. J. B. (**1995**). "Themes and Variations," in *Bird Song: Biological Themes and Variation*, edited by C. K. Catchpole and P. J. B. Slater (Cambridge University Press, Cambridge), pp. 164–173.

Coscia, E. M., Phillips, D. P., and Fentress, J. C. (**1991**). "Spectral analysis of noenatal wolf *Canis lupis* vocalizations," Bioacoustics **3**, 275–293.

De Wolfe, B. B., Baptista, L. F., and Petrinovich, L. (**1989**). "Song development and territory establishment in Nuttall's White-crowned sparrows," Condor **91**, 397–407.

Ford, J. K. (**1991**). "Vocal traditions among resident killer whales (*Orcinus orca*) in coastal waters of British Columbia," Can. J. Zool. **69**, 1454–1483.

Heil, P., and Allison, I. (**1999**). "The pattern and variability of Antarctic sea-ice drift in the Indian Ocean and western Pacific sectors," J. Geophys. Res., [Oceans] **104**, 15789–15802.

Janik, V. M., and Slater, P. J. B. (**1997**). "Vocal learning in mammals," Advances in the Study of Behavior **26**, 59–99.

Janik, V. M., and Slater, P. J. B. (**2000**). "The different roles of social learning in vocal communication," Anim. Behav. **60**, 1–11.

Jones, G., and Ransome, R. D. (**1993**). "Echolocation calls of bats change over a lifetime are influenced by maternal effects," Proc. R. Soc. London, Ser. B **252B**, 125–128.

Marler, P., and Peters, S. (**1981**). "Sparrows learn adult song and more from memory," Science **213**, 780–782.

Morrice, M. G., Burton, H. R., and Green, K. (**1994**). "Microgeographic variation and songs in the underwater vocalisation repertoire of the Weddell seal (*Leptonychotes weddellii*) from the Vestfold Hills, Antarctica," Polar Biology **14**, 441–446.

Nelson, B. S., and Marler, P. (**1993**). "Innate recognition of song in white-crowned sparrows: A role in selective vocal learning?," Anim. Behav. **46**, 806–808.

Nottebohm, F. (**1972**). "The origins of vocal learning," Am. Nat. **106**, 116–135.

Payne, K. B., and Payne, R. (**1985**). "Large scale changes over 19 years in songs of humpback whales in Bermuda," Z. Tierpsychol **86**, 89–114.

Payne, K. B., Tyack, P. L., and Payne, R. (**1983**). "Progressive changes in the songs of humpback whales (*Megaptera novaeangliae*): A detailed analysis of two seasons in Hawaii," in *Communication and Behaviour in Whales*, edited by R. Payne (Westview, Boulder, Co), pp. 9–57.

Podos, J., Nowicki, S., and Peters, S., (**1999**). "Permissiveness in the learning and development of song syntax in swamp sparrows," Anim. Behav. **58**, 93–103.

Ralls, K., Fiorelli, P., and Gish, S., (**1985**). "Vocalizations and vocal mimicry in captive harbor seals, Phoca vitulina," Can. J. Zool. **63**, 1050–1056.

Reiss, D., and McCowan, B., (**1993**). "Spontaneous vocal mimicry and production by bottlenose dolphins (*Tursiops truncatus*): Evidence for vocal learning," J. Comp. Psychol. **107**, 301–312.

Rogers, T. L., and Bryden, M. M., (**1997**). "Density and haul-out behaviour of leopard seals (*Hydrurga leptonyx*) in Prydz Bay, Antarctica," Marine Mammal Sci. **13**, 293–302.

Rogers, T. L., and Cato, D. H., (**2002**). "Individual variation in the acoustic behaviour of the adult male leopard seals, Hydrurga leptonyx," Behaviour **139**, 1267–1286.

Rogers, T. L., Cato, D. H., and Bryden, M. M., (**1995**). "Underwater vocal repertoire of the leopard seal (*Hydrurga leptonyx*) in Prydz Bay, Antarctica," in *Sensory Systems of Aquatic Mammals*, edited by R. A. Kastelein, J. A. Thomas, and P. E. Nachtigall (De Spil, Woerden, The Netherlands), pp. 223–236.

Rogers, T. L., Cato, D. H., and Bryden, M. M., (**1996**). "Behavioural significance of underwater vocalizations of captive leopard seals, Hydrurga leptonyx," Marine Mammal Sci. **12**, 414–427.

Rogers, T. L., Hogg, C. J., and Irvine, A., (**2005**). "Spatial movement of adult leopard seals (*Hydrurga leptonyx*) in Prydz Bay, Eastern Antarctica," Polar Biology **28**, 456–463.

Rounsevell, D., and Pemberton, D., (**1994**). "The status and seasonal occurrence of leopard seals, *Hydrurga leptonyx*, in Tasmanian waters," Australian Mammalogy **17**, 97–102.

Shipley, C., Hines, M., and Buchwald, J. S., (**1986**). "Vocalizations of northern elephant seal bulls: Development of adult call characteristics during puberty," J. Mammal. **67**, 526–536.

Siniff, D. B., and Stone, S., (**1985**). "The role of the leopard seal in the tropho-dynamics of the Antarctic marine ecosystem," in *Antarctic Nutrient Cycles and Food Webs*, edited by W. R. Siegfried, P. R. Condy, and R. M. Laws (Springer, Berlin), pp. 555–560.

Southwell, C., Kerry, K., Ensor, P., Woehler, E. J., and Rogers, T., (**2003**). "The timing of pupping by pack-ice seals in East Antarctica," Polar Biology **26**, 648–652.

Stirling, I., and Siniff, D. B., (**1979**). "Underwater vocalizations of leopard seals (*Hydrurga leptonyx*) and crabeater seals (*Lobodon carcinophagus*) near the South Shetland Islands, Antarctica," Can. J. Zool. **57**, 1244–1248.

Thomas, J. A., and DeMaster, D. P., (**1982**). "An acoustic technique for determining diurnal activities in leopard (*Hydrurga leptonyx*) and crabeater (*Lobodon carcinophagus*) seal," Can. J. Zool. **60**, 2028–2031.

Thomas, J. A., and Golladay, C. L., (**1995**). "Geographic variation in leopard seal (*Hydrurga leptonyx*) underwater vocalizations," in *Sensory Systems of Aquatic Mammals*, edited by R. A. Kastelein, J. A. Thomas, and P. E. Nachtigall (De Spil, Woerden, The Netherlands), pp 201–221.

Würsig, B., and Clark C., (**1993**). "Behavior," in *The Bowhead Whale*, edited by J. J. Burns, J. J. Montague, and C. J. Cowles (Society for Marine Mammalogy, Lawrence, KS), Vol. **2** pp. 157–199.

Underwater sound pressure variation and bottlenose dolphin (*Tursiops truncatus*) hearing thresholds in a small pool[a)]

James J. Finneran
U.S. Navy Marine Mammal Program, Space and Naval Warfare Systems Center, San Diego, Code 2351, 49620 Beluga Road, San Diego, California 92152

Carolyn E. Schlundt
EDO Professional Services, 3276 Rosecrans Street, San Diego, California 92110

(Received 11 January 2007; revised 15 April 2007; accepted 1 May 2007)

Studies of underwater hearing are often hampered by the behavior of sound waves in small experimental tanks. At lower frequencies, tank dimensions are often not sufficient for free field conditions, resulting in large spatial variations of sound pressure. These effects may be mitigated somewhat by increasing the frequency bandwidth of the sound stimulus, so effects of multipath interference average out over many frequencies. In this study, acoustic fields and bottlenose dolphin (*Tursiops truncatus*) hearing thresholds were compared for pure tone and frequency modulated signals. Experiments were conducted in a vinyl-walled, seawater-filled pool approximately 3.7 $\times 6 \times 1.5$ m. Acoustic signals were pure tone and linear and sinusoidal frequency modulated tones with bandwidths/modulation depths of 1%, 2%, 5%, 10%, and 20%. Thirteen center frequencies were tested between 1 and 100 kHz. Acoustic fields were measured (without the dolphin present) at three water depths over a 60×65 cm grid with a 5-cm spacing. Hearing thresholds were measured using a behavioral response paradigm and up/down staircase technique. The use of FM signals significantly improved the sound field without substantially affecting the measured hearing thresholds. [DOI: 10.1121/1.2743158]

PACS number(s): 43.80.Lb [WWA] Pages: 606–614

I. INTRODUCTION

The problems inherent in underwater bioacoustical testing in small tanks have been well known for years (e.g., Parvulescu, 1964, 1966): long acoustic wavelengths make sound absorption at tank boundaries difficult, resulting in complex reverberant sound fields in all but the largest tanks or specially designed enclosures (e.g., see Bobber and Beatty, 1959; Albers, 1965; Bobber, 1970; Miller, 2006), or at very high frequencies. Since, at any particular location, the sound pressure consists of the sum of the direct component and multiple reflected waves, single frequency tones will produce complex interference patterns. The resulting sound pressures may vary dramatically over relatively small spatial distances (fractions of the sound wavelength) and produce erratic pressure gradients (and thus sound particle motions), both of which confound attempts to accurately quantify sound stimuli during experimental manipulations. Despite these problems, however, relatively small tanks are often used for bioacoustic tests because of practical limitations, ambient noise concerns, or specific experimental design requirements. Although in most cases experimenters are aware of the potential issues created by complex acoustic fields, spatial variations in sound pressure or particle motion have rarely been quantified. In most cases received sound levels have been specified using a single hydrophone positioned at the subject's location, and descriptions of the expected sound level variability within the volume of interest have been vague.

To help mitigate sound propagation problems within small pools, some recent studies featured acoustic signals traditionally used in sonar systems, such as frequency modulated (FM) tones, in lieu of pure tone stimuli (Kastelein *et al.*, 2002a, b, 2003, 2005). FM stimuli possess a range of frequencies, so the overall sound pressure (summed across the FM signal bandwidth) exhibits less variability with spatial location compared to a pure tone. Sound at a particular frequency would still exhibit minima/maxima at different locations, but across a number of frequencies this effect would tend to be reduced, with a larger FM bandwidth resulting in less spatial variation in the overall sound pressure. Although the theory of operation of FM signals in this manner is well established, expected improvements in sound fields resulting from FM stimuli have not been quantified in small pools over the range of frequencies typically of interest in marine bioacoustics.

This paper describes a series of experimental measurements designed to quantify single frequency and FM sound fields within a relatively small pool. The resulting spatial variations in sound pressure are presented and compared over the frequency range from 1 to 100 kHz. The suitability of FM stimuli for hearing assessment was investigated by measuring hearing thresholds in a bottlenose dolphin (*Tursiops truncatus*) using FM and single frequency stimuli.

[a)]Portions of these data were presented in "Acoustic field measurements and bottlenose dolphin hearing thresholds using single frequency and frequency modulated tones," at the 4th Joint Meeting of the Acoustical Society of America and the Acoustical Society of Japan, December 2006, Honolulu, HI.

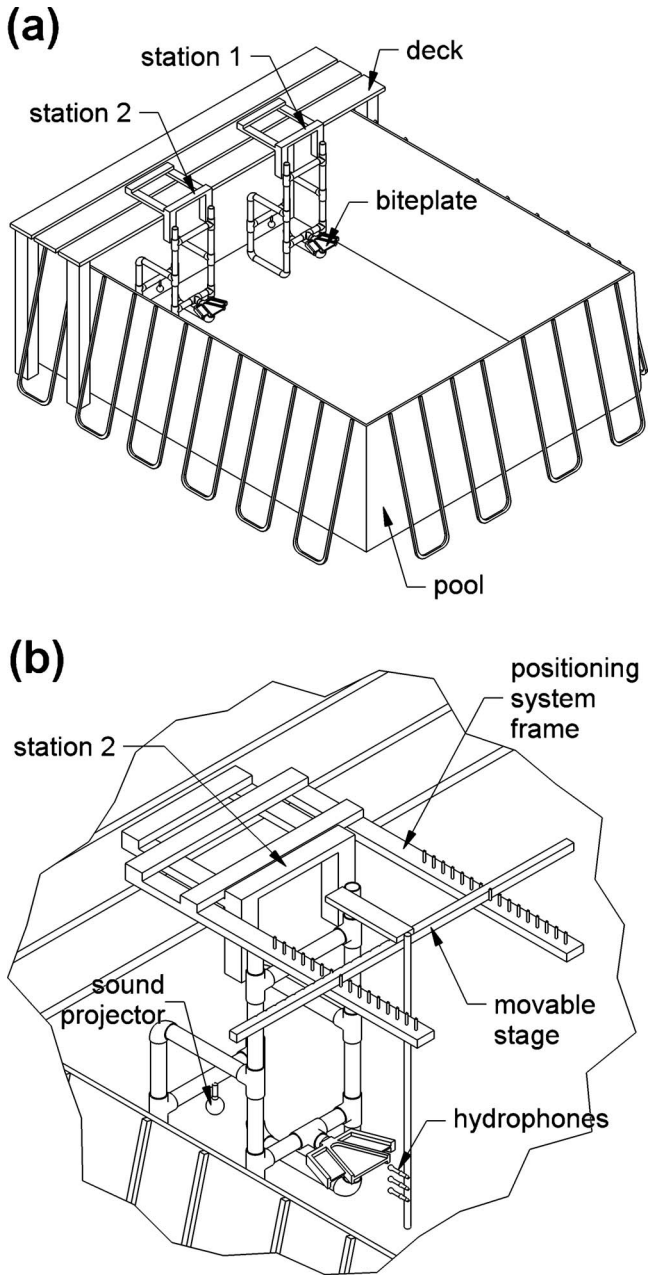


FIG. 1. (a) Schematic of the $3.7\times6\times1.5$ m pool showing the underwater stations. (b) Detail of station 2 showing the two-axis positioning system used to map the acoustic field. Three hydrophones were used to simultaneously measure the acoustic field in three horizontal planes.

II. METHODS

A. Test environment

Experiments were conducted in an approximately 3.7 $\times6\times1.5$ m vinyl-walled, above-ground pool (Fig. 1). The pool walls were supported only at the upper rim and floor and "bulged" in response to hydrostatic pressure; the pool walls were therefore not flat, parallel surfaces. The pool was filled with seawater, which was filtered using a closed-circuit pump and sand filter system. The pump was turned off during all experimental measurements. Water depth varied somewhat from day to day. The pool was located within a larger room whose walls and ceiling were treated with sound absorbing foam to reduce room reverberation. Ambient noise spectral density levels in the pool were approximately <60, <52, and <40 dB *re* 1 μPa^2/Hz over the frequency ranges 0.1–1, 1–3, and >3 kHz, respectively.

A wooden deck located at one end of the pool supported two "stations" constructed of PVC pipe. The pool contained two stations for use in temporary threshold shift (TTS) experiments where the fatiguing source used to induce TTS (located at station 1) is spatially separated from the hearing test location (station 2) (e.g., Schlundt *et al.*, 2000). Each station contained an underwater sound projector, receiving hydrophone(s), video camera, and a plastic, neoprene-covered "biteplate" upon which the dolphin subject was trained to position. The biteplates were located at mid-depth in the pool.

B. Acoustic field maps

1. Signals

Acoustic signals were digitally generated using a multifunction data acquisition card (National Instruments NI PCI-6070E) within a personal computer, filtered (SRS SR-560), attenuated (TDT PA5), amplified (Hafler P4000), and output through a piezoelectric sound projector (ITC 1032). Waveforms consisted of 500-ms single frequency (pure) tones (PT), linear FM (LFM) tones, and sinusoidal FM (SFM) tones. Center frequencies were 1, 3, 4.5, 6, 9, 10, 20, 30, 40, 50, 60, 80, and 100 kHz. The selection of 4.5 kHz was based on its use during recent TTS experiments (Finneran *et al.*, 2005; Schlundt *et al.*, 2006).

The signal instantaneous voltage $v(t)$ was defined as

$$v(t) = A \sin \theta(t), \tag{1}$$

where A was the signal amplitude and $\theta(t)$ was the instantaneous phase angle (rad). For PT signals,

$$\theta(t) = 2\pi f_c t, \tag{2}$$

where f_c was the frequency (Hz). For LFM signals,

$$\theta(t) = 2\pi\left(f_1 t + \frac{Bf_c}{2T}t^2\right), \tag{3}$$

where f_1 and f_2 were the lower and upper frequencies, respectively, f_c was the center frequency, $f_c=(f_1+f_2)/2$, B was the bandwidth, $B=(f_2-f_1)/f_c$, and T was the sound duration. The bandwidth $B=1\%$, 2%, 5%, 10%, or 20%. For SFM signals,

$$\theta(t) = 2\pi\left(f_c t - \frac{Df_c}{2f_m}\right)\cos(2\pi f_m t), \tag{4}$$

where f_c was the carrier (center) frequency, f_m was the modulation frequency, and D was the modulation depth. The modulation depth was defined as the peak-peak frequency deviation expressed as a percentage of the carrier frequency. For comparison to other definitions of SFM signals, the frequency deviation (Δf) and the modulation index (β) may be calculated from $\Delta f=Df_c/2$ and $\beta=Df_c/(2f_m)$. The SFM modulation depth (D) was 1%, 2%, 5%, 10%, or 20% and the modulation frequency (f_m) was 10 Hz. From Carson's rule (Carson, 1963), the approximate bandwidths of the SFM signals were Df_c+2f_m. Power spectral densities for the

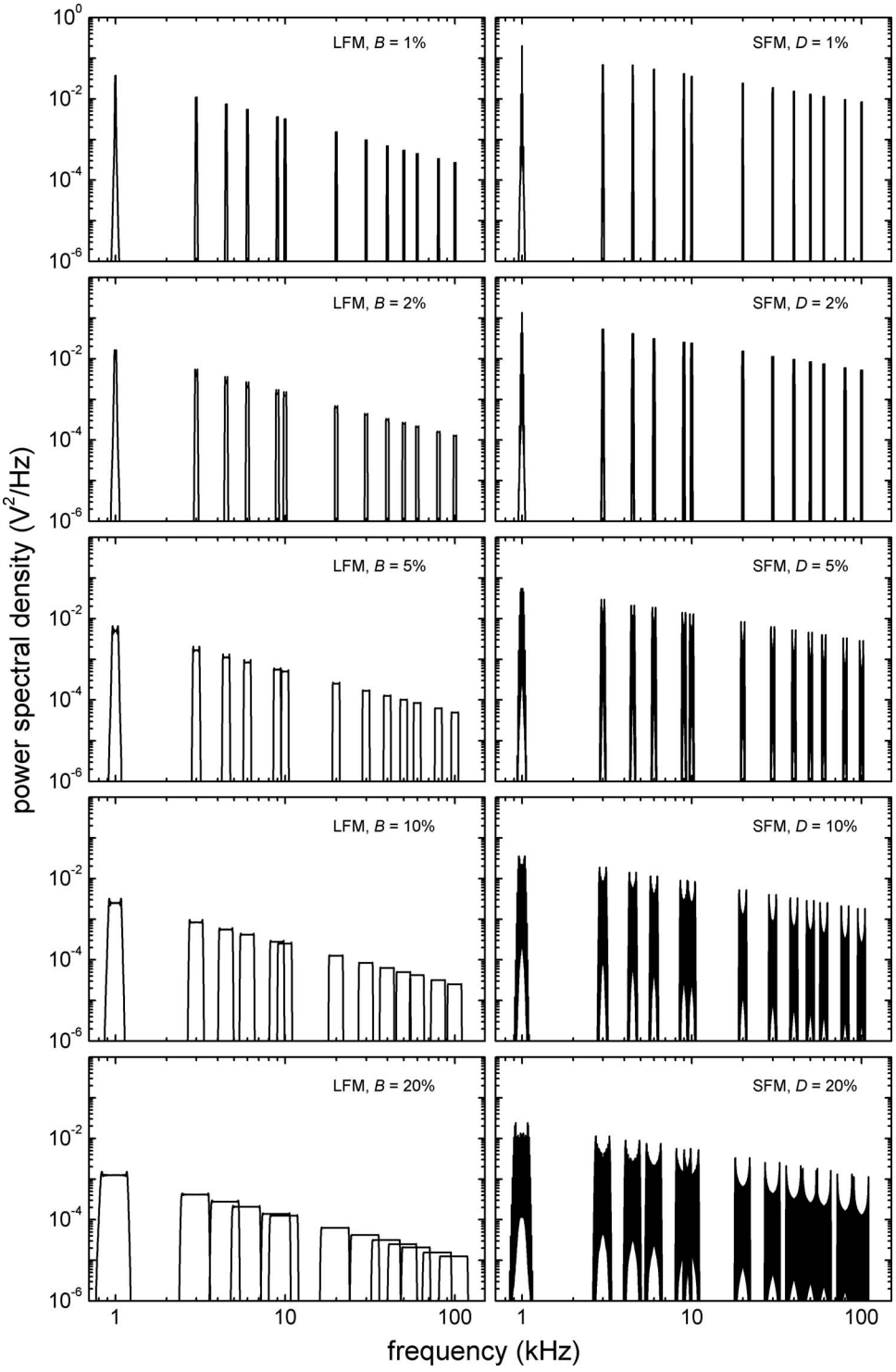


FIG. 2. Power spectral densities of the FM signals used for acoustic field measurements. Each signal had an amplitude of 1 V. LFM=linear frequency modulation, B=bandwidth, SFM =sinusoidal frequency modulation, D=modulation depth (see text).

waveforms as described in Eqs. (1)–(4) are shown in Fig. 2. For each waveform, the signal amplitude (A) was 1 V. For the specific values of f_c, D, and f_m, the modulation index varied from $0.5 \leq \beta \leq 1000$, thus in many cases the signals were not narrow-band FM (usually defined as $\beta < 1$) and at many f_c and D combinations would be considered wideband or ultra-wideband FM (Gerrits *et al.*, 2005). Many of the SFM spectra reveal the wideband FM nature—the lack of a peak at f_c and spectral peaks followed by sharp roll-offs at $f_c \pm D f_m/2$.

2. Acoustic measurements

To quantify the acoustic field within the pool, the sound pressure was measured as a function of position within three horizontal planes. The primary interest was on the region around station 2, where the auditory threshold testing occurred. Figure 3 shows the measurement geometry and coordinate system. The biteplate and dolphin outline were created from photos taken during threshold testing. Sound pressures were measured over a 65×60 cm grid at 5-cm intervals. A two-axis manual positioning system [Fig. 1(b)] was used to

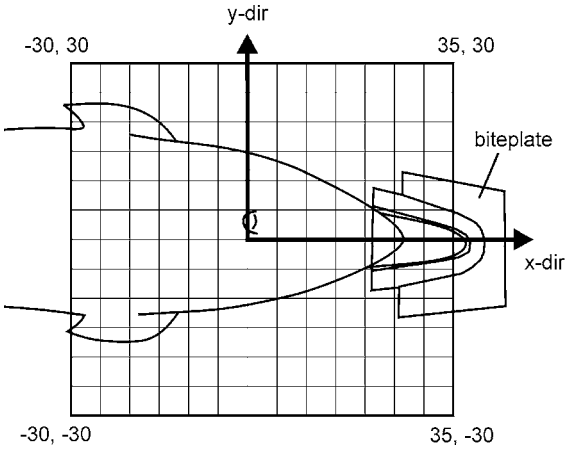


FIG. 3. Coordinate system and 65×60 cm grid used for the acoustic field measurements. The dolphin and biteplate outlines were created from overhead photographs taken during threshold testing.

simultaneously move three hydrophones (B&K 8105), arranged at different depths, point-by-point through the measurement grid in Fig. 3. The hydrophones were located at water depths $z=-10$, 0, and +10 cm relative to a depth just above the biteplate. Hydrophone signals were amplified and filtered (B&K 2635), digitized using the PCI-6070E, and stored on the computer. At each (x,y) position, custom software was used to sequentially generate acoustic signals and record the resulting sound pressure at each hydrophone along with the (x,y) position. The measurement position was then advanced and the process repeated. For measurements at $x \geq 20$ cm, removal of the lower portion of the vertical support post and the deepest hydrophone were required to allow clearance of the biteplate.

Measurements were made on two separate days separated by several months. Each measurement session, lasting approximately 9 h, featured seven different stimulus waveforms at 13 different center frequencies, resulting in 143 measurements at each (x,y) location (operator error and replication resulted in a total of 11 unique waveform conditions as described above).

3. Data analysis

The data consisted of 429 individual acoustic field maps, each defining the sound field for a single waveform and water depth. Each map consisted of a 13×14 ($z=0,10$ cm) or 10×14 ($z=-10$ cm) matrix of sound pressure levels (SPLs). Each matrix was normalized by subtracting the mean SPL over the matrix. The variance in the SPLs was quantified using the SPL range, defined as the maximum SPL minus the minimum SPL, and the SPL standard deviation.

C. Hearing tests

1. Subject

The subject (BLU) was a 41-year-old female bottlenose dolphin of approximately 200 kg. The subject had previously participated in cooperative psychophysical tasks, including auditory detection tasks. The subject was housed in floating netted enclosures located in San Diego Bay, CA, and trained to slide out of the water onto a foam mat for transport to the pool. The study followed a protocol approved by the Institutional Animal Care and Use Committee at the Space and Naval Warfare Systems Center, San Diego, and followed all applicable U.S. Department of Defense guidelines.

2. Stimuli

Sound stimuli were 500-ms (including 20 ms rise and fall times) single frequency, LFM ($B=10\%$), and SFM ($D=10\%$, $f_m=10$ Hz) tones as described in Eqs. (1)–(4). Center frequencies ranged from 3 to 80 kHz. Tones were generated using the same equipment used for the acoustic field measurements. Hearing test tones were calibrated before each session using two hydrophones (Reson TC4013) placed at (x,y,z) locations of $(-5,20,0)$ cm and $(-5,-20,0)$ cm without the subject present. Hydrophone signals were filtered and amplified (B&K 2692), then digitized using the PCI-6070E. Calibrated levels were based on the incoherent average of the two sound pressures. If calibration measurements revealed large (>10 dB) differences between the individual SPLs, station 2 was moved a few centimeters to either side along the deck and the calibration repeated. The goal was to avoid having either hydrophone at a known pressure maxima or minimum at the test frequency.

3. Procedure

Each session was divided into a number of trial blocks, each containing multiple trials. A light was toggled on/off to delineate each trial. Trial duration was 2 s, with a 4-s intertrial interval (trial start to next trial start). Fifty percent of the trials contained a hearing test tone beginning 500 ms after the light onset and 50% of the trials were no-stimulus (catch) trials, as determined by a Gellerman series (Gellerman, 1933). The subject was trained to whistle in response to hearing test tones and to remain quiet otherwise. Whistles occurring less than 50 ms after the tone onset or while the light was off were ignored (i.e., the response window was 0.55–2 s relative to the 2-s trial window). Tone presentations began at SPLs approximately 10 dB above expected thresholds and were adjusted from trial to trial using a modified up/down staircase technique (Cornsweet, 1962) with a 2-dB step size. After a variable number of trials, the subject was signaled to return to the surface for an amount of fish reward scaled to the performance during the trial block. The process was then repeated to obtain six reversal points, defined as a transition from a response to no-response, or vice versa, over consecutive stimulus present trials. Measurements at each frequency/waveform combination were repeated over at least three sessions to obtain a minimum of 18 reversals. Thresholds were based on the average SPL over the reversal points. False alarm rates were defined as the number of responses during stimulus-absent trials divided by the total number of stimulus-absent trials. Only the trials over which the reversal points occurred were used for the false alarm rate calculation. Custom software (Finneran, 2003) was used to generate and present the sound stimuli, record and view the subject's responses, analyze the results, and archive the data.

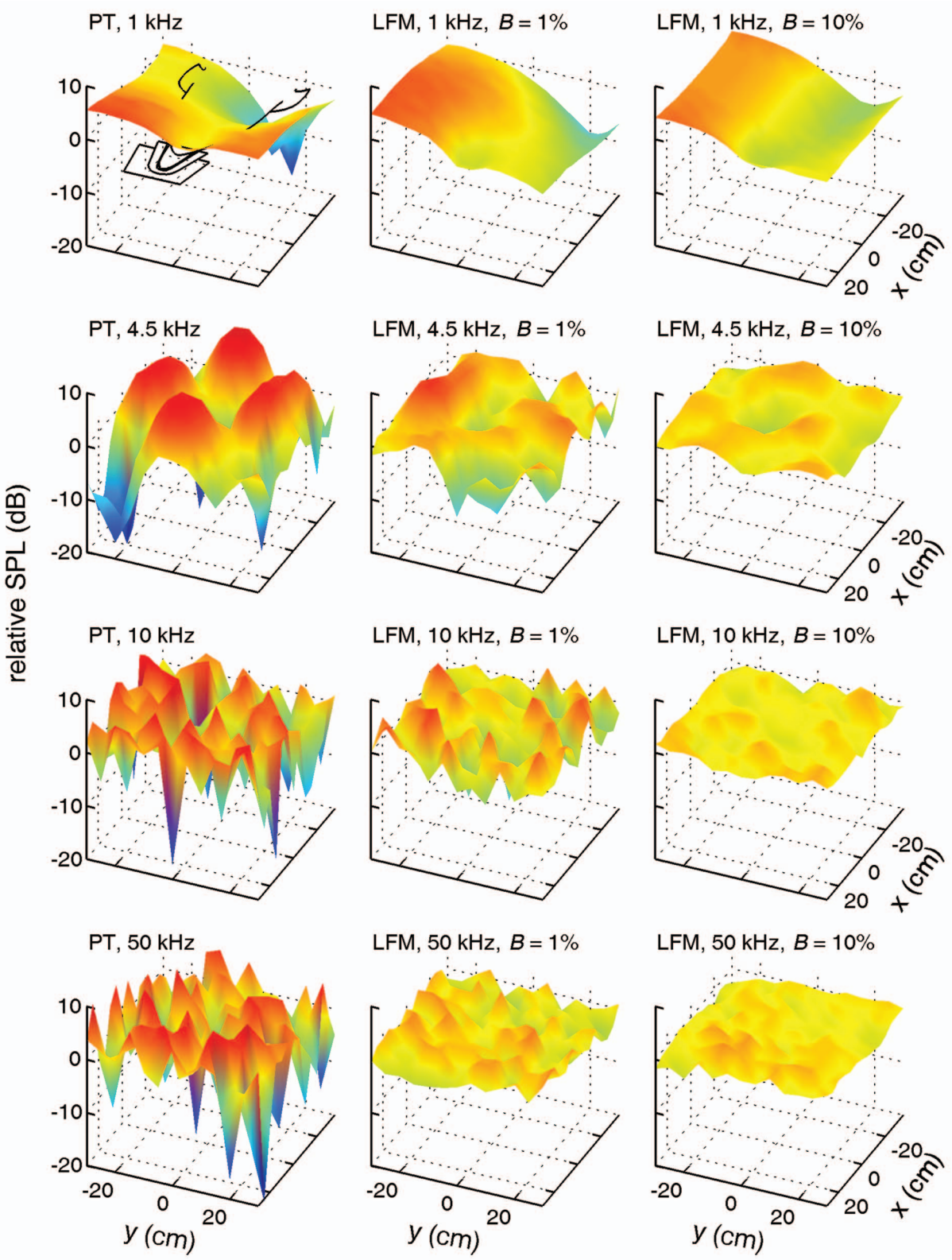


FIG. 4. (Color) Measured SPLs expressed as heights above the mean value for a subset of pure tone (PT) and linear frequency modulated (LFM) waveforms with bandwidths (B) of 1% and 10%. All data are from the middle water depth (z=0 cm). Waveform center frequencies are provided above each panel. The outline of the dolphin and biteplate are shown in the upper left panel for orientation purposes.

III. RESULTS

Figure 4 shows a representative subset of the 429 acoustic field maps showing the measured SPLs as vertical heights above the mean value (dB *re* mean SPL). Data are shown only for pure tones and linear FM tones (B=1% and 10%) at 1, 4.5, 10, and 50 kHz measured at the middle depth (z =0 cm). Acoustic field maps for SFM tones with identical

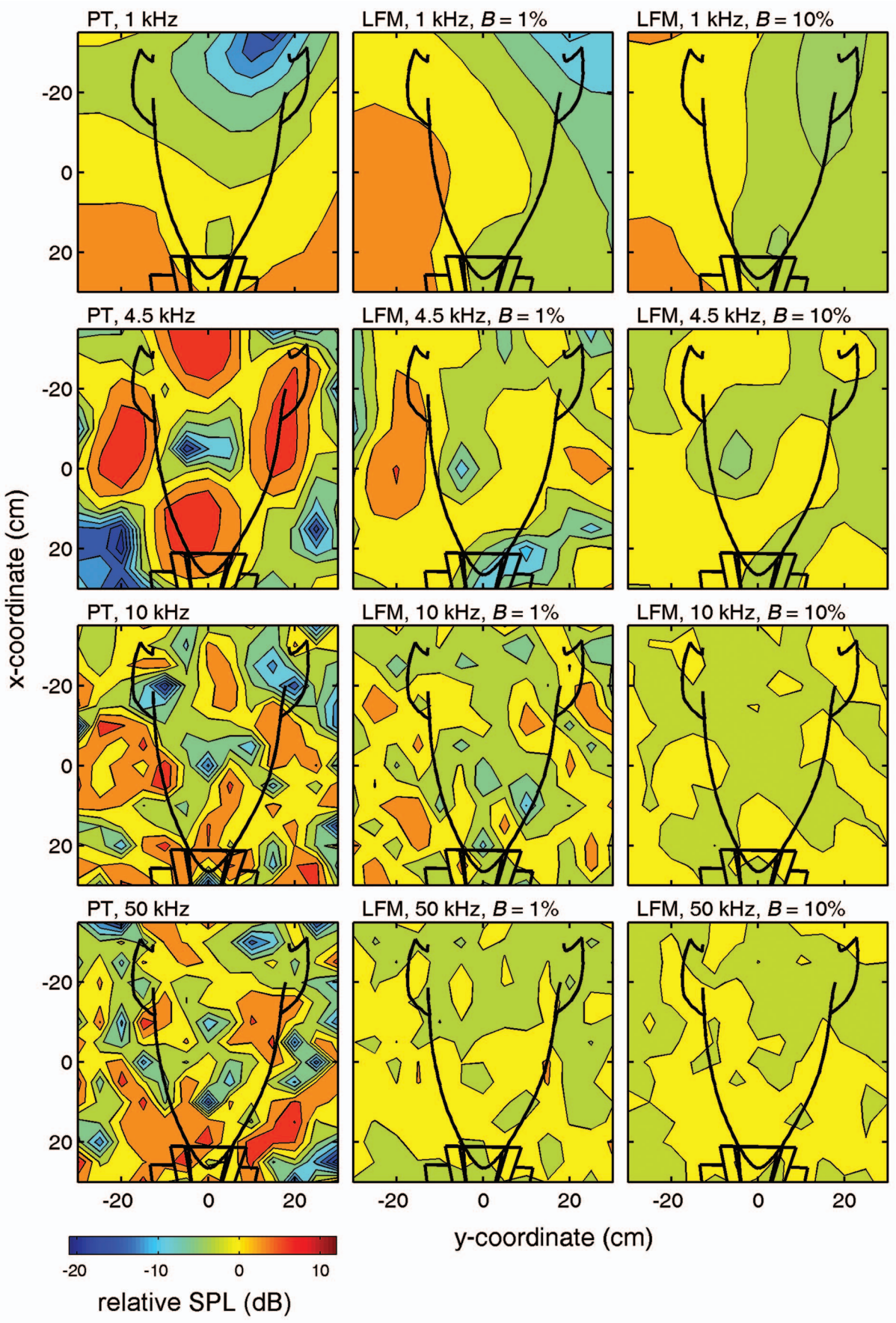


FIG. 5. (Color) Contour plots corresponding to the surface plots shown in Fig. 4. Levels represent deviations from the mean SPL for that frequency/waveform combination. Contour lines are separated by 3 dB. All data are from the middle water depth (z=0 cm).

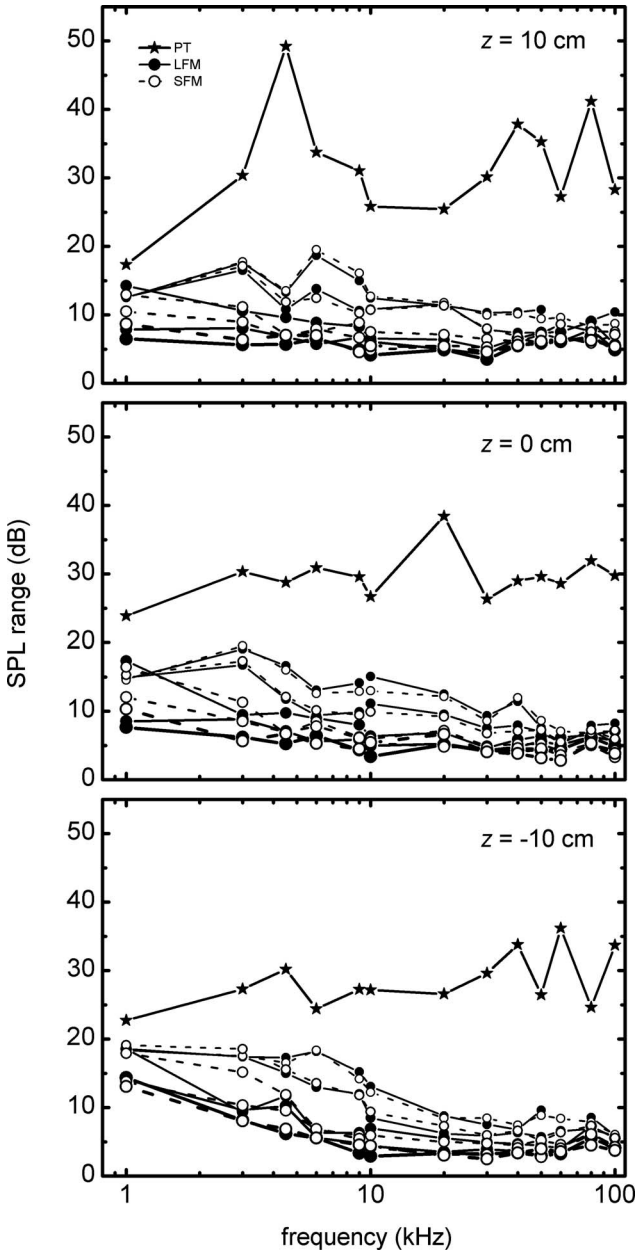


FIG. 6. SPL range, defined as the maximum SPL minus the minimum SPL, as a function of frequency for water depths of $z=10$ cm (upper), 0 cm (middle), and -10 cm (lower). Each curve shows the data for a different waveform: PT=stars, LFM=filled circles, SFM=open circles. The size of the circle represents the bandwidth (LFM) or modulation depth (SFM) and equaled 1%, 2%, 5%, 10%, and 20%.

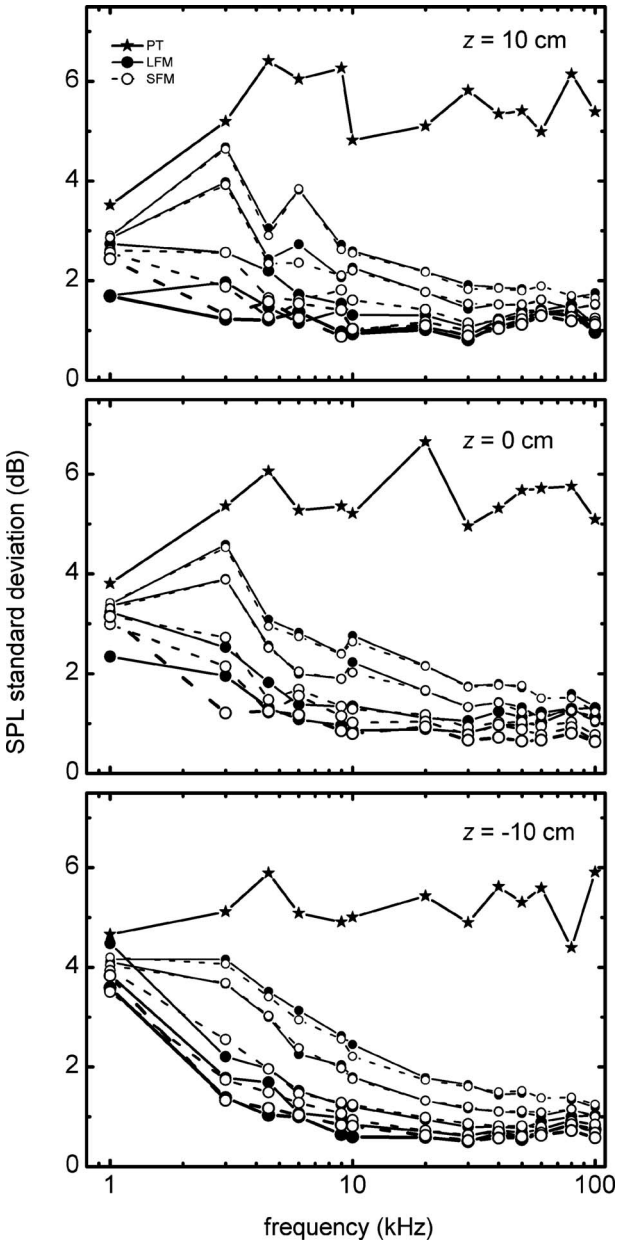


FIG. 7. SPL standard deviation as a function of frequency for water depths of $z=10$ cm (upper), 0 cm (middle), and -10 cm (lower). Each curve shows the data for a different waveform: PT=stars, LFM=filled circles, SFM =open circles. The size of the circle represents the bandwidth (LFM) or modulation depth (SFM) and equaled 1%, 2%, 5%, 10%, and 20%.

frequencies and comparable bandwidths exhibited similar characteristics, as did the measurements at the other two water depths. These same data are shown as contour plots in Fig. 5, with the dolphin and biteplate outline included in each plot for a spatial reference.

Figures 6 and 7 show the SPL range and standard deviation, respectively, for each signal waveform as a function of frequency. Data from the three water depths are shown in separate panels.

Hearing thresholds (upper panel) and false alarm rates (lower panel) are shown in Fig. 8. An example of the ambient noise spectral density measured in the pool is also included.

IV. DISCUSSION

Sound fields produced with PT signals exhibited pressure minim/maximum, as expected, with differences between minima and maxima almost always greater than 20 dB and at times more than 40 dB. Distances between adjacent pressure minima/maxima were related to the sound wavelength, so at high frequencies sound pressures fluctuated considerably over relatively small spatial distances. Because of the hydrophone dimensions (2.2 cm diameter) and 5-cm grid spacing, some existing variability in sound pressure may not have been captured in the measurements at high frequencies where the sound wavelength approached these dimensions.

Sound fields measured with FM signals exhibited similar features as the PT sound fields, except the pressure minima

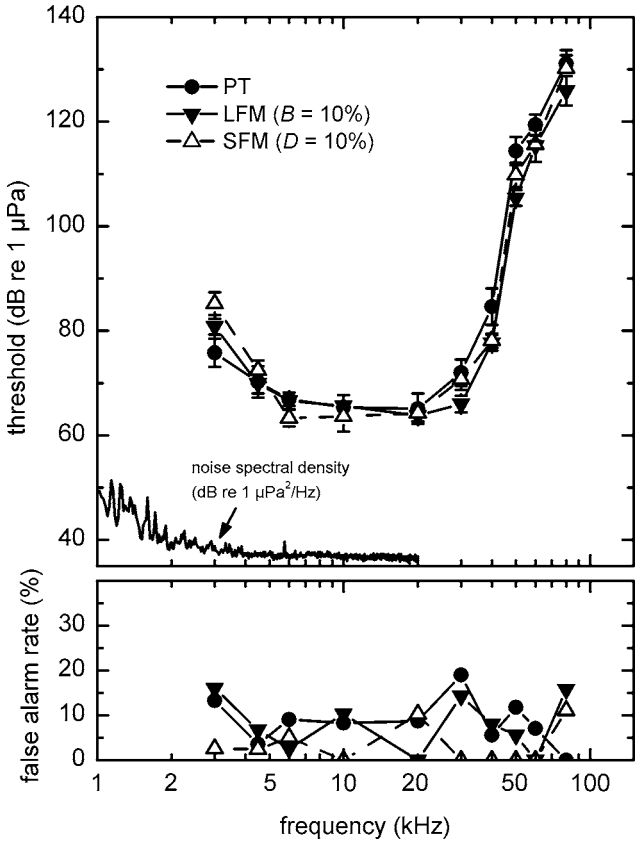


FIG. 8. Hearing thresholds (upper) and false alarm rates (lower) for BLU obtained behaviorally with PT, LFM, and SFM tones. Data shown are mean values ± SD over a total 18–30 reversals obtained on three or more sessions (six reversals each session). Ambient noise spectral density is referred to the same *y*-axis scale, but has units of dB *re* 1 μPa^2/Hz.

and maxima were reduced in amplitude (always $<$20 dB). This resulted in lower SPL ranges and SDs, making the FM sound fields more uniform than the PT sound fields. The improvement in the sound field variability was related to the FM bandwidth/modulation depth and the center frequency. Sound fields were more uniform with larger LFM bandwidth and SFM modulation depth; however, bandwidths/depths as low as 1% produced a noticeable improvement at all but the lowest frequency. An area of diminishing returns existed for B or $D \geq 5\%$, with little additional improvement in sound fields with B and D of 10% and 20%. There was a small improvement in the SPL range and SD with FM signals at 1 kHz, where the measurement grid dimensions were only a fraction of the acoustic wavelength (1.5 m). At higher frequencies the improvement with FM signals was more dramatic.

Sound fields obtained with the LFM and SFM signals were similar, especially for those conditions tested on the same day. LFM and SFM signal bandwidths were approximately equal for $f_c > 10$ kHz, where $2f_m \ll Df_c$, so the correspondence in sound fields at the higher frequencies was to be expected.

Mean hearing thresholds measured behaviorally with the PT, LFM, and SFM stimuli were similar across the range of frequencies tested. All three audiograms revealed significant high frequency hearing loss above 40 kHz, similar to that reported in prior behavioral and evoked potential hearing tests of BLU (Finneran and Houser, 2006). Thresholds were more than 20 dB above the ambient noise spectral density at all frequencies. Differences between PT and FM thresholds were generally $<$6 dB, with the largest differences (7 and 9 dB) occurring at 40 and 50 kHz where the audiogram slope was very steep. Differences between FM thresholds were less than 5 dB at all frequencies. At 3 kHz, both SFM and LFM thresholds were higher than the PT thresholds, possibly indicating FM bandwidths exceeding the auditory filter at 3 kHz, resulting in a lower detectability compared to the PT stimulus. False alarm rates were similar, except for the SFM stimuli, which typically resulted in lower false alarm rates. Although SFM thresholds were generally higher than LFM thresholds at the same frequency, there was no obvious relationship between SFM thresholds and the false alarm rates that would indicate a consistent shift in response bias.

Threshold SDs were all less than 3.5 dB. PT threshold SDs were usually larger than those of the FM thresholds, though the differences were relatively small (1–2 dB). This result is somewhat surprising considering the large variations in the PT SPLs compared to the FM SPLs. To some extent this validates the two-hydrophone technique used to calibrate the SPLs—despite large spatial variations in SPL, the calibration procedure yielded thresholds exhibiting small day-to-day variability. The subject's body also likely affected the sound pressures in the pool and may have reduced the SPL variability in the vicinity of the subject's head.

The similarity between PT and FM thresholds also raises the question of whether the use of FM signals is truly necessary. Although there were no obvious advantages to FM stimuli in terms of reducing the variability in measured thresholds, the calibration of the FM stimuli can be considerably easier because the need to use multiple hydrophones and manipulate the measurement locations to avoid pressure minima/maxima problems is eliminated. The use of FM stimuli also provides a much greater margin of error for subject placement and movement since the sound pressures are more uniform with spatial position and not as dependent on calibrating at the exact subject position. Of course, for some experimental questions the use of PT would be required; however, in circumstances where some finite stimulus bandwidth is acceptable, the use of FM stimuli offers significant advantages over pure tones for testing in small, enclosed volumes.

V. CONCLUSIONS

Acoustic fields produced by narrow-band and wideband FM signals exhibited less spatial variability compared to pure tone sound fields, especially for frequencies above 4 kHz. For many applications in small pools, FM signals may be good alternatives to pure tones because the sound fields are more uniform.

ACKNOWLEDGMENTS

We thank Randall Dear for assisting with the acoustic field measurements and Linda Green and Laura Lewis for animal training. Amanda Glazier also assisted with threshold

data collection. Financial support was provided by the U.S. Office of Naval Research, Marine Mammal Science and Technology Program.

Albers, V. M. (**1965**). *Underwater Acoustics Handbook* (Pennsylvania U. P., University Park, PA).

Bobber, R. J. (**1970**). *Underwater Electroacoustic Measurements* (Naval Research Laboratory, Washington, DC).

Bobber, R. J., and Beatty, L. G. (**1959**). "Impedance tube for underwater sound transducer evaluation," J. Acoust. Soc. Am. **31**, 832–833(A).

Carson, J. R. (**1963**). "Notes on the theory of modulation," Proc. IEEE **51**, 893–896.

Cornsweet, T. N. (**1962**). "The staircase method in psychophysics," Am. J. Psychol. **75**, 485–491.

Finneran, J. J. (**2003**). "An integrated computer-controlled system for marine mammal auditory testing," SSC San Diego TD 3159 (SSC San Diego).

Finneran, J. J., Carder, D. A., Schlundt, C. E., and Ridgway, S. H. (**2005**). "Temporary threshold shift (TTS) in bottlenose dolphins (*Tursiops truncatus*) exposed to mid-frequency tones," J. Acoust. Soc. Am. **118**, 2696–2705.

Finneran, J. J., and Houser, D. S. (**2006**). "Comparison of in-air evoked potential and underwater behavioral hearing thresholds in four bottlenose dolphins (*Tursiops truncatus*)," J. Acoust. Soc. Am. **119**, 3181–3192.

Gellerman, L. W. (**1933**). "Chance orders of alternating stimuli in visual discrimination experiments," J. Gen. Psychol. **42**, 206–208.

Gerrits, J. F. M., Kouwenhoven, M. H. L., van der Meer, P. R., Farserotu, J. R., and Long, J. R. (**2005**). "Principles and limitations of ultra-wideband FM communications systems," EURASIP J. Appl. Signal Process. **3**, 382–396.

Kastelein, R. A., Bunskoek, P., and Hagedoorn, M. (**2002a**). "Audiogram of a harbor porpoise (*Phocoena phocoena*) measured with narrow-band frequency-modulated signals," J. Acoust. Soc. Am. **112**, 334–344.

Kastelein, R. A., Hagedoorn, M., and de Haan, D. (**2003**). "Audiogram of a striped dolphin (*Stenella coeruleoalba*)," J. Acoust. Soc. Am. **113**, 1130–1137.

Kastelein, R. A., Mosterd, P., van Santen, B., and Hagedoorn, M. (**2002b**). "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.

Miller, L. (**2006**). "Recollections of Paul Sabine and the Harvard University Underwater Sound Lab," Acoust. Today **2**, 44–45.

Parvulescu, A. (**1964**). "Problems of propagation and processing," in *Marine Bioacoustics*, Vol. **1**, edited by W. N. Tavolga (Pergamon, Oxford).

Parvulescu, A. (**1966**). "The acoustics of small tanks," in *Marine Bioacoustics*, Vol. **2**, edited by W. N. Tavolga (Pergamon, Oxford), pp. 7–13.

Schlundt, C. E., Dear, R. L., Carder, D. A., and Finneran, J. J. (**2006**). "Growth and recovery of temporary threshold shifts in a dolphin exposed to mid-frequency tones with durations up to 128 s," J. Acoust. Soc. Am. **120**, 3227(A).

Schlundt, C. E., Finneran, J. J., Carder, D. A., and Ridgway, S. H. (**2000**). "Temporary shift in masked hearing thresholds of bottlenose dolphins, *Tursiops truncatus*, and white whales, *Delphinapterus leucas*, after exposure to intense tones," J. Acoust. Soc. Am. **107**, 3496–3508.

Simultaneously measured behavioral and electrophysiological hearing thresholds in a bottlenose dolphin (*Tursiops truncatus*)

Carolyn E. Schlundt[a)]
EDO Professional Services, 3276 Rosecrans Street, San Diego, California 92110

Randall L. Dear and Linda Green
Science Applications International Corporation, 4065 Hancock Street, San Diego, California 92110

Dorian S. Houser
Biomimetica, 7951 Shantung Drive, Santee, California 92071

James J. Finneran
U.S. Navy Marine Mammal Program, Space and Naval Warfare Systems Center, San Diego, Code 2351, 53560 Hull Street, San Diego, California 92152

(Received 29 December 2006; revised 17 April 2007; accepted 18 April 2007)

Dolphin auditory thresholds obtained via evoked potential audiometry may deviate from behavioral estimates by 20 dB or more. Differences in the sound source, stimulus presentation method, wave form, and duration may partially explain these discrepancies. To determine the agreement between behavioral and auditory evoked potential (AEP) threshold estimates when these parameters are held constant, behavioral and AEP hearing tests were simultaneously conducted in a bottlenose dolphin. Measurements were made in-air, using sinusoidal amplitude-modulated tones continuously projected via a transducer coupled to the pan region of the dolphin's lower jaw. Tone trials were presented using the method of constant stimuli. Behavioral thresholds were estimated using a 50% correct detection. AEP thresholds were based on the envelope following response and 50% correct detection. Differences between AEP and behavioral thresholds were within ±5 dB, except at 10 kHz (12 dB), 20 kHz (8 dB), 30 kHz (7 dB), and 150 kHz (24 dB). In general, behavioral thresholds were slightly lower, though this trend was not significant. The results demonstrate that when the test environment, sound source, stimulus wave form, duration, presentation method, and analysis are consistent, the magnitude of the differences between AEP and behavioral thresholds is substantially reduced. [DOI: 10.1121/1.2737982]

PACS number(s): 43.80.Lb [WWA] Pages: 615–622

I. INTRODUCTION

Behaviorally based audiometric data exist for approximately 14% of odontocete species [the harbor porpoise, *Phocoena phocoena* (Andersen, 1970; Kastelein *et al.*, 2002), killer whale, *Orcinus orca* (Hall and Johnson, 1972; Szymanski *et al.*, 1999), Amazon River dolphin, *Inia geoffrensis* (Jacobs and Hall, 1972), beluga, *Delphinapterus leucas* (White *et al.*, 1978; Awbrey *et al.*, 1988; Ridgway *et al.*, 2001; Finneran *et al.*, 2005), Eastern Pacific bottlenose dolphin, *Tursiops spp.* (Ljungblad *et al.*, 1982), false killer whale, *Pseudorca crassidens* (Thomas *et al.*, 1988; Yuen *et al.*, 2005), Chinese River dolphin, *Lipotes vexillifer* (Wang *et al.*, 1992), Risso's dolphin, *Grampus griseus* (Nachtigall *et al.*, 1995), tucuxi, *Sotalia fluviatilis guianensis* (Sauerland and Dehnhardt, 1998), and the striped dolphin, *Stenella coeruleoalba* (Kastelein *et al.*, 2003)]. In each of these studies the number of subjects was limited to one or several animals because of the effort and time required to train the behaviors for testing and to perform the data collection. As a result, variability in the hearing sensitivity of most odontocete species is largely undetermined.

An alternative method for rapidly assessing hearing sensitivity in small odontocetes is the measurement of auditory evoked potentials (AEPs). AEP methods are attractive because hearing thresholds can be measured quickly (minutes to hours) with minimal training on the part of the animal. AEP-based audiometric data exist for 12 species of odontocetes, including the bottlenose dolphin (Ridgway *et al.*, 1981; Popov and Supin, 1990b; Finneran and Houser, 2006; Houser and Finneran, 2006b), beluga (Popov and Supin, 1987; Dolphin, 1997b), beaked whale, *Mesoplodon europaeus* (Cook *et al.*, 2006), Amazon River dolphin (Popov and Supin, 1990c), false killer whale (Dolphin *et al.*, 1995; Supin *et al.*, 2003; Yuen *et al.*, 2005), common dolphin, *Delphinus delphis* (Popov and Klishin, 1998), harbor porpoise (Popov and Supin, 1986), killer whale (Szymanski *et al.*, 1999), tucuxi (Popov and Supin, 1990a), striped dolphin (Andre *et al.*, 2003), Risso's dolphin (Dolphin, 1997a; Nachtigall *et al.*, 2005), and the finless porpoise, *Neophocaena phocaenoides asiaeorientalis* (Popov *et al.*, 2005). The technique is ideal for increasing the sample size within a species for which hearing data already exist, and for appli-

[a)]Electronic mail: carolyn.melka@edocorp.com

cation to stranded and rehabilitating marine mammals not likely to be kept in captivity.

Several recent efforts have focused on quantifying differences between behavioral and AEP thresholds within the same subjects in order to benchmark the electrophysiological measurements for use as a predictor of behavioral sensitivity (Szymanski *et al.*, 1999; Yuen *et al.*, 2005; Finneran and Houser, 2006; Houser and Finneran, 2006a). In these cases, behavioral and AEP data from the same individual have yielded average deviations of 11 dB and as large as 41 dB. It is generally acknowledged that differences in stimulus wave form used in the behavioral and AEP techniques influence the results, with the relatively short duration stimuli used for AEP audiometry leading to higher thresholds than those behaviorally obtained with longer duration stimuli (Szymanski *et al.*, 1999; Yuen *et al.*, 2005). A review of the experimental parameters used to assess hearing in each of the behavioral-AEP comparison studies reveals that differences were not only limited to stimulus duration but also included differences in the sound source (e.g., free-field or direct field versus jawphone), the stimulus wave form [e.g., pure tone versus sinusoidal amplitude modulated (SAM) tone burst], the stimulus presentation method (e.g., staircase versus descending limits), the test environment (e.g., in-water versus in-air), and even in the analysis used to estimate threshold [e.g., average of reversals versus regression of the envelope following response (EFR) amplitude versus stimulus level], all of which could contribute to the differences between thresholds observed in these studies. Even when differences between experimental parameters are minimized, for example Yuen *et al.* (2005) who used the same test environment and sound source in their comparison study, threshold differences exist.

The objective of the present study was to examine the extent to which eliminating differences in methodologies might reduce differences between behavioral and AEP thresholds. This goal was accomplished by comparing hearing thresholds simultaneously measured using behavioral and AEP methods. Behavioral and AEP thresholds were measured using the same experimental parameters, including the sound source (jawphone), stimulus wave form (SAM tone), stimulus duration (12 s), stimulus presentation method (method of constant stimuli), test environment (in-air), and analysis (50% correct detection). This is the first study to use a SAM tone and an in-air test environment for both behavioral and AEP measurements. Using this approach, thresholds were expected to show closer agreement than in previous comparisons.

II. METHODS

A. Subject

The subject (WEN) was a male bottlenose dolphin (*Tursiops truncatus*), 21-years and approximately 210 kg. The subject had previously participated in cooperative psychophysical tasks, including auditory detection tasks and AEP measurements. He was housed in floating netted enclosures, 9×9 to 12×24 m, located in San Diego Bay, CA. The study followed a protocol approved by the Institutional Animal Care and Use Committee of the Biosciences Division, Space and Naval Warfare Systems Center, San Diego, and followed all applicable U.S. Department of Defense guidelines for the care of laboratory animals.

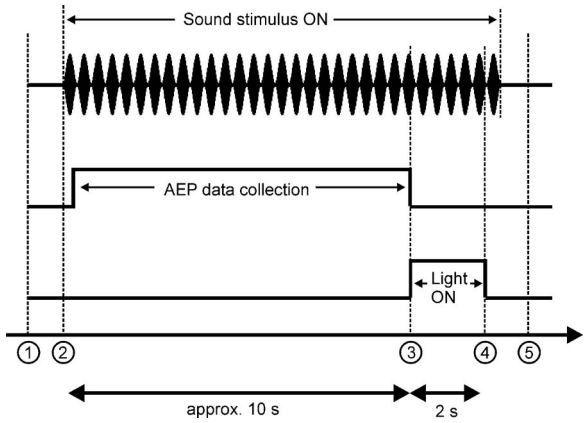


FIG. 1. Relationship between sound stimulus, AEP data collection interval, and behavioral response interval. (1) Each trial was started with the trainer touching the subject's rostrum. (2) The sound stimulus then began, followed by AEP data collection. The AEP data collection ignored the first epoch to avoid any potential transient response. (3) At the conclusion of the AEP data collection, the light was illuminated signifying the time during which the subject reported the presence or absence of the stimulus. (4) After 2 s, the light was turned off. The sound stimulus was turned off after the light. (5) At the end of the trial, the subject was rewarded for correct responses; no reward was offered for incorrect responses.

B. Approach

The approach to simultaneously measure AEP and behavioral thresholds centered on training the subject to provide an unambiguous behavioral response (a whistle) when it heard the sound stimulus and to remain quiet otherwise. To prevent the subject's behavioral response from producing artifacts and contaminating the concurrent AEP data collection, a light was used to delineate a trial period for the subject's behavioral response. The light was illuminated after the AEP recording was complete but while the sound stimulus was still presented to the subject. The subject was trained to whistle only if it heard the sound stimulus while the light was illuminated and to otherwise remain quiet. An overview of the experimental test sequence is shown in Fig. 1.

Measurements were conducted in air, with the subject positioned on a foam mat. Sound stimuli were coupled to the lower jaw using a "jawphone"—a piezoelectric sound projector embedded in a silicon rubber suction cup (Moore *et al.*, 1995; Brill *et al.*, 2001; Finneran and Houser, 2006). In-air measurements were chosen for this study because (1) AEP amplitudes are typically larger in-air compared to underwater, and (2) in-air testing allowed the effects of ambient noise in San Diego Bay to be avoided. Measurements of the rms sound pressure level (SPL) underwater at a distance of 15 cm from the suction cup face were used to estimate received levels experienced by the subject (see Finneran and Houser, 2006). Comparison measurements in five dolphins, including WEN, have shown that in-air AEP thresholds obtained with jawphones calibrated in this manner are comparable to underwater hearing thresholds (Finneran and Houser,

2006; Houser and Finneran, 2006a). Since the primary goal of this study was to compare relative hearing thresholds obtained with AEP and behavioral techniques under identical stimulation, the specifics of the jawphone calibration technique, and the uncertainties therein, do not influence the major findings. The jawphone was placed on the subject's lower left jaw, in the pan region, for all measurements. Previous research has shown this site to be important for the reception of high frequency sounds in dolphins (Bullock *et al.*, 1968; McCormick *et al.*, 1970; 1980; Brill *et al.*, 1988; Møhl *et al.*, 1999). Eleven frequencies were tested, from 10 to 150 kHz. Two jawphones were used during the study: one (JP1) for frequencies greater than 100 kHz and another (JP4) for frequencies from 10 to 100 kHz (Finneran and Houser, 2006).

Tone trials were presented using the method of constant stimuli. Either one or two 30-min sessions were conducted per day, with a single frequency tested per session. Sessions were divided into six blocks of ten trials each, with five trials being stimulus-present and five trials being stimulus-absent. The order of stimulus-present/stimulus-absent trials was randomized using a Gellerman configuration (Gellerman, 1933). The SPL was held constant during each trial block. The six SPLs tested for each frequency were selected in advance, based on previous behavioral and electrophysiological hearing studies with the same subject (Finneran and Houser, 2006). An attempt was made to include SPLs that would result in probabilities of detection spanning the range of 0–100%. Each session resulted in 60 trials; 30 stimulus-present and 30 stimulus-absent trials at six different SPLs (five stimulus-present and five stimulus-absent at each SPL). Each frequency was tested over six different sessions, so for each frequency/SPL combination a total of 30 stimulus-present and 30 stimulus-absent trials were obtained.

During each session, a trainer was seated immediately in front of the dolphin and indicated the start of a trial by placing a finger on top of the dolphin's rostrum. Once the trainer had given this cue, the computer operator began the trial by initiating the sound stimulus and AEP data collection. The trainer was blind to the trial conditions (SPL; present/absent), though the highest SPL stimulus at 10 kHz was usually audible. After AEP data collection, with the stimulus still "on," the light was illuminated for 2 s to delineate the behavioral response interval. The light was positioned so that its beam would illuminate the left side of the subject's rostrum (i.e., not pointing directly into the eyes, yet clearly visible). After the light turned off (signaling the end of the response interval), the computer operator sounded a buzzer if the subject's behavioral response was correct (a hit or a correct rejection) and did nothing if the subject was incorrect (a miss or a false alarm). The sound stimulus remained on until after the light was off. The subject's whistle responses were clearly audible and unambiguous. For correct responses the trainer removed his/her finger from the dolphin's rostrum and fed the subject two fish (capelin). If the behavioral response was incorrect, the trainer's finger remained on the dolphin's rostrum, no fish were offered, and the computer operator began the next trial. A short break occurred after each block of ten trials while the computer operator prepared for the next block and the subject was rinsed with water to keep cool. This sequence continued until all six stimulus SPLs at the test frequency were presented (i.e., 60 trials).

Finally, up to eight additional trials were presented, still in keeping with the Gellerman configuration, so that three to four additional stimulus-present trials were presented to the subject. The SPLs of these stimuli were selected to ensure sufficient data for post-hoc evoked potential threshold estimates based on a linear regression utilizing data points contained within the AEP input-output function (see Supin *et al.*, 2001). This allowed a comparison of behavioral and AEP thresholds when just one experimental parameter, the analysis, was varied. These additional data were excluded from the derivation of the behavioral and AEP percent correct versus SPL functions. At the conclusion of the session, equipment (see the following) was removed and the subject returned to the water.

C. Measurement system

Stimulus generation and AEP recording equipment are described in detail in Finneran and Houser (2006). Stimuli were digitally generated and output at 12-bit resolution and 500-kHz sampling rate using a personal computer (PC) with a multifunction data acquisition board (National Instruments PCI-MIO-16E-1). Stimulus levels were controlled by a Programmable Attenuator (Tucker Davis PA5). All stimuli were sinusoidal amplitude modulated tones with a 1-kHz modulation rate. Tones were continuously presented to the subject, with no temporal gaps (total duration was about 12 s).

AEPs were measured using 10-mm gold cup surface electrodes (e.g., Grass FH-E6G series) embedded in 25-mm-diam silicon rubber suction cups. The noninverting (+) electrode was placed approximately 10 cm posterior of the blowhole and offset approximately 2 cm to the right side (i.e., contralateral to the left ear being tested). The inverting (−) electrode was placed just posterior to the right external auditory meatus. A ground (common) electrode was placed on the subject's back just anterior of the dorsal fin. Conductive paste was applied to each electrode before attachment. A differential amplifier (Grass P511) was used to amplify ($\times 10^5$) and filter (0.3–3 kHz) the voltage difference between the noninverting and inverting electrodes. The amplified signal was digitized at 10-kHz sampling rate with a 12-bit resolution and stored on the PC. AEP data were divided into 19-ms epochs for analysis. To avoid artifacts, epochs with peak instantaneous voltage greater than 8 μV were rejected from further analysis. This reject level was below that observed to occur when breathing or whole animal movement took place, thus ensuring that myogenic artifacts related to these events were not recorded. AEP data collection proceeded until 500 nonrejected epochs were acquired.

D. Analysis

For both the behavioral and AEP tasks, a binary result was obtained for each trial. For the behavioral test, each trial was scored by the computer operator as a response/no response. For each AEP measurement, magnitude-squared coherence (MSC) was calculated and used to objectively determine if the measured AEP component at the modulation

frequency was statistically different from background noise (Dobie and Wilson, 1989; Dobie, 1993; Dobie and Wilson, 1996). MSC relies on segment analysis, where the collection of recorded epochs is divided into a number of individual segments, each of which is then averaged (either synchronously in the time domain or coherently in the frequency domain) to yield a number of "subaverages." In this case, the MSC calculation used 20 subaverages, each based on 25 individual epochs. Critical values for MSC, using $\alpha=0.01$, were obtained from Amos and Koopmans (1963) and Brillinger (1978). If the calculated MSC was greater than the critical value, the AEP at the modulation frequency was considered to be detected.

For each test frequency and SPL combination, the percentage correct was calculated as the total number of responses to stimulus present trials at that frequency and SPL, divided by 30. A nonlinear regression was used to fit the percentage correct versus SPL data with a sigmoidal dose-response curve (a four parameter logistic equation):

$$P_c(\mathrm{SPL}) = B + \frac{T-B}{1+10^{H(L_{50}-\mathrm{SPL})}}, \quad (1)$$

where P_c is the percent correct and T, B, H, and L_{50} are fitting parameters (GraphPad Software, 2003). For this study, T ("top") and B ("bottom") were constrained to 100% and 0%, respectively, resulting in only two free parameters: L_{50}, the SPL corresponding to the 50% correct point, which was defined as the threshold for this study, and H, the "Hill slope" (GraphPad Software, 2003). The regression process also produced confidence intervals for the threshold (L_{50}).

False alarm rates were defined as the number of responses (a whistle during the behavioral data collection or a detected response during the AEP data collection) during all stimulus absent trials during a particular session, divided by 30. False alarm rates from the six sessions at each frequency were then averaged.

An additional, post-hoc analysis method was used to estimate evoked potential hearing thresholds for each of the frequencies tested. A linear regression technique was performed using a minimum of four consecutive detected responses beginning with the lowest stimulus SPL (Finneran and Houser, 2006). At SPLs with more than one detected response, the detection with the median response amplitude (or the *higher* median for SPLs with an even number of detections) was used for the regression. Detections occurring at SPLs below two nondetections were excluded from threshold calculations. If the regression r^2 was less than 0.9 using the first four data points, additional points obtained at consecutively higher stimulus levels were added to the regression analysis until the r^2 criterion of 0.9 was met or an obvious plateau was reached in the AEP amplitude data. The AEP regression threshold was defined as the stimulus SPL at which the extrapolated regression line reached 0 V. AEP regression thresholds estimated from the six sessions at each frequency were then averaged.

III. RESULTS

The percentage of correct responses as a function of stimulus SPL is shown in Fig. 2 for the behavioral and AEP tests. Agreement between the AEP and behavioral data sets was generally good within the midrange of frequencies tested (e.g., 40–115 kHz), but deteriorated toward the lowest and highest frequencies tested. The solid and dashed lines represent the curve fits as described by Eq. (1). Goodness of fit parameters for the curve fits are provided in Table I; 14/22 data sets had $R^2>0.99$, 21/22 had $R^2>0.96$.

Table II provides numerical values for the behavioral, AEP 50% correct, and AEP regression thresholds. Figure 3 shows the hearing thresholds based on the 50% correct detection point from data in Fig. 2 (upper panel) and false alarm rates (lower panel) for the behavioral and AEP tests. Though both audiograms resembled the U-shaped curve typically seen in mammalian hearing, each revealed a "notch" in WEN's hearing ability at 60–80 kHz. The most sensitive portion of the hearing curve was at 50 kHz for both the behavioral and AEP methods. Both audiograms showed a sharp upper frequency cut-off beyond 130 kHz. Thresholds increased by 25 dB between 130 and 150 kHz in behavioral testing, and by 44 dB in AEP testing. The upper cutoff frequency and the 60- to 80-kHz notch agree with behavioral hearing thresholds for WEN obtained in San Diego Bay under direct-field conditions and reported earlier (Finneran and Houser, 2006). The overall false alarm rate for the behavioral tests was 4.6%. The overall false alarm rate for the AEP response detection was 0.6%, reflecting the $\alpha=1\%$ criterion used to define the critical value for the MSC.

The differences (in dB) between the 50% correct AEP, regression AEP, and behavioral thresholds are shown in Fig. 4. Figure 4 also includes data from Yuen *et al.* (2005, false killer whale) and Szymanski *et al.* (1999, killer whale) for comparison. Overall, there was very close agreement between the behavioral and 50% correct AEP thresholds, which were within 5 dB of each other at 7/11 frequencies, and within 8 dB of each other at 9/11 frequencies. Behavioral thresholds were lower than the 50% correct AEP thresholds at 9 of the 11 frequencies tested, though these differences were >5 dB only at four frequencies (10, 20, 30, and 150 kHz). The largest differences occurred at the extreme ends of the audiogram; the behavioral thresholds were 12 and 24 dB lower than the 50% correct AEP thresholds at 10 and 150 kHz, respectively.

Behavioral and AEP regression thresholds were within 5 dB of each other at 3/11 frequencies, and within 10 dB of each other at 6/11 frequencies. The largest deviations were distributed throughout the audiogram; the behavioral thresholds were 9 and 21 dB lower than the AEP regression thresholds at 10 and 150 kHz, respectively, and the behavioral threshold was 16 dB higher than the AEP regression threshold at 60 kHz. Behavioral thresholds were higher than AEP regression thresholds at 8 of the 11 frequencies tested.

IV. DISCUSSION

The approach used in this study was an effective means of obtaining both AEP and behavioral data without signifi-

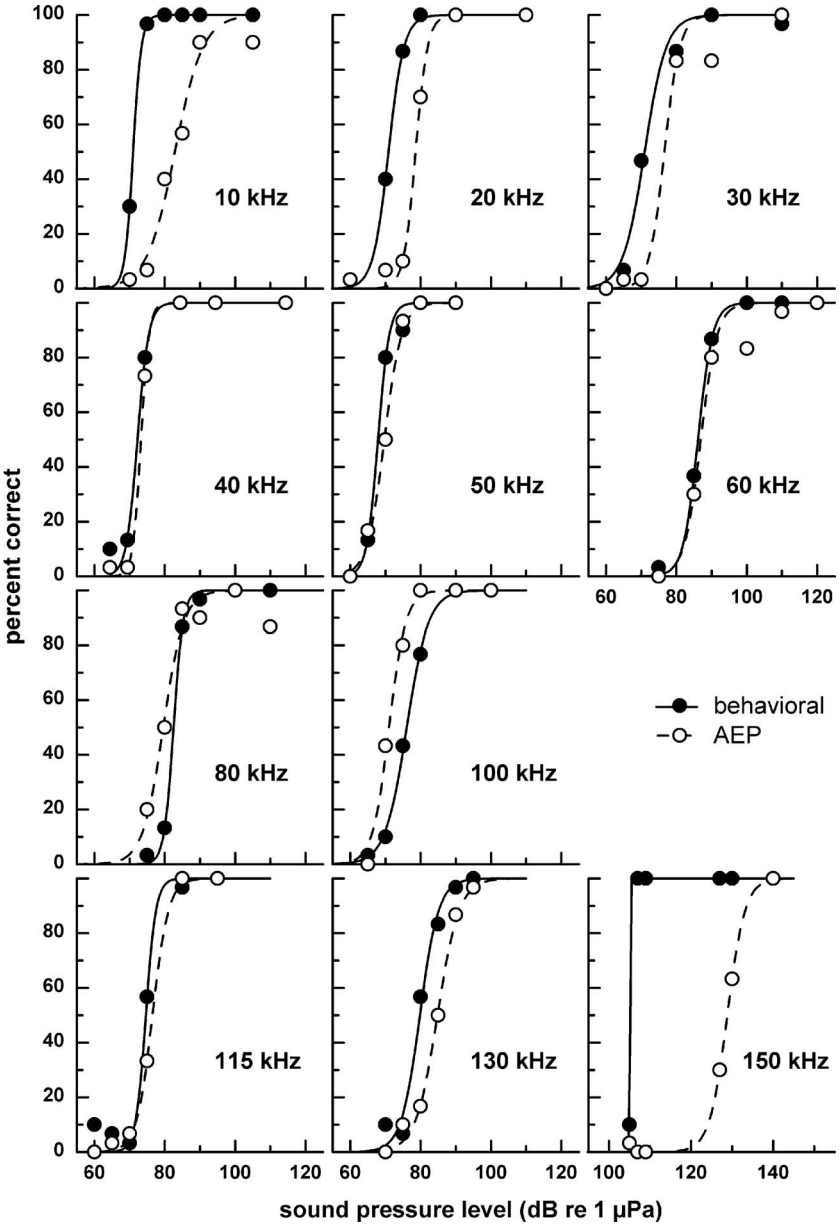


FIG. 2. Percentage of correct behavioral (solid circles, solid lines) and AEP (open circles, dotted lines) responses to stimulus-present trials for each stimulus level presented. Each panel shows data for a single frequency. The lines show the curve fit to the data. The stimulus level corresponding to the 50% correct detection value was defined as the threshold.

cantly extending the test time or allowing the subject's behavioral response to interfere with the AEP data collection. The use of the vocal response was important in this respect, since responses can be made rapidly by dolphins and do not require extensive physical movement, reducing the chance of disrupting the electrodes or jawphone placement. Furthermore, the distribution of false alarm rates as a function of frequency indicates that the animal had no response bias toward any particular test frequency and that the method suitably minimizes experimental artifacts on which the subject may capitalize.

TABLE I. Goodness of fit (R^2) for nonlinear fit of sigmoidal dose-response curve to experimental data.

Frequency (kHz)	R^2, behavioral	R^2, AEP
10	1.000	0.968
20	0.999	0.996
30	0.986	0.975
40	0.991	0.999
50	0.992	0.996
60	0.999	0.966
80	0.998	0.940
100	0.999	0.989
115	0.985	0.999
130	0.982	0.995
150	1.000	0.999

Differences previously observed between AEP and behavioral thresholds are attributed in part to differences in the stimulus wave form or duration. This study lends support to this conclusion: When the sound source, test environment, stimulus wave form, stimulus duration, stimulus presentation method, and analysis were held constant and measurements made simultaneously, behavioral and the 50% correct AEP data were very close (average difference of 6 dB). This agreement is much better than that obtained in the present study using the regression technique and that previously reported when experimental parameters between the AEP and

TABLE II. Comparison of thresholds and false alarm rates (FAR) during simultaneous behavioral and AEP hearing tests for dolphin WEN at 11 frequencies. Testing was conducted in-air. Behavioral thresholds were based on the 50% correct point on the psychometric function.

Frequency (kHz)	Behavioral threshold (dB re 1 μPa)	Behavioral FAR (%)	AEP threshold, 50% correct (dB re 1 μPa)	AEP FAR (%)	AEP threshold, regression (dB re 1 μPa)
10	71.0	8.3	82.9	1.1	80.4
20	70.9	2.2	78.6	0.0	72.6
30	70.0	8.9	76.6	0.6	65.6
40	72.2	7.2	73.2	1.1	63.3
50	67.9	7.2	69.7	0.0	55.1
60	86.1	1.7	86.9	0.6	69.7
80	82.5	3.9	79.6	0.6	71.9
100	76.1	1.7	71.0	0.6	64.7
115	74.6	5.0	76.5	1.7	68.1
130	79.8	0.6	84.8	0.6	74.5
150	105.2	3.9	128.8	0.0	126.3

behavioral test conditions varied (Finneran and Houser, 2006; Houser and Finneran, 2006a). AEP regression thresholds were expected to be less than the 50% correct AEP thresholds, though it was somewhat surprising that the regression thresholds were much lower than the behavioral thresholds at many frequencies. Nevertheless, behavioral thresholds in prior AEP and behavioral comparisons were obtained via modified staircase procedures (Finneran and Houser, 2006; Houser and Finneran, 2006a). Given the stimulus used in this study, combined with a modified staircase presentation procedure, we might expect the behavioral threshold to be lower and possibly closer to the AEP thresholds calculated via the regression technique. Certainly another of the major sources of discrepancy in previous studies was the use of underwater direct field stimulation (in San Diego Bay) for behavioral measurements and in-air jawphones for AEP tests. The agreement in the present study is also better than previously reported by Yuen *et al.* (2005), who used SAM tones and pure tones for the AEP and behavioral tests, respectively, and Szymanski *et al.* (1999), who used tone pips for AEP measurements and relatively long duration pure tones for behavioral thresholds. In both of these studies, the stimulus presentation was made in the direct field with the animal underwater, thus minimizing some of the methodological differences between the AEP and behavioral approaches. The average difference between AEP and behavioral data reported by Yuen *et al.* (2005, false killer whale) was 10 dB with a maximum difference of 15 dB. The average difference measured by Szymanski *et al.* (1999, killer whale) was 9 dB with a maximum difference of 41 dB at 100 kHz.

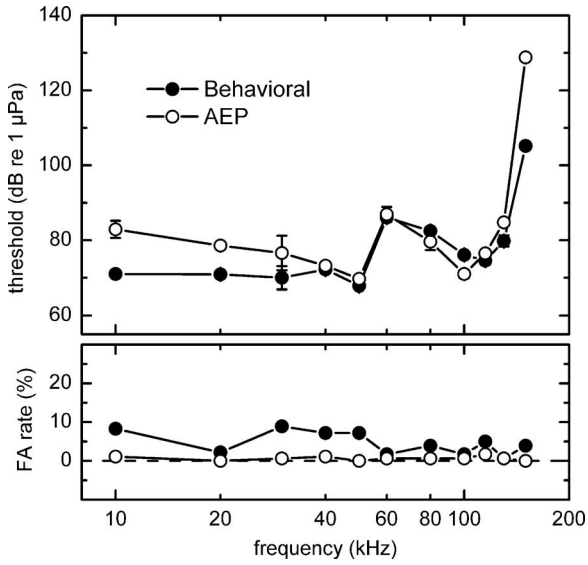


FIG. 3. Comparison of behavioral (closed circles) and AEP (open circles) thresholds collected simultaneously and when all experimental parameters were held constant for dolphin WEN at 11 frequencies. Testing was conducted in-air. Test tones were presented using the method of constant stimuli and were continuous, 9.5 s SAM tones based on 500 averages and 19-ms epochs. Stimuli were projected via a jawphone. Symbols represent 50% correct detection values; error bars represent the 95% confidence interval. False alarm rates for WEN during behavioral (closed triangles) and AEP (open triangles) testing.

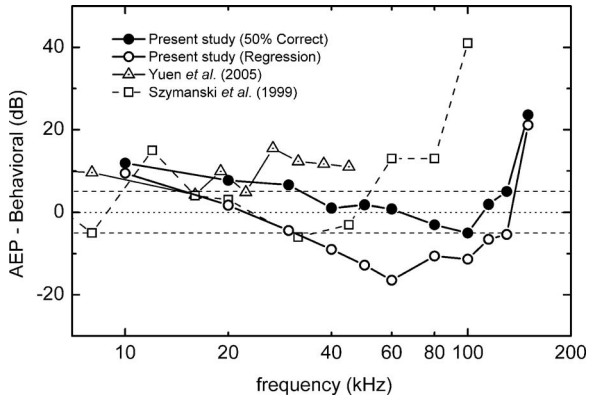


FIG. 4. Differences between AEP and behavioral thresholds. Data from Yuen *et al.* (2005) and Szymanski *et al.* (1999) are included for comparison. The dashed lines are at ±5 dB.

Odontocete evoked potential audiometry often uses the regression of EFR amplitudes to 0 V or a lowest detectable hit to estimate AEP thresholds. The design of the current study necessitated an analysis technique that could be applied to both behavioral and AEP measurements. The lowest-detectable level could have been applied if the stimuli had been presented using the method of limits. The behavioral disadvantages of using this method include anticipation by the subject of the next stimulus level and/or habituation of the subject to making a certain response. There is also the problem of determining the optimum step size between stimulus levels (Gelfand, 1990). Alternatively, the stimuli could have been presented using a modified up-down staircase procedure and threshold estimation could have been based on a number of reversals; however, the process would be biased as the staircase would have to follow the hits and misses of either the behavioral responses or the AEP responses. In either case, the number of reversals would not likely be the same. To address these issues, the present study used the method of constant stimuli, and consequently the 50% correct detection analysis, so that stimuli could be presented randomly and the response bias could be tracked.

The largest differences between AEP and behavioral thresholds in the present study were obtained at the lowest and highest frequencies tested, where AEP thresholds exceeded the behavioral values by 12–24 dB. At the lower frequencies, the differences increased as the frequency decreased, from 7 dB at 30 kHz, to 8 dB at 20 kHz, to 12 dB at 10 kHz. This result suggests that at the lower frequencies the discrepancy is due to the nature of the EFR response itself. Lower frequencies are not as effective in generating the EFR, causing lower amplitude signals and a loss of detection at lower stimulus levels, resulting in higher thresholds. Yuen *et al.* (2005) reported somewhat better agreement between AEP and behavioral data at lower frequencies; however, these data were masked by ambient noise in Kaneohe Bay and the thresholds themselves were rather high (about 100 dB re 1 μPa at 4 kHz).

At 150 kHz, the difference between AEP and behavioral data was dramatic, not only in terms of thresholds, but also in the relationship between correct detection and stimulus level (Fig. 2). Although the AEP percent correct data looked similar to that at the other frequencies, the behavioral data at 150 kHz exhibited an extremely sharp cutoff, with the percent correct changing from 10% to 100% over only a 2-dB step (105–107 dB re 1 μPa). This behavior was unlike that at any of the other frequencies and raises some questions as to whether stimulus artifacts may have been present and audible to the subject. The stimulus update rate was 500 kHz (the maximum possible rate), meaning that at 150 kHz only 3.3 samples/cycle existed in the digitally created wave form. Although above the Nyquist limit, this rate may have been insufficient to produce a truly "clean" signal free from audible artifacts. This explanation, however, seems unlikely as a similar issue would be expected at 130-kHz signals where the behavioral and AEP detections are in close agreement. Furthermore, the false alarm rate is low at 150 kHz and similar to other false alarm rates suggesting that no artifact existed on which the subject could capitalize.

An alternative explanation is that some physiological process, such as a nonlinearity of the basilar membrane input-output function, occurs at 150 kHz resulting in a pronounced mismatch of the tonal perception and AEP threshold. Near frequencies where thresholds are impaired, perceived loudness can grow more quickly than at frequencies with normal or lower thresholds. This phenomenon of loudness recruitment is generally associated with cochlear impairment (Hallpike and Hood, 1951; Phillips, 1987) but may also have origin at cochlear regions associated with the upper frequency limit of hearing. Steep AEP input-output functions at 150 kHz are suggestive of recruitment in dolphins at this frequency. Prior work has demonstrated that humans exhibiting loudness recruitment also demonstrate a selective perceptual magnification of modulation depths in amplitude modulated signals (Moore *et al.*, 1996). It is unknown whether any electrophysiological measures correlate with this perceptual enhancement. However, if such a phenomenon exists in the bottlenose dolphin at the upper frequency limit of hearing, it may explain the discrepancy in AEP and behavioral thresholds observed here. It is interesting to note that Szymanski *et al.* (1999) also show a similar trend of significantly elevated AEP thresholds at the higher frequencies near the subject's upper cutoff frequency.

V. CONCLUSIONS

Some of the larger deviations between behavioral and AEP thresholds found in previous comparison studies can be attributed to differences in experimental parameters, such as sound source, stimulus wave form, duration, and presentation, test environment, and threshold analysis. In this study, when test parameters were held constant, there was good agreement between behavioral and electrophysiological methods, except at the lower frequencies, where the stimulus is less effective at evoking the EFR, and at 150 kHz.

ACKNOWLEDGMENTS

We thank the animal training staff at SSC San Diego, especially C. Quintos and L. Lewis. E. Hernandez and H. London assisted with the data collection. Financial support was provided by the U.S. Office of Naval Research, Marine Mammal Science and Technology Program and the SSC San Diego In-house Laboratory, Independent Research program.

Amos, D. E., and Koopmans, L. H. (**1963**). *Tables of the Distribution of the Coefficient of Coherence for Stationary Bivariate Gaussian Processes* (Sandia Corporation, Livermore, CA).

Andersen, S. (**1970**). "Auditory sensitivity of the harbour porpoise *Phocoena phocoena*," in *Investigations on Cetaceans*, edited by G. Pilleri (Berne, Switzerland), pp. 255–259.

Andre, M., Supin, A., Delory, E., Kamminga, C., Degollada, E., and Alonso, J. M. (**2003**). "Evidence of deafness in a striped dolphin, *Stenella coeruleoalba*," Aquat. Mamm. **29**, 3–8.

Awbrey, F. T., Thomas, J. A., and Kastelein, R. A. (**1988**). "Low-frequency underwater hearing sensitivity in belugas, *Delphinapterus leucas*," J. Acoust. Soc. Am. **84**, 2273–2275.

Brill, R. L., Moore, P. W. B., and Dankiewicz, L. A. (**2001**). "Assessment of dolphin (*Tursiops truncatus*) auditory sensitivity and hearing loss using jawphones," J. Acoust. Soc. Am. **109**, 1717–1722.

Brill, R. L., Sevenich, M. L., Sullivan, T. J., Sustman, J. D., and Witt, R. E. (**1988**). "Behavioral evidence for hearing through the lower jaw by an echolocating dolphin (*Tursiops truncatus*)," Marine Mammal Sci. **4**, 223–230.

Brillinger, D. R. (**1978**). "A note on the estimation of evoked response," Biol. Cybern. **31**, 141–144.

Bullock, T. H., Grinnell, A. D., Ikezono, E., Kameda, K., Katsuki, K., Nomoto, M., Sato, O., Suga, N., and Yanagisawa, K. (**1968**). "Electrophysiological studies of central auditory mechanisms in cetaceans," Zeitschrift fur Vergleichende Physiologie **59**, 117–156.

Cook, M. L. H., Varela, R. A., Goldstein, J. D., McCulloch, S. D., Bossart, G. D., Finneran, J. J., Houser, D., and Mann, D. A. (**2006**). "Beaked whale auditory evoked potential hearing measurements," J. Comp. Physiol., A **192**, 489–495.

Dobie, R. A. (**1993**). "Objective response detection," Ear Hear. **14**, 31–35.

Dobie, R. A., and Wilson, M. J. (**1989**). "Analysis of auditory evoked potentials by magnitude-squared coherence," Ear Hear. **10**, 2–13.

Dobie, R. A., and Wilson, M. J. (**1996**). "A comparison of *t* test, *F* test, and coherence methods of detecting steady-state auditory-evoked potentials, distortion-product otoacoustic emissions, or other sinusoids," J. Acoust. Soc. Am. **100**, 2236–2246.

Dolphin, W. F. (**1997a**). "Electrophysiological measures of auditory processing in odontocetes," Bioacoustics **8**, 79–101.

Dolphin, W. F. (**1997b**). "The envelope following response to multiple tone pair stimuli," Hear. Res. **110**, 1–14.

Dolphin, W. F., Au, W. W., Nachtigall, P. E., and Pawloski, J. (**1995**). "Modulation rate transfer functions to low-frequency carriers in three species of cetaceans," J. Comp. Physiol., A **177**, 235–245.

Finneran, J. J., Dear, R., Carder, D. A., Belting, T., McBain, J., Dalton, L.,

and Ridgway, S. H. (**2005**). "Pure tone audiograms and possible aminoglycoside-induced hearing loss in belugas (*Delphinapterus leucas*)," J. Acoust. Soc. Am. **117**, 3936–3943.
Finneran, J. J., and Houser, D. S. (**2006**). "Comparison of in-air evoked potential and underwater behavioral hearing thresholds in four bottlenose dolphins (*Tursiops truncatus*)," J. Acoust. Soc. Am. **119**, 3181–3192.
Gelfand, S. A. (**1990**). *Hearing: An Introduction to Psychological and Physiological Acoustics* (Marcel Dekker, New York).
Gellerman, L. W. (**1933**). "Chance orders of alternating stimuli in visual discrimination experiments," J. Genet. Psychol. **42**, 206–208.
GraphPad Software (**2003**). "GraphPad Prism," GraphPad Software, San Diego, CA.
Hall, J. D., and Johnson, C. S. (**1972**). "Auditory thresholds of a killer whale *Orcinus orca Linnaeus*," J. Acoust. Soc. Am. **51**, 515–517.
Hallpike, C. S., and Hood, J. D. (**1951**). "Some recent work on auditory adaptation and its relationship to the loudness recruitment phenomenon," J. Acoust. Soc. Am. **23**, 270–274.
Houser, D. S., and Finneran, J. J. (**2006a**). "A comparison of underwater hearing sensitivity in bottlenose dolphins (*Tursiops truncatus*) determined by electrophysiological and behavioral methods," J. Acoust. Soc. Am. **120**, 1713–1722.
Houser, D. S., and Finneran, J. J. (**2006b**). "Variation in the hearing sensitivity of a dolphin population obtained through the use of evoked potential audiometry," J. Acoust. Soc. Am. **120**, 4090–4099.
Jacobs, D. W., and Hall, J. D. (**1972**). "Auditory thresholds of a fresh water dolphin, *Inia geoffrensis Blainville*," J. Acoust. Soc. Am. **51**, 530–533.
Kastelein, R. A., Bunskoek, P., and Hagedoorn, M. (**2002**). "Audiogram of a harbor porpoise (*Phocoena phocoena*) measured with narrow-band frequency-modulated signals," J. Acoust. Soc. Am. **112**, 334–344.
Kastelein, R. A., Hagedoorn, M., and de Haan, D. (**2003**). "Audiogram of a striped dolphin (*Stenella coeruleoalba*)," J. Acoust. Soc. Am. **113**, 1130–1137.
Ljungblad, D. K., Scroggins, P., D., and Gilmartin, W. G. (**1982**). "Auditory thresholds of a captive Eastern Pacific bottle-nosed dolphin, *Tursiops spp.*," J. Acoust. Soc. Am. **72**, 1726–1729.
McCormick, J. G., Wever, E. G., Palin, J., and Ridgway, S. H. (**1970**). "Sound conduction in the dolphin ear," J. Acoust. Soc. Am. **48**, 1418–1428.
McCormick, J. G., Wever, E. G., Ridgway, S. H., and Palin, J. (**1980**). "Sound reception in the porpoise as it relates to echolocation," in *Animal Sonar Systems*, edited by R. G. Busnel, and J. F. Fish (Plenum, New York).
Møhl, B., Au, W. W. L., Pawloski, J., and Nachtigall, P. E. (**1999**). "Dolphin hearing: Relative sensitivity as a function of point of application of a contact sound source in the jaw and head region," J. Acoust. Soc. Am. **105**, 3421–3424.
Moore, B. C. J., Wojtczak, M., and Vickers, D. A. (**1996**). "Effect of loudness recruitment on the perception of amplitude modulation," J. Acoust. Soc. Am. **100**, 481–489.
Moore, P. W. B., Pawloski, D. A., and Dankiewicz, L. (**1995**). "Interaural time and intensity difference thresholds in the bottlenose dolphin (*Tursiops truncatus*)," in *Sensory Systems of Aquatic Mammals*, edited by R. A. Kastelein, J. A. Thomas, and P. E. Nachtigall (De Spil, Woerden, The Netherlands).
Nachtigall, P. E., Au, W. W. L., Pawloski, J., and Moore, P. W. B. (**1995**). "Risso's dolphin (*Grampus griseus*) hearing thresholds in Kaneohe Bay, Hawaii," in *Sensory Systems of Aquatic Mammals*, edited by R. A. Kastelein, J. A. Thomas, and P. E. Nachtigall (DeSpil, Woerden, The Netherlands), pp. 49–53.
Nachtigall, P. E., Yuen, M. M. L., Mooney, T. A., and Taylor, K. A. (**2005**). "Hearing measurements from a stranded infant Risso's dolphin, *Grampus griseus*," J. Exp. Biol. **208**, 4181–4188.
Phillips, D. P. (**1987**). "Stimulus intensity and loudness recruitment: Neural correlates," J. Acoust. Soc. Am. **82**, 1–12.
Popov, V., and Supin, A. (**1990a**). "Electrophysiological studies of hearing in some cetaceans and a manatee," in *Sensory Abilities in Cetaceans*, edited by J. A. Thomas, and R. A. Kastelein (Plenum, New York), pp. 405–415.
Popov, V. V., and Klishin, V. O. (**1998**). "EEG study of hearing in the common dolphin, *Delphinus delphis*," Aquat. Mamm. **24**, 13–20.
Popov, V. V., and Supin, A. Y. (**1986**). "Evoked potentials of auditory cortex in dolphin recorded from body surface," Dokl. Akad. Nauk SSSR **288**, 756–759.
Popov, V. V., and Supin, A. Y. (**1987**). "Characteristics of hearing in the beluga Delphinapterus leucas," Dokl. Akad. Nauk SSSR **294**, 1255–1258.
Popov, V. V., and Supin, A. Y. (**1990b**). "Auditory brainstem responses in characterization of dolphin hearing," J. Comp. Physiol., A **166**, 385–393.
Popov, V. V., and Supin, A. Y. (**1990c**). "Electrophysiological investigation of hearing in the freshwater dolphin *Inia geoffrensis*," Dokl. Biol. Sci. **313**, 238–241.
Popov, V. V., Supin, A. Y., Wang, D., Wank, K., Xiao, J., and Li, S. (**2005**). "Evoked-potential audiogram of the Yangtze finless porpoise *Neophocaena phocaenoides asiaeorientalis*," J. Acoust. Soc. Am. **117**, 2728–2731.
Ridgway, S. H., Bullock, T. H., Carder, D. A., Seeley, R. L., Woods, D., and Galambos, R. (**1981**). "Auditory brainstem response in dolphins," Neurobiology **78**, 1943–1947.
Ridgway, S. H., Carder, D. A., Kamolnick, T., Smith, R. R., Schlundt, C. E., and Elsberry, W. R. (**2001**). "Hearing and whistling in the deep sea: Depth influences whistle spectra but does not attenuate hearing by white whales (*Delphinapterus leucas*) (Odontoceti, Cetacea)," J. Exp. Biol. **204**, 3829–3841.
Sauerland, M., and Dehnhardt, G. (**1998**). "Underwater audiogram of a tucuxi (*Sotalia fluviatilis guianensis*)," J. Acoust. Soc. Am. **103**, 1199–1204.
Supin, A. J., Popov, V. V., and Mass, A. M. (**2001**). *The Sensory Physiology of Aquatic Mammals* (Kluwer Academic, Boston).
Supin, A. Y., Nachtigall, P. E., Pawloski, J., and Au, W. W. L. (**2003**). "Evoked potential recording during echolocation in a false killer whale *Pseudorca crassidens*," J. Acoust. Soc. Am. **113**, 2408–2411.
Szymanski, M. D., Bain, D. E., Kiehl, K., Pennington, S., Wong, S., and Henry, K. R. (**1999**). "Killer whale (*Orcinus orca*) hearing: Auditory brainstem response and behavioral audiograms," J. Acoust. Soc. Am. **106**, 1134–1141.
Thomas, J., Chun, N., Au, W., and Pugh, K. (**1988**). "Underwater audiogram of a false killer whale (*Pseudorca crassidens*)," J. Acoust. Soc. Am. **84**, 936–940.
Wang, D., Wang, K., Xiao, Y., and Sheng, G. (**1992**). "Auditory sensitivity of a Chinese River dolphin, *Lipotes vexillifer*," in *Marine Mammal Sensory Systems*, edited by J. A. Thomas, R. A. Kastelein, and A. Y. Supin (Plenum, New York), pp. 213–221.
White, M. J., Norris, J., Ljungblad, D. K., Baron, K., and di Sciara, G. N. (**1978**). *Auditory Thresholds of Two Beluga Whales (Delphinapterus leucas)*, (Hubbs Sea World Research Institute, San Diego).
Yuen, M. M. L., Nachtigall, P. E., Breese, M., and Supin, A. Y. (**2005**). "Behavioral and auditory evoked potential audiograms of a false killer whale (*Pseudorca crassidens*)," J. Acoust. Soc. Am. **118**, 2688–2695.

The effects of high-intensity, low-frequency active sonar on rainbow trout

Arthur N. Popper[a),b)] and Michele B. Halvorsen[b)]
Department of Biology and Center for Comparative and Evolutionary Biology of Hearing, University of Maryland, College Park, Maryland 20742

Andrew Kane
Department of Epidemiology and Preventive Medicine, University of Maryland, Baltimore, Maryland 21201 and Aquatic Pathobiology Center, VA-MD Regional College of Veterinary Medicine, 8075 Greenmead Drive, College Park, Maryland 20742

Diane L. Miller and Michael E. Smith
Department of Biology and Center for Comparative and Evolutionary Biology of Hearing, University of Maryland, College Park, Maryland 20742

Jiakun Song
Department of Biology and Center for Comparative and Evolutionary Biology of Hearing, University of Maryland, College Park, Maryland 20742 and Division of Fishes, National Museum of Natural History, Smithsonian Institution, Washington, DC 20013-7012

Peter Stein
Scientific Solutions, Inc., 99 Perimeter Road, Nashua, New Hampshire 03049

Lidia E. Wysocki
Department of Biology and Center for Comparative and Evolutionary Biology of Hearing, University of Maryland, College Park, Maryland 20742

(Received 3 October 2006; revised 30 March 2007; accepted 4 April 2007)

This study investigated the effects on rainbow trout (*Oncorhynchus mykiss*) of exposure to high-intensity, low-frequency sonar using an element of the standard Surveillance Towed Array Sensor System Low Frequency Active (LFA) sonar source array. Effects of the LFA sonar on hearing were tested using auditory brainstem responses. Effects were also examined on inner ear morphology using scanning electron microscopy and on nonauditory tissues using general pathology and histopathology. Animals were exposed to a maximum received rms sound pressure level of 193 dB re 1 μPa^2 for 324 or 648 s, an exposure that is far in excess of any exposure a fish would normally encounter in the wild. The most significant effect was a 20-dB auditory threshold shift at 400 Hz. However, the results varied with different groups of trout, suggesting developmental and/or genetic impacts on how sound exposure affects hearing. There was no fish mortality during or after exposure. Sensory tissue of the inner ears did not show morphological damage even several days post–sound exposure. Similarly, gross- and histopathology observations demonstrated no effects on nonauditory tissues. © *2007 Acoustical Society of America.* [DOI: 10.1121/1.2735115]

PACS number(s): 43.80.Lb, 43.80.Nd, 43.64.Wn, 43.50.Qp, 43.50.Rq [WWA] Pages: 623–635

I. INTRODUCTION

There is growing concern among investigators, regulators, and the public that human-generated (anthropogenic) sounds in the marine environment could have deleterious impacts on aquatic organisms (e.g., National Research Council, 1994, 2000, 2003; Richardson *et al.*, 1995; Popper, 2003; Popper *et al.*, 2004; Wartzog *et al.*, 2004). However, only a few peer-reviewed studies address these concerns. Although much of the interest in the effects of marine anthropogenic sound has focused on marine mammals, concerns are now being raised about the effects of these sounds on fish (see Myrberg, 1980; Popper, 2003; Popper *et al.*, 2004, 2005; Hastings and Popper, 2005). Many of the physiological systems in fishes are similar to those in marine mammals, suggesting that environmental sounds could impact the survival and/or health and well-being of fishes as well as mammals.

The potential effects of high-intensity sounds on fishes (as on marine mammals—Richardson *et al.*, 1995) may include temporary threshold shift, increased stress levels (e.g., Hattingh and Petty, 1992), and/or damage to other organ systems including the circulatory system, neural tissue, etc. (e.g., Cernak *et al.*, 1996; Dodd *et al.*, 1997). There is also the potential for behavioral effects including movement away from the source and alterations in feeding and mating (e.g., Løkkeborg *et al.*, 1996; Engås and Løkkeborg, 2002). Although most sound exposures do not cause immediate

[a)]Electronic mail: apopper@umd.edu
[b)]Dr's. Popper and Halvorsen contributed equally to this paper and serve as co-first authors.

death, as may occur as a consequence of exposure to the concussive forces of blasting (e.g., Yelverton *et al.*, 1975), the long-term effects of any of the stimuli may reduce the viability of animals or their reproductive success, with potential consequences at the population level.

Data on the direct effects of loud sounds on fishes are limited. Exposure to very loud sound for several hours results in damage to the sensory hair cells of the inner ear (Enger, 1981; Hastings *et al.*, 1996; McCauley *et al.*, 2003). Exposure to sounds for shorter periods can produce temporary hearing loss [i.e., temporary threshold shift] that lasts for some time after the termination of exposure (e.g., Popper and Clarke, 1976; Scholik and Yan, 2001; Amoser and Ladich, 2003; Smith *et al.*, 2004a, b; Popper *et al.*, 2005; Wysocki and Ladich, 2005). Lastly, almost nothing is known about the potential effects of sounds to other organ systems.

The current study examined whether exposure to high-intensity, low-frequency sonar, such as the Navy's Surveillance Towed Array Sensor System (SURTASS) Low Frequency Active (LFA) sonar, affects hearing sensitivity, the structure of the ear, and nonauditory systems in rainbow trout (*Oncorhynchus mykiss*). A LFA sonar array produces a signal with an effective source level of over 220 dB re 1 μPa^2 at 1 m and has the potential to ensonify fishes with sound pressure levels over 180 dB re 1 μPa^2 within 1 km from the ship-board array. Moreover, LFA sonar uses frequencies from 100 to 500 Hz, a range in which most fish are able to detect sound and the range of best hearing of many species (Fay, 1988; Fay and Popper, 1999; Popper *et al.*, 2003). Thus, LFA sonar not only has the potential to damage any number of organ systems in fishes due to the sound intensity but also has the direct potential of affecting hearing because the ears of fishes detect the operational frequency range of the sonar.

II. METHODS

A. Species

Species of great concern to users of LFA sound sources (as well as other human-generated sources) are listed and endangered salmonids on the West Coast of the U.S. While it would have been ideal to use those species in this study, their status precludes their availability for research. (Moreover, only species that are endemic to Seneca Lake could be used in this study.) At the same time, rainbow trout was available for these studies. This species has some potential for being a surrogate for a number of salmonids since it is of the same genus of most of the endangered and listed species. Furthermore, analysis of the structure of the ears of several different salmonid species, including rainbow trout (*Salmo gairdneri* Richardson), Atlantic salmon (*Salmo salar* Linnaeus) (Popper, 1977), Chinook salmon (*Oncorhynchus tshawytscha* Walbaum) (Popper, 1977; Song, private communication, and other salmonids (Popper, 1976, 1977) show that the ear is very similar in all of these species. Moreover, the general structure of the auditory system, and the lack of specializations for enhanced hearing, is the same in all salmonids.

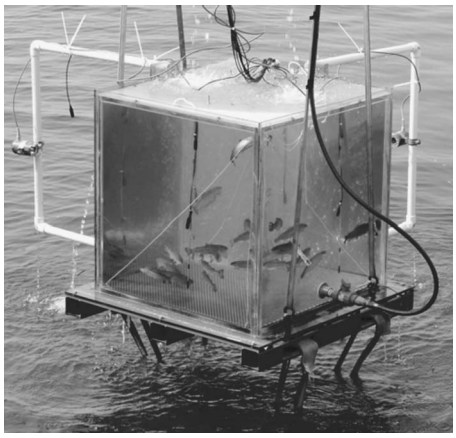


FIG. 1. The test tank with rainbow trout being removed from the water alongside the barge. The hose on the right brings cool water into the test tank. Hydrophones can be seen hanging inside the test tank and TV cameras were mounted on the two PVC brackets located on the outside. A 238.1-kg weight (below the water level) is suspended by slings from the bottom to stabilize deployment.

B. Test site

Experiments were conducted at the US Navy Sonar Test Facility in Seneca Lake, Dresden, NY (http://www.npt.nuwc.navy.mil/Seneca/index.html). The facility consists of a large barge in the center of the lake where the lake is 140.2 m deep. The facility includes a nearby shore support area with space for holding animals and for all post–sound exposure studies.

C. Test tank

Exposure to the LFA sound was conducted with fish in 1-m^3 test tanks made of 1.27-cm-thick clear Lexan® (Fig. 1) that were lowered into Seneca Lake from the barge (Fig. 2). Cool lake water was pumped into the test tank both during

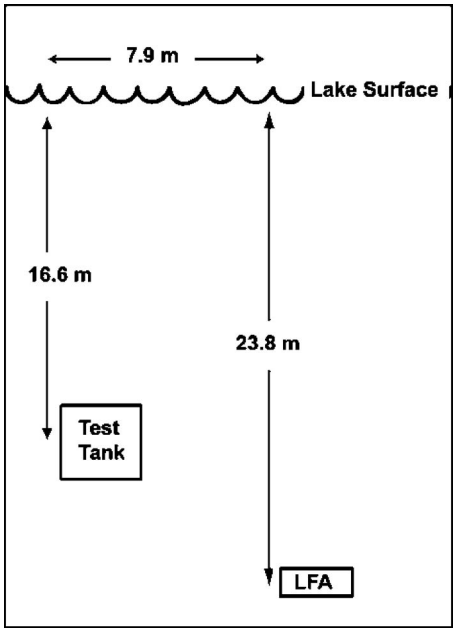


FIG. 2. Geometry of experimental setup. Distances are center of LFA-to-center of test tank and their respective depths. (Image is not to scale.)

experiments and on shore through an inlet valve located near the bottom of the test tank. Another valve placed near the top of the test tank allowed water to flow out. The test tank sat on a metal palette with 1.58-cm shoulder eyebolts that enabled the whole unit to be lifted by forklift or crane. The test tank was covered with a perforated polypropylene cover that allowed water to flow in and out of the test tank when it was submerged and for air to escape from within the test tank.

Six hydrophones (HighTech, Inc., model HTI-96-Min) were placed in the test tank during experiments. They were attached onto three fixed posts that were mounted to the perforated lid, with two hydrophones per post. Three of the hydrophones were 30.5 cm below the top of the tank and the other three were 20 cm lower. The output of each hydrophone was fed to a Wavebook (DAS model 516E) data-acquisition board using a sampling rate of 8 kHz and then to a Dell Inspiron 8500 laptop computer. An additional hydrophone was placed outside the test tank, 50.8 cm laterally to the test tank and 30.5 cm from the top. The hydrophone data-collection hardware was the same as for the calibration test (see the following). MATLAB was used to analyze the digitized signal for all experiments. Background noise spectral levels were measured daily during control tests and were consistently 100±2 dB re 1 μPa^2/Hz in the frequency band of interest.

Two Sea Viewer 650-420TVL video cameras were mounted on PVC brackets at right angles to each other on the outside of the tank (Fig. 1). These were used to monitor and record the behavior of the fish from the time the test tank entered the water until it was removed from the water. Data from the behavioral observations will be presented in a separate study.

D. Acoustic calibration in the test tank

During experiments and calibration, a crane lowered the test tank so that the center of the tank was 16.6 m below the surface of the water. A single LFA LTS (Low-Frequency Transmit System) transducer was at a depth of 23.8 m from its center to the water surface. The configuration of the setup is shown in Fig. 2.

Calibration of the system was performed to determine the maximum sound pressure level inside the submerged test tank and the uniformity of the sound field to provide a proper environment for the fish experiments.

Data from transmitted flat-spectra broadband noise (Fig. 3) in 1-Hz bins showed that the variation in level inside the test tank was ±3 dB. By comparison with the exterior hydrophone, it was also determined that the test tank was essentially acoustically transparent over the frequency range of interest. For a single tone, the maximum rms sound level was 193±3 dB re 1 μPa^2 at 196 Hz. Sound exposure level (SEL) is a measurement of a constant level of energy for 1 s. The calculated SEL was approximately 188.5 dB re 1 μPa^2 s. Finally, assuming particle velocity would be that for a propagating planar wave, the rms particle velocity is $v=p/\rho_0 c$, where p is rms pressure (~4500 Pa), ρ_0 is density

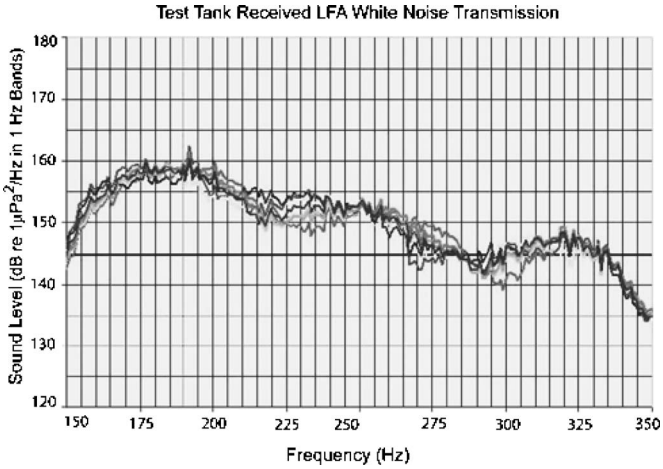


FIG. 3. White noise transmitted by the LFA as recorded on eight different hydrophones distributed throughout the TT.

(~1000 kg/m^3), and c is the speed of sound (~1500 m/s); thus the calculated maximum particle velocity inside the test tank was 0.003 m/s.

E. LFA test sounds

Sounds were produced using the single LFA transducer driven by 1600 V by two s11-48 amplifiers (Instrument, Inc.) connected in series. The output of the transducer gave an approximate source level of 215 dB re 1 μPa^2 at 1 m. The signal used was generated via MATLAB and played out through the Dell Inspiron internal sound card. The signal was slightly modified from the actual classified sonar signal train used by the US Navy. The output of the transducer was constantly monitored to ensure the integrity of the test signal.

The frequency range of the signal was from 170 to 320 Hz (Fig. 4), and it consisted of three hyperbolic frequency-modulated (HFM) sweeps. Each hyperbolic sweep produces a curved output, spending more time on the lower

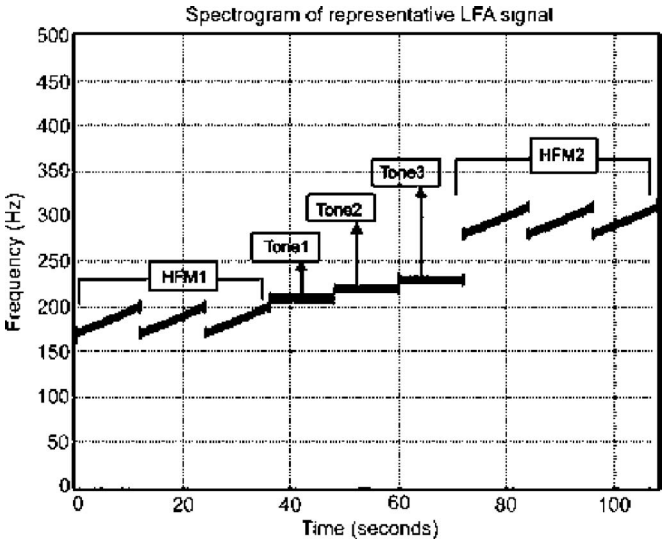


FIG. 4. Sound used in all experiments. The signal started with three HFM sweeps followed by three tones and concluded with three higher frequency HFM sweeps. (H=hyperbolic) (Note, in the MAX*2 experiments, this sequence was presented twice to give a continuous 216-s signal.)

TABLE I. Experimental groups defined.

Experimental groups	Definition
Baseline	Animals taken directly from the holding tank and not subject to any experimental handling. These served as "controls" for the actual control animals to ensure that handling and the experience in the experimental tanks did not cause any of the results ultimately found.
Control animals	Animals given the identical treatment as sound-exposed animals except that they were not presented with sound. They were kept at depth for the same length of time as the experimental animals.
MAX	Animals subject to three ensonifications of the signal (108 s duration) with a quiet period of 9 min between them.
MAX_24	Animals exposed to the maximum source level and tested 24 h postexposure.
MAX_48	Animals exposed to the maximum source level and tested 48 h postexposure.
MAX*2	Signal duration was twice that of MAX as was the off period. Source level was the same as in MAX.
MAX*2_24h	Animals exposed to MAX*2 and tested 24 h postexposure.
MAX*2_48	Animals exposed to MAX*2 and tested 48 h postexposure.
MAX-12	Animals exposed to MAX source level attenuated by 12 dB.
MAX-12_24	Animals exposed to the MAX source level attenuated by 12 dB; tested 24 h postexposure.
MAX-18	Animals exposed to MAX source level attenuated by 18 dB.

frequencies. This type of signal is a re-creation of the way the Navy transmits the signal when it uses the LFA sonar. The first part of the signal was a tone swept from 170 to 210 Hz three times followed by three tones at 210, 220, and 230 Hz, finally followed by three more HFM sweeps from 280 to 310 Hz.

Tests were conducted with several different sound exposures as described in Table I. Test tones in all but the MAX*2 experiments consisted of three repetitions of test sound (108 s each) separated by a 9-min silent period. In the MAX*2 trials, each sound presentation was repeated twice in succession for a 216-s-long signal with an 18-min quiet period. The longer quiet interval was required with MAX*2 in order to maintain a 16.7% duty cycle for all experiments and to allow the LFA transducer to cool. The overall sequence for each test was lower test tank to depth-transmit the signal-quiet-repeat signal-quiet-repeat signal-lift test tank out of the water onto barge.

F. Fish handling

Fish (170–275 mm total length; 90–212 g) were obtained from the Fingerlakes Fish Farm, Seneca Falls, NY. The history of the fish, including the source of eggs, could not be ascertained. Prior to transport to the Seneca Lake test facility, fish were gradually temperature acclimated to test conditions. They arrived at the study site in good health. Upon delivery, fish were transferred to holding tanks into which fresh, cool lake water was continuously circulated to maintain water quality. All fish handling and experiments were supervised and approved by the Institutional Animal Care and Use Committee (IACUC) of the University of Maryland, College Park, MD.

Prior to each experiment, and at least 12 h before being used in an actual experiment, 20 fish were transferred from the holding tanks into one of three test tanks. Until the test tank was actually submerged for tests, it was enclosed in insulating foam to help maintain water temperature and to prevent the fish from receiving external visual stimuli, thereby minimizing stress in response to movements of the investigators.

When a test tank was to be used for a trial, it was lifted by a crane, placed onto a workboat, and transferred to the barge. Noise spectral levels within the test tank during transport to the barge were measured as 105 ± 5 dB re 1 $\mu Pa^2/Hz$. Once on the barge, the insulation was removed and the test tank was fitted with the perforated cover that included the hydrophones and video cameras (Fig. 1). The test tank was again hooked to a flow-through water system to maintain aeration and a stable temperature for the fish.

Setup of the test tank on the barge, which involved putting hydrophones and cameras in place and hooking to the crane for lowering into the water, took approximately 10 min. It was then lifted by a crane and slowly lowered to the required depth over 15 min to allow the fish to acclimate to the change in water pressure. After the sound presentation, the test tank was lifted to the surface over 15 min and placed on the barge, the electronics were removed, and the insula-

tion was replaced. The test tank was then placed on a workboat and returned to shore where it was again hooked up to the lake water flow-through system.

G. Determination of hearing sensitivity

Once the test tank was returned to shore, three to six fish were removed sequentially for hearing tests at the US Navy shore facility on Seneca Lake. Once hearing tests were completed, fish were prepared for analysis of ear structure (see the following). Additional fish from each tank were used for pathology (see the following). The remaining fish were transferred to holding pens and tested up to 48 h postexposure to determine if there were changes in hearing capabilities and/or tissue damage. Classification of experimental groups follows the notation in Table I.

Hearing capabilities were measured using auditory brainstem response (ABR), a noninvasive method (e.g., Corwin *et al.*, 1982; Kenyon *et al.*, 1998). The ABR tank is a 35.6-cm-deep steel cylinder, 35.6 cm in diameter, with 1.27-cm-thick walls. An underwater speaker (UW30) is mounted through the bottom of the ABR tank. The tank has three legs 50.8 cm long, each with a vibration-isolation damping foot.

Fish were lightly anesthetized with buffered MS-222 and restrained in a mesh sling where the fish were suspended so that the top of the head was approximately 13 cm below the surface of the water.

A stainless steel electrode (Rochester Electro-Medical, Inc.) was inserted subdermally into the medial dorsal surface of the head between the nares as a reference electrode, while another electrode was inserted subdermally into the dorsal midline surface of the fish directly over the brainstem to record the brainstem neural response to sound stimulation. A ground electrode was placed in the water. All exposed metal surfaces of the reference and recording electrode tips that were not below the fish skin were insulated with cured fingernail polish.

Signals and the levels were produced and calibrated using a Tucker-Davis-Technologies (TDT) System 3 physiology apparatus with SigGen and BioSig software (TDT). The signals were played out through TDT RP2.1 and passed through a power amplifier (Hafler P1000) connected to the underwater speaker. Presented signals were 100-, 200-, and 400-Hz tone bursts with a 2-ms rise and fall time, 10 ms in duration, and were gated through a Hanning window. Brainstem responses were collected using the BIOSIG software package with at least 200 responses averaged for each presentation. The resulting levels of each presented frequency were confirmed using a hydrophone (G.R.A.S. 10CT hydrophone with a receiving sensitivity of -211 dB ± 3 dB re 1 V/μPa connected to a Kistler 5010 dual-mode amplifier).

The signal levels at each frequency were changed in 5-dB steps until a typical ABR wave form could not be seen above the noise. Threshold is defined as the lowest intensity level where a response above the background noise could be seen. The traditional determination of threshold is achieved through visual inspection, which provides results that are virtually identical to those determined using statistical approaches (Mann *et al.*, 2001; Higgs *et al.*, 2003). In addition, two experimenters independently evaluated all data and found agreement with the determined thresholds.

The focus of these hearing experiments was on relative measures of hearing sensitivity (Hawkins and Johnstone, 1978), i.e., *changes* in hearing sensitivity as a result of exposure to the LFA source. Thus, data are provided as relative measures of threshold in terms of differences between experimental and control hearing thresholds. Relative measures were maintained by placing fish at the same depth and position in the ABR tank and by keeping the water depth the same for all experiments.

H. Statistical analysis of hearing tests

All data sets were tested for normal distribution using the Kolmogorov-Smirnov test. When data were normally distributed and showed homogeneous variances, audiograms of the different exposure groups were compared using a two-way analysis of variance (ANOVA) following a general linear model where one factor was treatment (exposure group) and the other factor was frequency (100, 200, and 400 Hz). The treatment factor alone indicates overall differences between exposed groups of animals. When the treatment factor is combined with the frequency factor, it indicates overall differences at the tested frequencies.

In order to determine which frequencies differed in their thresholds, one-way ANOVAs were followed by Scheffé's multiple comparison procedure at each test frequency. The p values of statistical significance for these one-way ANOVAs and Scheffé's were adjusted (using the Bonferroni correction) by 3, for the number of frequencies tested ($^{*}p \leqslant 0.0167$, $^{**}p \leqslant 0.0033$, $^{***}p \leqslant 0.0003$).

For nonparametrical data, Kruskal-Wallis tests were used at each frequency to test overall group differences, followed by pairwise Mann-Whitney U tests. Levels of significance were adjusted using the Bonferroni correction, as above.

I. Necropsy and histopathology

Fish from the first two experimental weeks (Group I—see the following) for Controls ($n=30$), MAX ($n=22$), MAX*2 ($n=19$), and Baseline ($n=6$) were sacrificed using buffered MS-222. Gross pathological observations and length and weight data were taken for each fish after sacrifice (Kane, 1996; Kane *et al.*, 2000). Gross pathological observations examined the condition of the skin, eyes, fins, mesenteric vasculature, and the swim bladder to assess the general condition of the fish and look for possible exposure-related hemorrhage. General parasitology was conducted from skin scrapes and gill biopsies and ranked based on incidence of the different taxa observed.

Blood samples were drawn from the caudal vein, transferred to heparinized hematocrit tubes, sealed, and centrifuged to determine hematocrit (packed red blood cell volume). Hematocrit tubes were also observed for sanguineous color in the plasma layer that might be indicative of possible LFA effect on red blood cells.

TABLE II. Auditory threshold differences between postexposure means and controls (combined baseline and control) means at each test frequency (mean dB ± SEM).

	Δ 100 Hz	*n*	Δ 200 Hz	*n*	Δ 400 Hz	*n*
			Group I. May and June, 2004			
MAX	4±1.95	16	5±1.83	17	21±2.20	20
MAX_24h	2±1.63	9	4±1.57	10	25±2.33	10
MAX*2	−3±1.42	6	−2±1.26	6	20±1.50	6
MAX*2_24h	0±1.73	4	−3±1.73	4	17±1.73	4
MAX*2_48h	0±1.94	4	1±1.88	4	25±1.63	4
			Group II. October, 2004 and May, 2005			
MAX	−2±2.34	12	−1±1.91	12	3±2.37	12
MAX_24h	−2±2.29	12	−2±2.30	12	1±2.59	12
MAX_48h	2±3.19	6	−2±3.20	6	−5±3.26	6
MAX-12	4±2.09	18	5±2.07	18	5±2.08	18
MAX-12_24h	−3±2.36	12	−3±2.32	12	3±2.42	12
MAX-18	0±2.64	10	−3±2.80	10	5±2.61	10

Histopathology samples ($n=8-10$ per exposure group) were taken from necropsied specimens immediately after sacrifice and external examination (Kane, 1996). During necropsy, the swim bladder was examined for any signs of hemorrhage that might be associated with sound exposure. The presence of any external or internal anomalies were noted and digitally photographed. Tissues, including eye, skin, gill, muscle, heart, posterior kidney, liver, spleen, and swim bladder, were preserved in 10% neutral-buffered formalin. Preserved tissues were processed for routine histology and stained with hematoxylin and eosin prior to coverslipping.

Glass slides were observed blindly (without knowing the exposure group) to evaluate the various tissues for histopathology. Tissue-level anomalies were noted and ranked on a scale of 0–5, where 0=no pathology and 5=severe. Examples of lesions typical of an exposure group, if any, were digitally photographed and recorded.

J. Examination of inner ear tissue

Immediately after ABR, each fish was euthanized with buffered MS-222. Some fish ear tissue was obtained 24, 48, and 96 h postexposure. The cranium was opened, and fixative (2.5% gluteraldehyde and 2.5% paraformaldehyde in 0.1 Millonig's buffer at pH 7.4) was injected into the cranial cavity as the initial fixation. The cranium and the ear were then placed in cold fixative solution for at least 24 h and then transferred to cold buffer for storage until use.

The ears were then dissected from the cranium, counterstained with 1% osmium tetroxide, and dehydrated through a series of ethanols to 70%. The sensory epithelia were then removed for SEM, dehydrated in 100% ethanol, and critical point dried using liquid carbon dioxide as the intermediary fluid. After mounting on aluminum stubs, the epithelia were coated with a 25-nm-thick layer of platinum and studied using an AMRAY model 1820 scanning electron microscope.

The sensory cells of the three otolithic end organ epithelia (saccule, lagena, and utricle) were examined for possible morphological damage resulting from LFA exposure. Tissue was examined in terms of the integrity of the sensory tissues by comparing exposure groups to control and baseline animals. In addition, the results were compared to morphological hair cell damage that was encountered in a number of earlier studies (e.g., Enger, 1981; Hastings *et al.*, 1996; McCauley *et al.*, 2003).

K. Grouping experimental weeks

The results reported here were determined during four different experimental weeks at Seneca Lake—May, June, and October, 2004 and May, 2005. The hearing data for May and June, 2004 were similar and they are pooled in the following discussions. Similarly, the hearing data for October, 2004 and May, 2005 were similar and are pooled. For ease of discussion and because differences existed in each group's respective baseline thresholds, May-June (Group I) and October-May (Group II) data are reported separately. The difference between the combined control/baseline thresholds of Group II from Group I was 20 dB at 100 Hz, 12 dB at 200 Hz, and −18 dB at 400 Hz (i.e., Group I control thresholds were 20 dB greater than Group II, etc.). Mean data for the two groups are presented in Table II. Nonparametrical statistical tests were applied to Group I results, whereas parametrical tests were used for Group II.

III. RESULTS

A. Baseline and control hearing thresholds

In order to determine whether there were threshold shifts due to LFA exposure or due to the handling of the fish, effects on hearing and on body tissues were measured in control and baseline animals. Controls were treated precisely as exposed animals other than for actual ensonification. Baseline animals were from the same stock as control and exposed animals but they were not placed in test tanks or transported to the barge. The baseline served as a control for handling. There were no differences in the detection sensitivity between control and baseline for either group (Group I: Mann-Whitney U tests at each frequency, all p values >0.05; Group II: unpaired t-tests at each frequency, all p values >0.05). Consequently, data for the baseline and control animals within each group were pooled.

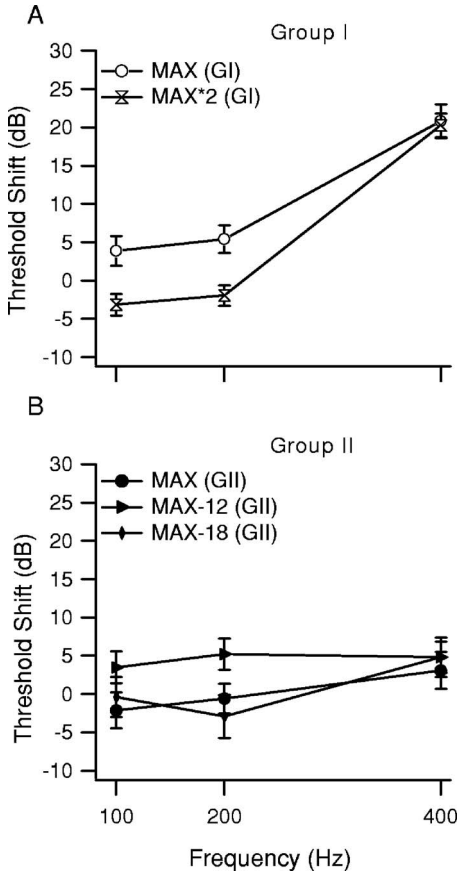


FIG. 5. Relative thresholds shifts immediately after LFA exposure; (A) Group I (GI) includes MAX and MAX*2 exposures; significant shifts were at 400 Hz for both exposures. (B) Group II (GII) includes MAX, MAX-12, and MAX-18 exposures. A significant shift was at 200 Hz for MAX-12. (Note: Auditory thresholds after exposure were subtracted from the combined baseline/control auditory thresholds giving threshold shifts.)

B. Hearing thresholds immediately after exposure to LFA sounds

Kruskal-Wallis tests at each frequency (levels of significance had Bonferroni correction) in Group I animals revealed significant differences in the ABR thresholds between control, MAX and MAX*2 animals at 200 Hz [Fig. 5(a)] ($\chi^2_{2,41}=8.62$, $p=0.013$) and at 400 Hz ($\chi^2_{2,43}=27.56$, $p=\leqslant 0.001$).

At 400 Hz, the thresholds of MAX and MAX*2 animals were 20 dB higher than those of the control animals and significantly differed from the control values ($p\leqslant 0.001$) and not from each other ($p=0.322$). The difference at 200 Hz was mainly due to the MAX group having an average threshold 5 dB higher than that of controls [Fig. 5(a), Table II]; however, subsequent Mann-Whitney *U* tests did not reveal this to be significantly different.

Comparisons between audiograms of Group II animals showed significant overall differences between the exposure groups (two-way ANOVA: $F_{3,191}=7.14$, $p\leqslant 0.001$) but no significant interaction between treatment (exposure groups) and frequency ($F_{6,191}=1.64$, $p=0.138$), indicating that the changes in thresholds showed the same trend at all frequencies tested. This difference was due to the MAX-12 animals being significantly different from all the other groups (Scheffé's post hoc tests: $p<0.05$ in all cases). Subsequent one-way ANOVAs at the separate frequencies showed that this difference was due to a shift in thresholds at 200 Hz ($F_{3,63}=5.61$, $p=0.002$), (Fig. 5).

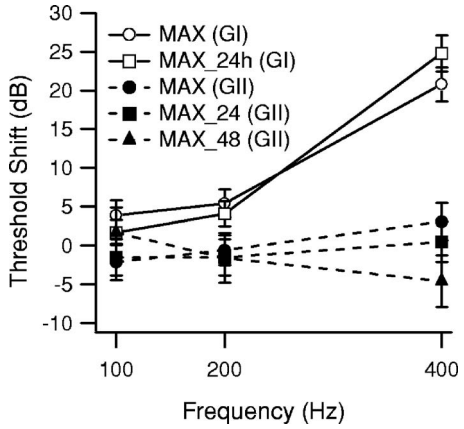


FIG. 6. Postexposure threshold shift for Group I (GI) and Group II (GII) trout following MAX exposure. Group I (open symbols) showed significant threshold shift at 400 Hz out to 24 h. Group II (closed symbols) did not show threshold shifts and there was no threshold change up to 48 h postexposure.

C. Postexposure hearing thresholds

In order to determine the time course of threshold shift after exposure, rainbow trout were tested up to 48 h postexposure. Group I showed significant threshold shifts in postexposure auditory measurements with threshold shifts at 400 Hz of 25 dB for MAX_24h, 17 dB for MAX*2_24h, and 25 dB for MAX*2_48h (Table II). Controls, MAX, and MAX_24h animals showed significant differences at 400 Hz (Kruskal-Wallis: $\chi^2_{2,47}=30.26$, $p\leqslant 0.001$) and none at 100 or 200 Hz. At 400 Hz for both MAX and MAX_24h, thresholds significantly differed from control thresholds (Mann-Whitney *U* tests: $p\leqslant 0.001$ in both cases): indicating significant threshold shifts remained 24 h post-LFA exposure (Fig. 6).

Control, MAX*2, MAX*2_24h, and MAX*2_48h animals showed significant differences at 400 Hz (Kruskal-Wallis: $\chi^2_{3,32}=25.87$, $p\leqslant 0.001$) and none at 100 or 200 Hz. Subsequent Mann-Whitney *U* tests showed that MAX*2 as well as MAX*2_24h and MAX*2_48h were significantly different from control thresholds ($p\leqslant 0.001$ in all cases), indicating that auditory thresholds at 400 Hz were still elevated relative to controls even 48 h postexposure (Fig. 7).

In Group II, the auditory thresholds of MAX animals were not different from the controls. A two-way ANOVA (including controls, MAX, MAX_24h, and MAX_48h) showed no difference in thresholds out to 48 h postexposure ($F_{3,161}=0.52$, $p=0.671$). In contrast, the control and MAX-12 auditory thresholds differed significantly (two-way ANOVA: $F_{2,161}=12.33$, $p\leqslant 0.001$ for the factor treatment group). MAX-12 differed significantly from the control values ($p<0.001$).

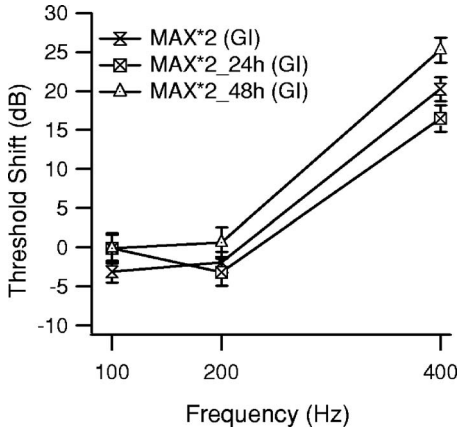


FIG. 7. Post-MAX*2 exposure threshold shifts for Group I trout. Significant shifts were at 400 Hz for all three test groups.

D. Effects on inner ear morphology

The morphology of the sensory epithelia of the end organs of LFA-exposed fish were examined and compared to control and baseline fish. Figure 8 shows the results from MAX exposure, while Fig. 9 shows MAX*2 exposure. There was no apparent morphological damage to any end organ sensory epithelia or sensory cells in any of the exposure groups, even in fish that were examined 96-h postexposure. (Note, 96 h animals were not tested for hearing sensitivity due to time constraints during the weeks at Seneca Lake.)

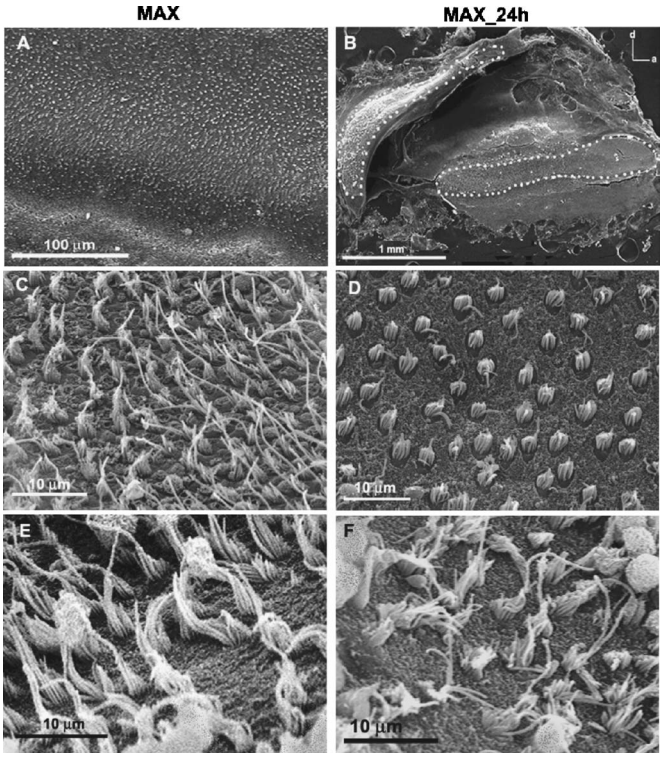


FIG. 8. SEM of auditory sensory epithelia and hair cells from Group I trout. (A) MAX exposure: Low-power image of a region of the left saccule. Note that there were no missing ciliary bundles. (B) MAX_24h: Low-power image of a complete lagena (left) and complete saccule (right). The sensory epithelia are outlined with a dashed line. Both end organs were completely intact, and there was no indication at low power of damage. (C) MAX exposure: High-power image of ciliary bundles on a region of the right utricle. (D) MAX_24h: A region of the left saccule. There was no indication of damage. Note the ciliary bundles on the saccule are shorter than those on the utricle in C. (E) MAX exposure: A region of the left lagena. (F) MAX_24h: A region of the left lagena.

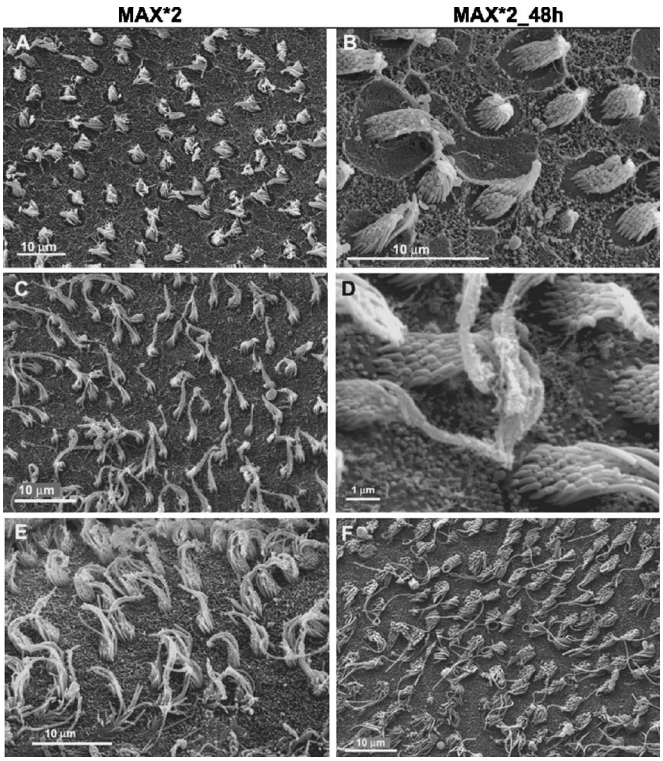


FIG. 9. SEM of auditory sensory epithelia and sensory hair cells of Group I. (A) MAX*2 exposure: Low-power image of the right saccule. (B) MAX*2_48h: left saccule; (C) MAX*2: right utricle; (D) MAX*2_48h: right utricle at higher power to show details of the stereocilia and how they are graded in size. (E) MAX*2: left lagena. (F) MAX*2_48h: left lagena.

E. Gross pathology and histopathology

Eighty-two rainbow trout were examined (26 controls; 56 exposed) from Group I experiments for general observations and gross pathological changes associated with exposure to LFA, while eight animals from each sound exposure parameter were analyzed for histopathology. All comparisons were between exposed animals versus baseline and control animals. Several animals exhibited mild to moderate skin or fin abrasion but these alterations were likely due to transport and/or netting since similar abrasions were found in baseline, control, and exposed animals. There were no changes to the external anatomy of any of the sound-exposed specimens and no observations of bleeding from the swim bladder or any vasculature. Swim bladders from all groups (exposed, baseline, and control) were intact and still inflated upon necropsy.

Blood was taken from baseline, control, and exposed fish to observe the plasma for the presence of pink or red coloration as a sign of possible erythrocyte membrane disruption in association with LFA exposure. Hematocrit data were derived from two control groups ($n=15$), two MAX groups ($n=20$), and two MAX*2 groups ($n=16$). The mean hematocrit values for the control, MAX, and MAX*2 groups were 43.6%, 43.5%, and 44.4%, respectively. A parametric ANOVA was used to analyze hematocrit between exposed

groups because the data met the assumptions of normality and homoscedasticity. There were no statistical differences between the hematocrits of these groups ($p=0.92$) and the hematocrit values derived were within normal range for rainbow trout (Rehulka and Minarik, 2004). The plasma of all of the blood samples was clear to straw colored and was not pink or red, indicating that erythrocytes remained largely intact throughout the different exposures as well as the blood collection.

Parasitology data from rainbow trout gill biopsies and skin scrapes indicated the presence of several different taxa of external parasites. There were three protozoan ciliate parasites observed (*Trichodina, Ichthyopthirius*, and *Ambiphyra*), and one flagellate (*Ichthyobodo*). Metazoan parasites included monogenean trematodes (skin and gills) and copepods (gills only). Ranking of parasite observations ranged from 0 to 3 for any given parasite. For statistical analysis, each exposure group from the different experimental replicates (control, MAX, and MAX*2) were tested for differences using nonparametric Kruskal-Wallis. No significant differences were found among the exposure groups; therefore, the data for each exposure were pooled for analysis. A Kruskal-Wallis ANOVA was used to test for differences between exposure groups (control, MAX, and MAX*2) and parasite rank. There were no statistical or empirically observed differences in parasite rankings between control and LFA-exposed groups, indicating that acoustic exposure did not alter the incidence of parasite observations.

Observations were made on histological tissues from brain, eye, heart, gills, stomach, intestine, liver, pancreas, spleen, anterior kidney, posterior kidney, swim bladder, skin, muscle, and gonad. There were no exposure-related pathologies observed in any of the tissues from exposed or control animals (Fig. 10).

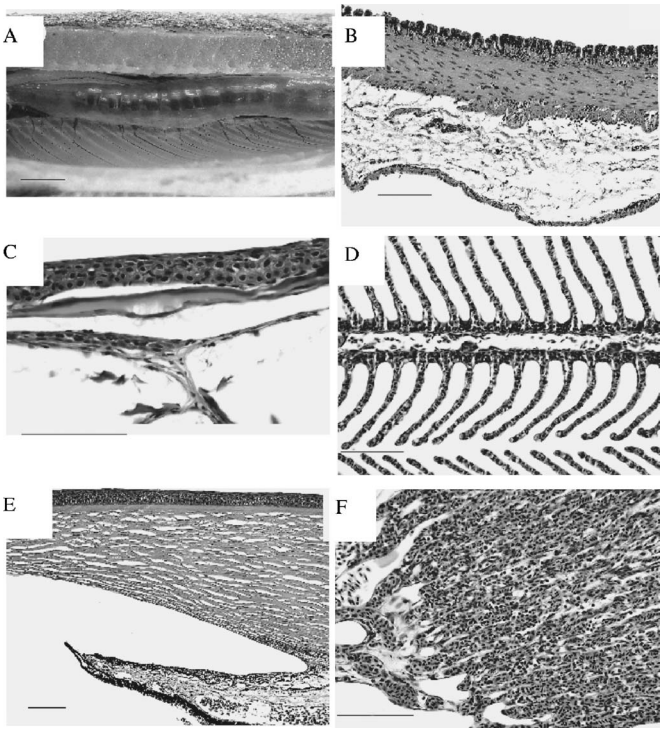


FIG. 10. Typical gross and histological observations from MAX-exposed rainbow trout. Observations of various tissues focused on different epithelia as well as areas of microvascularization. (A) Grossly intact, inflated swim bladder after removal of visceral organs (scale bar 10 mm). (B) Cross section of swim bladder (scale bar 100 μm). (C) Cross section of skin showing normal epithelium and scale pocket (scale bar 100 μm). (D) Gill tissue showing primary and secondary lamellae with intact epithelium and supporting cells (scale bar 100 μm). (E) Cross section of eye showing corneal epithelium and iris. (F) chorid (rete) of the eye (scale bar 100 μm).

IV. CONCLUSIONS

The order of conclusions begins with the experimental paradigm this study generated and followed, which serves as a guideline for future research studying the effects of noise exposure on fish. Next is a discussion of results of hearing sensitivity and the differences between Group I and Group II, followed by histopathology results. Finally, there is a discussion of the effects of LFA sonar on fish and the application of the results reported here to other species and other sources of intense sounds.

A. Guidelines for noise exposure studies

The Naval Sonar Test Facility (the experimental site) enabled development of an experimental paradigm that is optimal in terms of acoustic environment, calibration of the sounds to which the fish were exposed, and the ability to use a sound source that is virtually the same intensity as under actual use. In essence, this study was designed to eliminate the problems in earlier studies that resulted from lack of ability to calibrate the sound field and have control over most experimental parameters. The study provides a general experimental approach applicable to experiments on the effects of other anthropogenic sounds. Of course, the facilities available for this study were exceptionally conducive to this kind of work and access to comparable facilities for studies of other types of sounds may be limited. Thus, while future studies may not be able to precisely replicate the procedures used here vis-à-vis sound presentation, there are several experimental "principles" developed in this study that are applicable to other investigations. These include the following.

(1) Use of a fully calibrated and consistently monitored sound source. To characterize the sound field, pressure measurements were simultaneously made at six locations. The measured pressure was within 3 dB at all locations, indicating little pressure gradients. This is consistent with expectations of being in the free field and justifies an assumption that the particle velocity is related to the pressure by $(1/\rho c)$. Thus, signal characteristics were known and compared between trials, allowing for assurance that the signal was consistent.
(2) The test tanks were thoroughly analyzed in terms of acoustics (i.e., sound pressure, particle velocity, uniformity of signal inside the test tank, signal itself, etc). Thus, all factors of the sound field to which the fish were exposed were known.
(3) There was constant monitoring and digital storage of LFA signal presentations inside the test tank during all experiments. The sound pressure levels, frequency, envelope of the signal, and other sound characteristics were analyzed after each presentation.

(4) Each experiment, from the moment the test tank containing fish was placed in the water to the time it came out was video recorded and digitally stored. This provided a full record of the experiments for later analysis and allowed for detailed behavioral analyses to determine how fish responded to the sound.
(5) Relative hearing studies were done on fish immediately after sound exposure and continued for several days postexposure to investigate delayed effects and/or potential recovery.
(6) At least two ABR experts independently examined the hearing data sets. Thus, there was no bias in determining hearing thresholds based upon any preconceived expectations. The nature of the experiments did not allow for a double-blind study because the results from each day influenced the subsequent test days within each experimental week.
(7) Morphological analysis of the inner ear epithelia and sensory hair cells was performed on test animals by an expert anatomist.
(8) Gross pathology and follow-up histopathology were done by an expert fish pathologist. Internal organs, including the swim bladder, and associated microvasculature of exposed fish were examined and compared with control and baseline animals.
(9) The studies were done in a realistic environment that provided an excellent sound field as well as control of most variables that typically impact a "field" study.
(10) A significant number of control tests were performed along with an additional group of controls (referred to as baseline) for handling of the experimental and control animals.

B. Effects on hearing sensitivity

This is the first study to examine the effects of high-intensity low-frequency sonar on any fish species and is one of only two published studies to examine effects of any high-intensity sound on fish hearing (Popper *et al.*, 2005). The results from the hearing tests using ABR demonstrate that exposure to the LFA may affect hearing thresholds. However, as discussed in the following, the exposure to the LFA signal paradigm used in this study is more severe than fish would encounter in the wild. Therefore, the hearing losses found here represent an absolute worst-case scenario for hearing exposure.

There is evidence for threshold shifts in Group I at 400 Hz in rainbow trout lasting 48 h postexposure. In Group I at 400 Hz, the MAX and MAX_24h exposures had 20–25 dB shifts and the MAX*2, MAX*2_24h, and MAX*2_48h had shifts ranging from 17 to 25 dB (Table II). However, it should be noted that limited numbers ($n=4$) of the MAX*2_24h and MAX*2_48h were available and thus may not offer sufficient statistical power. Regardless, the overall results from Group I show that LFA sonar exposure does affect trout hearing thresholds and may last for at least two days.

In Group II, the only significant threshold shifts occurred at MAX-12 exposure. The fact that the maximum exposure (MAX) did not cause threshold shifts is perplexing. The Group II control fish had a 20-dB lower threshold at 200 Hz than the Group I fish, which could account for the shift that occurred at 200 Hz but does not offer any explanations as to why the MAX exposures had no effect on hearing thresholds. Also perplexing was the finding that Group I MAX had a large threshold shift at 400 Hz and Group II MAX had no change in threshold sensitivity. But, for Group I control fish-thresholds at 400 Hz were 20 dB lower than Group II control. This could somewhat account for the lack of an effect seen in Group II MAX at 400 Hz. Overall, Group II's data supports Group I's results that the LFA does affect hearing sensitivity, but the results from Group II call into question the degree to which the very long exposure to LFA signals affects fish hearing sensitivity.

It is possible that a 10–20 dB threshold shift at one frequency has an impact on survival of fish due to a decreased ability to detect biologically relevant sounds. However, not enough is known about the use of sounds by rainbow trout (or any other species) to determine whether such a decrement in hearing at a limited point of its hearing bandwidth would impair the detection of relevant sounds and the consequences of this for fish survival.

C. Differences in exposed groups

An issue with this study was the substantial differences in thresholds between Group I (May/June, 2004) experiments and Group II (October, 2004/May, 2005) experiments. The basis for these differences is not immediately clear. All animals came from the same supplier and were maintained and handled in an identical manner once they were brought to the experimental site, and the actual experiments were identical during each test week.

Factors that could not be controlled were the way the fish were raised at the fish hatchery and the genetic stock from which the animals came (something not controllable by fish farms). While all fish used in this study were of similar size, it is possible that their husbandry treatment had differing conditions at various times of the year or factors such as egg manipulation, water quality, feed, or other variables that could not be controlled. How such factors might affect hearing is unknown and would be almost impossible to control for.

Recent studies have shown that a hatchery-raised salmonid may differ in hearing sensitivity depending upon whether their otoliths (ear stones) have the typical aragonite crystal structure or if the otoliths are the far less common (and perhaps abnormal) vaterite crystalline form of calcium carbonate (Oxman, private communication). The discovery of these structural differences appears to be related to the way the fish were raised in hatcheries or to genetic diversities. Thus, if Groups I and II, used in the LFA study, were raised in dissimilar ways, it is possible that they had varying otolith structures and this could, in part, explain the different results reported here.

Additionally, a recent study on hearing sensitivity of two cohorts of rainbow trout from the same eggs but with slightly different handling prior to hatching showed significant dif-

ferences in hearing sensitivity, and this difference continued to show up over nine months of repeated testing (Wysocki, private communication). It is certainly the case that the LFA fish from Groups I and II came from different stocks, and it is possible that there are some genetic factors which impact hearing sensitivity and/or sensitivity to potential effects on hearing by anthropogenic sound.

The differences between Group I and II results are of interest and concern, and it may be key in demonstrating that reaching conclusions about the effects of sound on fish is not as simple as calculating sound levels or examining a single species. There is variability among fish within a species and going a step further to compare one species to another species becomes even more laden with complexities. These results may point to potential variability in effects based on a broad range of physiological issues such as water temperature and/or quality, general physiology of an animal, animal size, and numerous other factors that cannot be quantified or predicted. For example, fish of the same species but of different sizes could be affected by intense sounds differently.

D. Pathology and histopathology

Results indicate that LFA exposure at the levels described in this study caused no acute gross- or histopathology in rainbow trout. There were no changes in hematocrit between control and any of the LFA-exposed animals. LFA exposure did not cause shearing of red blood cell membranes as indicated by a lack of sanguineous color in the plasma. The rainbow trout is called a physostome because it has an esophageal connection to the swim bladder (i.e., a pneumatic duct) and is thus able to exchange gas from within the swim bladder directly through the mouth. It would be interesting to observe possible changes in a species that does not have a connection between the esophagus and the swim bladder (i.e., physoclistus condition without a pneumatic duct). Without the ability to "burp" swim bladder gas upon traumatic sound wave impingement, it may be more likely to observe changes in the microvasculature of the swim bladder and other organ systems.

The histological findings are in contrast to other studies using explosives or pile driving, many of which are in the gray literature, that suggest damage to the swim bladder and other body tissues because of exposure to explosives (e.g., Yelverton *et al.*, 1975; Sverdrup *et al.*, 1994; reviewed in Hastings and Popper, 2005). It is very hard, however, to compare the LFA results with results from other types of acoustic stimuli since the sounds are so notably different.

E. Inner ear morphology

The sensory epithelia and sensory hair cells showed no apparent morphological damage after any of the LFA exposures, even though there were shifts in hearing thresholds. However, it should be noted that tip link integrity was not accessed for this project. Results of the ultrastructural analysis using scanning electron microscopy suggest that the LFA sonar sounds do not cause gross morphological damage to the sensory epithelia or sensory hair cells in the inner ear. The effect LFA sounds might have on the physiological or metabolic changes of the sensory hair cells themselves is unknown and would be difficult to determine.

F. Effects of SURTASS LFA on fish

A single LFA sonar signal used in this study was made up of three parts. The LFA signal is turned on for 108 s, followed by a quiet of period of 9 min. This "signal followed by quiet" is presented three times, giving a total of 324 s of signal exposure for our MAX tests and 648 s of signal exposure for our MAX*2 tests.

Typical use of LFA sonar is on a ship moving about 3 knots (1.5 m/s). The critical issue from these studies is whether LFA sonar exposure impairs the survival of fishes and, more important, whether survival would be impaired in a normal environment when a ship using LFA is in the vicinity of a fish. In answering this question, several factors must be taken into consideration.

First, the sound level to which fish were exposed in these experiments was 193 dB re 1 μPa^2, a level that is only found within about 100 m of a ship using LFA. Thus, the likelihood of exposure to this or higher sound levels is small, considering all the possible places a fish might be relative to the sound source. In effect, the area of the ocean ensonified by a single LFA source at 193 dB re 1 μPa^2 or higher is small compared to the ocean area ensonified by the LFA source at lower sound levels.

Second, the presentation of the LFA sound in this study represents substantially longer exposure than might be encountered by fishes in the wild because during actual LFA use, the source is on a moving ship. An absolute worst-case scenario would be a fish following the ship during ensonification however, this is very unlikely. A realistic worst-case scenario is a stationary fish. Assuming that the ship is moving at 3 knots (the general speed of ships with the LFA device) and ensonifies a stationary fish at 193 dB, the fish would be exposed to the maximum sound level for only a few seconds. The next received exposure of the stationary fish would be at a significantly lower sound level since the ship would have moved on during ensonification and during the quiet period for the sonar.

The results presented here are representative of a worst-case scenario (albeit highly unlikely) of hearing for fish following exposure to LFA sonar. These results represent two extremes of effects—a minimal effect of 5-dB threshold shifts to a large effect of 20-dB threshold shifts at a single frequency that may or may not recover. The temporary exposure to sounds of SURTASS LFA sonar does affect rainbow trout auditory thresholds, and how this deficiency in hearing at one frequency impacts their ability to survive in the wild is unknown. Ultimately, the results showed no mortality associated with sound exposure and all fish appeared healthy and active until the end of our experimental week, when they were sacrificed or returned to the fish farm from which they were purchased.

The rainbow trout exhibited a behavioral response to the onset of the LFA signal, the description of which is presented elsewhere, but included a rapid burst of swimming immediately after the sound onset (Wysocki, private communica-

tion). How such responses may affect a fish's long-term behavior and survival and the potential cumulative effects of repeated exposure are not known.

G. Relevance to other anthropogenic sources

The relevance of these studies to other high-intensity sound sources is tenuous. In particular, the LFA sound is ramped on relatively slowly and consists of FM sweeps and pure tones. In contrast, other anthropogenic sources such as pile driving and seismic air-guns have rapid onsets and may have significantly different effects on fish due to their transient nature. Thus, extrapolation from LFA to other sources should be done with the utmost caution.

At the same time, it may be possible to compare exposures using SEL as a measure of total energy in 1 s of exposure (see Hamernik and Hsueh, 1991; Finneran *et al.*, 2002; Hastings and Popper, 2005). The SEL provides a measure of the total energy in 1 s of a signal, and this is a more appropriate and encompassing comparison between signals of different types (e.g., air-gun versus sonar) than is peak or rms pressure, particularly for very short signals with significant energy peaks. SEL is calculated by summing the cumulative pressure squared (p^2) over time and is expressed as dB re 1 μPa^2 s.

In this study, the 1-s SEL was 189 dB re 1 μPa^2 s. To date, the only other report of SEL for fish was in exposure to a seismic air-gun (Popper *et al.*, 2005), where the maximum SEL to which another salmonid, the broad whitefish (*Coregonus nasus*), was exposed was about 180 dB re 1 μPa^2 s. The maximum peak pressure in that study was around 197 dB re 1 μPa^2, a level considerably above that used in this study, but that signal was of a lower SEL calculation. In that seismic study, exposure to relatively high SEL level signals resulted in little or no physiological effects on fish, suggesting that salmonids are potentially able to withstand higher sound levels than have been tested to date.

It is also hard to compare the exposure to the LFA signal to the effects of generally elevated background noise because of increased shipping and other sources. It is known that increased background noise can have long-term effects on hearing sensitivity in fishes that have auditory specializations (Smith *et al.*, 2004a, b), but unless the sound was exceedingly loud, it may not affect hearing in fishes, such as salmonids, that do not have such specializations (Smith *et al.*, 2004a). However, noise can also affect other aspects of fish physiology, e.g., ship noise has been shown to induce increased cortisol secretion (i.e., increases stress levels) in different species of freshwater fishes regardless of their hearing sensitivity (Wysocki *et al.*, 2006).

H. Extrapolation to other species

As mentioned earlier, the effects of LFA sonar on salmonids are of concern, but most are listed or endangered species. Thus, this study used rainbow trout as a representative salmonid. Among most salmonid species, the ears, lack of auditory specializations, and hearing sensitivity are all very similar to those of rainbow trout.

Extrapolation of these rainbow trout results to other nonsalmonids or fish with hearing specializations is not possible since there are differences in ear structures and hearing sensitivity between various groups. Studies need to be performed on several other species (these are planned for the near future) to even consider extrapolation to other species or groups of fish.

ACKNOWLEDGMENTS

This work was supported by funding from the United States Chief of Naval Operations. This project could not have happened were it not for the immense logistic and technical support provided by numerous people. We are grateful to Kevin White for very creative assistance at the Seneca Lake Facility. Nelson Elias and his staff at Seneca Lake went out of their way to ensure that our every need was met throughout the study and to ensure that the study was done efficiently and with maximum safety. We are grateful to Armen Bahlavouni of Scientific Solutions Inc. (SSI) for design of the test tanks and the sound and video systems and for developing the detailed logistical procedures required to move fish around and run the tests. Abigail Philips and Mark Poston of SSI worked with us at each of the test sessions and provided invaluable electronic support, calibration of the setup, and signal analysis. Clay Spikes of Marine Acoustics Inc. provided significant logistical support and guidance and was critical to ensuring that all parts of the study came together and Geoffrey Sisson offered technical support. We also thank Dr. James Salierno for his field assistance. We are grateful to Dr. Richard Fay, Dr. Dennis Higgs, Dr. David Mann, and Dr. David Zeddies for providing valuable comments on earlier versions of this manuscript.

Finally, we want to thank Joseph Johnson of the office of the Chief of Naval Operations not only for supporting the project but also for leaving the science to the scientists.

Amoser, S., and Ladich, F. (**2003**). "Diversity in noise-induced temporary hearing loss in otophysine fishes," J. Acoust. Soc. Am. **113**, 2170–2179.

Cernak, I., Savic, J., Malicevic, Z., Zunic, G., Radosevic, P., Ivanovic, I., and Davidovic, L. (**1996**). "Involvement of the central nervous system in the general response to pulmonary blast injury," J. Trauma: Inj., Infect., Crit. Care **40**, S100–S104.

Corwin, J. T., Bullock, T. H., and Schweitzer, J. (**1982**). "The auditory brainstem response in five vertebrate classes," Electroencephalogr. Clin. Neurophysiol. **54**, 629–641).

Dodd, T. K. T., Mundie, G., Lagutchik, M. S., and Morris, J. R. (**1997**). "Cardiopulmonary effects of high-impulse noise exposure," J. Trauma: Inj., Infect., Crit. Care **43**, 656–666.

Engås, A., and Løkkeborg, S. (**2002**). "Effects of seismic shooting and vessel-generated noise on fish behaviour and catch rates," Bioacoustics **12**, 313–315.

Enger, P. S. (**1981**). "Frequency discrimination in teleosts—central or peripheral?," in *Hearing and Sound Communication in Fishes*, edited by W. N. Tavolga, A. N. Popper, and R. R. Fay (Springer, New York), pp. 243–255.

Fay, R. R. (**1988**). *Hearing in Vertebrates, A Psychophysics Databook* (Hill-Fay Associates, Winnetka, IL).

Finneran, J. J., Schlundt, C. E., Dear, R., Carder, D. A., and Ridgway, S. H. (**2002**). "Temporary shift in masked hearing thresholds in odontocetes after exposure to single underwater impulses from a seismic watergun," J. Acoust. Soc. Am. **111**, 2929–2940.

Hamernik, R. P., and Hsueh, K. D. (**1991**). "Impulse noise: Some definitions, physical acoustics, and other considerations," J. Acoust. Soc. Am. **90**, 189–196.

Hastings, M. C., and Popper, A. N. (**2005**). "Effects of sound on fish," Report to California Department of Transportation Contract No. 43A0139, Task Order 1, http://www4.trb.org/trb/crp.nsf/reference/boilerplate/Attachments/$file/EffectsOfSoundOnFish1-28-05(FINAL).pdf. Viewed 5/11/07.

Hastings, M. C., Popper, A. N., Finneran, J. J., and Lanford, P. J. (**1996**). "Effect of low frequency underwater sound on hair cells of the inner ear and lateral line of the teleost fish *Astronotus ocellatus*," J. Acoust. Soc. Am. **99**, 1759–1766.

Hattingh, J., and Petty, D. (**1992**). "Comparative physiological responses to stressors in animals," Comp. Biochem. Physiol. A **101A**, 113–116.

Hawkins, A. D., and Johnstone, A. D. F. (**1978**). "The hearing of the Atlantic salmon, *Salmo salar*," J. Fish Biol. **13**, 655–673.

Higgs, D. M., Rollo, A. K., Souza, M. J., and Popper, A. N. (**2003**). "Development of form and function in peripheral auditory structures of the zebrafish (*Danio rerio*)," J. Acoust. Soc. Am. **113**, 1145–1154.

Kane, A. S. (**1996**). *Fish Guts: A Multimedia Guide to the Art and Science of Fish Anatomy, Health and Necropsy*, edited by A. S. Kane (APC Press, Baltimore, MD, **1996**), ⟨http://aquaticpath.umd.edu/fg⟩.

Kane, A. S., Dykstra, M. J., Noga, E. J., Reimschuessel, R., Baya, A., Driscoll, C., Paerl, H. W., and Landsberg, J. (**2000**). "Etiologies, observations and reporting of estuarine finfish lesions," Mar. Environ. Res. **50**, 473–477.

Kenyon, T. N., Ladich, F., and Yan, H. Y. (**1998**). "A comparative study of hearing ability in fishes; the auditory brainstem response approach," J. Comp. Physiol. **182**, 307–318.

Løkkeborg, S., Ona, E., and Soldal, A. V. (**1996**). "Effects of seismic shooting on local abundance and catch rates of cod (*Gadus morhua*) and haddock (*Melanogrammus aeglefinus*)," Can. J. Fish. Aquat. Sci. **53**, 2238–2249.

Mann, D. A., Higgs, D. M., Tavolga, W. N., Souza, M. J., and Popper, A. N. (**2001**). "Ultrasound detection by clupeiform fishes," J. Acoust. Soc. Am. **109**, 3048–3054.

McCauley, R. D., Fewtrell, J., and Popper, A. N. (**2003**). "High intensity anthropogenic sound damages fish ears," J. Acoust. Soc. Am. **113**, 638–642.

Myrberg, A. A., Jr., (**1980**). "Ocean noise and the behavior of marine animals," in *Advanced Concepts in Ocean Measurements for Marine Biology*, edited by F. P. Diemer, F. J. Vernberg, and D. V. Mirkes (University of South Carolina Press, Columbia, SC), pp. 461–491.

National Research Council (**1994**). *Low-Frequency Sound and Marine Mammals: Current Knowledge and Research Need* (National Academy Press, Washington, DC).

National Research Council (**2000**). *Marine Mammals and Low Frequency Sound: Progress Since 1994* (National Academy Press, Washington, DC).

National Research Council (**2003**). *Ocean Noise and Marine Mammals* (National Academy Press, Washington, DC).

Popper, A. N. (**1976**). "Ultrastructure of the auditory regions in the inner ear of the lake whitefish," Science **192**, 1020–1023.

Popper, A. N. (**1977**). "A scanning electron microscopic study of the sacculus and lagena in the ears of fifteen species of teleost fishes," J. Morphol. **153**, 397–418.

Popper, A. N. (**2003**). "Effects of anthropogenic sound on fishes," Fisheries **28**, 24–31.

Popper, A. N., and Clarke, N. L. (**1976**). "The auditory system of the goldfish (*Carassius auratus*): Effects of intense acoustic stimulation," Comp. Biochem. Physiol. A **53A**, 11–18.

Popper, A. N., and Fay, R. R. (**1999**). "The auditory periphery in fishes," in *Comparative Hearing: Fish and Amphibians*, edited by R. R. Fay and A. N. Popper (Springer, New York), pp. 43–100.

Popper, A. N., Fay, R. R., Platt, C., and Sand, O. (**2003**). "Sound detection mechanisms and capabilities of teleost fishes," in *Sensory Processing in Aquatic Environments*, edited by S. P. Collin and N. J. Marshall (Springer, New York), pp. 3–38.

Popper, A. N., Fewtrell, J., Smith, M. E., and McCauley, R. D. (**2004**). "Anthropogenic sound: Effects on the behavior and physiology of fishes," Mar. Technol. Soc. J. **37**, 35–40.

Popper, A. N., Smith, M. E., Cott, P. A., Hanna, B. W., MacGillivray, A. O., Austin, M. E., and Mann, D. A. (**2005**). "Effects of exposure to seismic airgun use on hearing of three fish species," J. Acoust. Soc. Am. **117**, 3958–3971.

Rehulka, J., and Minarik, B. (**2004**). "Effect of polychlorinated biphenyls (Delor 103) on haematological and enzyme parameters of the rainbow trout *Oncorhynchus mykiss*," Dis. Aquat. Org. **62**, 147–53.

Richardson, W. J., Greene, C. R., Jr., Malme, C. L., and Thomson, D. H. (**1995**). *Marine Mammals and Noise* (Academic, New York).

Scholik, A. R., and Yan, H. Y. (**2001**). "Effects of underwater noise on auditory sensitivity of a cyprinid fish," Hear. Res. **152**, 17–24.

Smith, M. E., Kane, A. S., and Popper, A. N. (**2004a**). "Noise-induced stress response and hearing loss in goldfish (*Carassius auratus*)," J. Exp. Biol. **207**, 427–435.

Smith, M. E., Kane, A. S., and Popper, A. N. (**2004b**). "Acoustical stress and hearing sensitivity in fishes: Does the linear threshold shift hypothesis hold water?" J. Exp. Biol. **207**, 3591–3602.

Sverdrup, A., Kjellsby, E., Krüger, P. G., Fløysand, R., Knudsen, F. R., Enger, P. S., Serck-Hanssen, G., and Helle, K. B. (**1994**)."Effects of experimental seismic shock on vasoactivity of arteries, integrity of the vascular endothelium and on primary stress hormones of the Atlantic salmon." J. Fish Biol. **45**, 973–995.

Wartzog, D., Popper, A. N., Gordon, J., and Merrill, J. (**2004**). "Factors affecting the responses of marine mammals to acoustic disturbance," Mar. Technol. Soc. J. **37**, 6–15.

Wysocki, L. E., Dittami, J. P., and Ladich, F. (**2006**). "Ship noise and cortisol secretion in European freshwater fishes." Biol. Conserv. **128**, 501–508.

Wysocki, L. E., and Ladich, F. (**2005**). "Effects of noise exposure on click detection and the temporal resolution ability of the goldfish auditory system." Hear. Res. **201**, 27–36.

Yelverton, J. T., Richmond, D. R., Hicks, W., Saunders, K., and Fletcher, E. R. (**1975**). *The Relationship Between Fish Size and Their Response to Underwater Blast*, Report No. DNA 3677T, Director, Defense Nuclear Agency, Washington, DC.

The dependence of time-domain speed-of-sound measurements on center frequency, bandwidth, and transit-time marker in human calcaneus *in vitro*

Keith A. Wear[a)]
U.S. Food and Drug Administration, Center for Devices and Radiological Health, HFZ-140, 12720 Twinbrook Parkway, Rockville, Maryland 20852

(Received 25 February 2007; revised 7 April 2007; accepted 9 April 2007)

Time-domain speed-of-sound (SOS) measurements in calcaneus are effective predictors of osteoporotic fracture risk. High attenuation and dispersion in bone, however, produce severe distortion of transmitted pulses that leads to ambiguity of time-domain SOS measurements. An equation to predict the effects of system parameters (center frequency and bandwidth), algorithm parameters (pulse arrival-time marker), and bone properties (attenuation coefficient and thickness) on time-domain SOS estimates is derived for media with attenuation that varies linearly with frequency. The equation is validated using data from a bone-mimicking phantom and from 30 human calcaneus samples *in vitro*. The data suggest that the effects of dispersion are small compared with the effects of frequency-dependent attenuation. The equation can be used to retroactively compensate data. System-related variations in SOS are shown to decrease as the pulse-arrival-time marker is moved toward the pulse center. Therefore, compared with other time-domain measures of SOS, group velocity exhibits the minimum system dependence. © *2007 Acoustical Society of America.* [DOI: 10.1121/1.2735811]

PACS number(s): 43.80.Qf, 43.20.Hq [CCC] Pages: 636–644

I. INTRODUCTION

Speed of sound (SOS) in calcaneus is very useful for characterization of bone and diagnosis of osteoporosis (Rossman *et al.*, 1989; Tavakoli and Evans, 1991; Zagzebski *et al.*, 1991; Laugier *et al.*, 1993; Njeh *et al.*, 1996; Bouxsien and Radloff, 1997; Bauer *et al.*, 1997; Laugier *et al.*, 1997; Strelitzki and Evans, 1996; Strelitzki *et al.*, 1996; Strelitzki *et al.*, 1997; Nicholson *et al.*, 1996, 1998; Hans *et al.*, 1999; Trebacz and Natali, 1999; Hoffmeister *et al.*, 2000, 2002b, a; Chaffai *et al.*, 2002; Lee *et al.*, 2003a, b; Laugier, 2004; Hakulinen *et al.*, 2005, Yamoto *et al.*, 2006). A landmark prospective clinical study of 5662 women showed that a linear combination of SOS and broadband ultrasound attenuation (BUA) predicts hip fracture risk in elderly women as well as dual energy x-ray absorptiometry (Hans *et al.*, 1996). Other prospective (Huopio *et al.*, 2004; Krieg *et al.*, 2006) and retrospective (Schott *et al.*, 1995; Turner *et al.*, 1995; Mautalen *et al.*, 1995; Thompson *et al.*, 1998; Krieg *et al.*, 2003; Glüer *et al.*, 2004; Maggi *et al.*, 2006) clinical studies have further established the utility of SOS for fracture risk prediction.

In laboratories and in commercial bone sonometers, SOS is usually measured in through-transmission. The ultrasound pulse transit time through the bone is compared with the transit time through a water path. SOS in bone is then computed from SOS in water, bone thickness, and transit-time differential. In order to measure transit time from digitized radio-frequency wave forms, a marker on the wave forms such as a zero crossing is usually designated. Unfortunately, frequency-dependent attenuation and dispersion can distort wave forms, shift locations of zero crossings, and produce variations in SOS measurements. See Fig. 1. This is particularly troublesome for highly attenuating media such as bone.

Ambiguities in time-domain SOS measurements in cancellous bone were first reported by Laugier *et al.* (1993) and subsequently by Strelitzki and Evans (1996), Nicholson *et al.* (1996), and Wear (2000a). Laugier *et al.* showed that SOS measurements in cancellous bone are highest for markers near the leading edge of the pulse wave form and lowest for markers near the trailing edge. As previously reported (Wear, 2000a) and also shown in Table I, there is a considerable range of choices for markers. Therefore, different investigators using different measurement systems and different analysis algorithms obtain disparate values for SOS, often precluding meaningful comparison and/or pooling of measurements obtained from different studies. This problem may contribute to the substantial disparity in SOS measurements from different commercial bone sonometers reported by Njeh *et al.* (2000). See Fig. 2. Glüer (2007) recently emphasized that improvement in standardization methods will be required before quantitative ultrasound can play a major role in clinical practice.

Velocity measurement methods that do not rely on designated markers may be more resistant to frequency-dependent-attenuation-induced artifacts. One method estimates relative time delay from the cross correlation of the two wave forms. Unfortunately, this method becomes less robust when signals are severely distorted. In bone, substantial distortion arises from frequency-dependent attenuation, dispersion, multiple scattering, and phase cancellation. An-

a)Electronic mail: kaw@cdrh.fda.gov

 0001-4966/2007/122(1)/636/9/$23.00

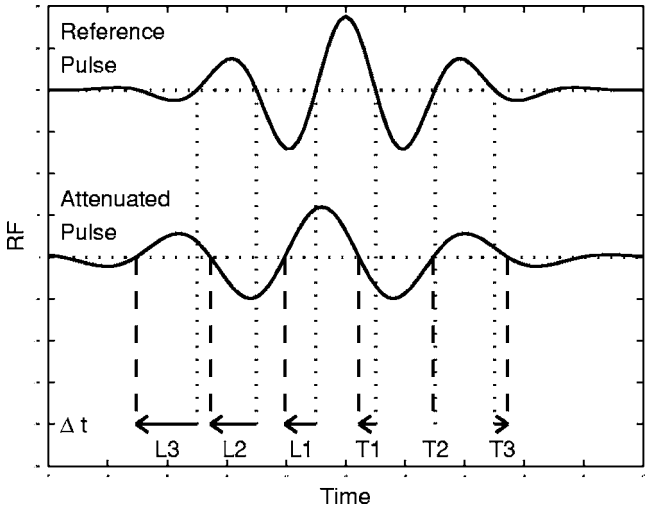


FIG. 1. The effect of frequency-dependent attenuation on transit-time differential, Δt. Markers are labeled with an L for leading half or a T for trailing half and are numbered outward from the pulse center. The variation in Δt with marker location is due to the fact that the attenuated pulse is stretched in time as a consequence of the low-pass filtering effect of frequency-dependent attenuation.

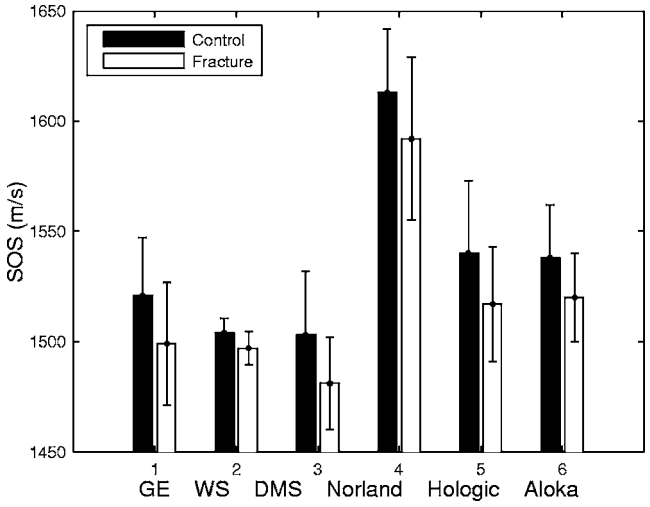


FIG. 2. SOS measurements reported by Njeh *et al.* (2000) in 35 normal postmenopausal women and 35 age-matched women with fractures using six commercial bone sonometers.

other method utilizes phase velocity, which is measured in the frequency domain, rather than a time-domain SOS measure. Although these alternatives may have some advantages, the ultrasonic bone characterization literature is dominated by the marker-time-shift methodology discussed earlier (see Table I). Therefore, it is important to explore sources of variability inherent in these methods so that they may be taken into account when interpreting and comparing existing results. Moreover, although bone sonometer manufacturers generally do not publish their SOS-measurement algorithms in the public domain, it is likely that many commercial systems utilize marker-time-shift SOS measurements due to their prevalence in the literature. (Note that a senior author for two of the references in Table I was the founder of Lunar—now GE Lunar—Corporation, a major manufacturer of clinical bone sonometers.) Therefore, these same sources of variability are likely to affect commercial bone sonometer measurements.

A mathematical model was previously developed to predict variations in SOS estimates for Gaussian pulses propagating through media with weak dispersion and attenuation that varies linearly with frequency (Wear, 2000a, 2001). The model was validated with measurements on 24 human calcaneus samples *in vitro*. It was shown that in human calcaneus *in vitro* the effects of attenuation are far greater than the effects of dispersion. Haïat *et al.* (2006) used an approach based on this model to investigate SOS measurements in 38 human femur samples *in vitro*. Rather than using a fixed value for dispersion, however, they compensated SOS measurements based on individual dispersion measurements. They found this approach to be very accurate in numerically predicting disparity between a zero-crossing-based SOS estimate and group velocity. Their analysis was very thorough in delineating the effects of dispersion and attenuation on SOS estimates in femur.

Although the numerical approach accurately predicts system-related and algorithm-related variations in SOS measurements, it requires substantial effort to implement and is

TABLE I. A nonexhaustive listing of marker designations for SOS measurements in bone.

General marker type	Specific location	References
Leading edge (first detectable deviation from zero)		Njeh *et al.*, 1996, 1997; Nieh and Langton, 1997
Thresholding	At three times the noise standard deviation	Alves *et al.*, 1996a
	At 20% of amplitude of first half cycle	Hakulinen *et al.*, 2005
Zero crossings	First	Nicholson *et al.*, 1998; Trebacz and Natali, 1999; Lee *et al.*, 2003a, b
	"Specific"	Rossman *et al.*, 1989
	Of first negative slope	Zagzebski *et al.*, 1991
	First after threshold of 10%	Haïat *et al.*, 2006
	First after threshold of 15%	Haïat *et al.*, 2005
Maximum absolute value		Alves *et al.*, 1996b.
Maximum Envelope		Wear, 2000b

TABLE II. Locations of zero crossings for the reference and attenuated pulses shown in Fig. 1. The maximum of the reference pulse is halfway between L1 and T1 at $t_0=0$. The period of the reference pulse is $1/f_0$. The period of the attenuated pulse is $1/f_1$.

Zero crossing	L3	L2	L1	T1	T2	T3
Reference pulse (t_0)	$-5/4f_0$	$-3/4f_0$	$-1/4f_0$	$1/4f_0$	$3/4f_0$	$5/4f_0$
Attenuated pulse (t_1)	$-5/4f_1+s$	$-3/4f_1+s$	$-1/4f_1+s$	$1/4f_1+s$	$3/4f_1+s$	$5/4f_1+s$

unlikely to be widely adopted by the research community or by industry. In order to make this problem more tractable, a simpler model, which accounts for frequency-dependent attenuation but not dispersion, is derived in the following. The analysis produces an analytic formula that is easy to use for the adjustment of SOS measurements performed with different measurement systems and different marker locations. The formula provides insight into the effects of system parameters (center frequency and bandwidth), algorithm parameters (marker location), and bone properties (attenuation coefficient and thickness) on time-domain SOS estimates. The formula may be used to retroactively compensate existing data. The model is validated in a bone-mimicking phantom and in 30 human calcaneus samples *in vitro*.

II. THEORY

SOS is usually measured in through transmission. Two transducers, one transmitter and one receiver, are arranged in a "pitch-catch" orientation. First, a calibration measurement is performed by propagating a pulse from the transmitter to the receiver through a water path. Then a sample (e.g., bone) is placed between the two transducers, and a second measurement is performed. SOS may then be computed from

$$\mathrm{SOS} = \frac{c_w}{1 + \frac{c_w \Delta t}{d}}, \tag{1}$$

where c_w is the acoustic velocity in water, Δt is the difference in arrival times of the two pulses, and d is the thickness of the sample. Since the pulse has finite duration, a marker (e.g., the maximum or a zero crossing) is designated for the measurements of arrival times. (Of course, for the sake of consistency, the same marker must be chosen for both calibration and sample measurements.) As illustrated in Fig. 1, however, frequency-dependent attenuation (which is a low-pass filter) stretches the attenuated signal in time and causes Δt to vary with marker designation.

In this section, a formula for the disparity in SOS measurement due to a disparity in marker designation is derived. Suppose that two markers, n and m, yield Δt values Δt_n and Δt_m, then the resulting disparity in SOS estimates is

$$\begin{aligned}\Delta \mathrm{SOS}_{n,m} &= \frac{c_w}{1 + \frac{c_w \Delta t_n}{d}} - \frac{c_w}{1 + \frac{c_w \Delta t_m}{d}} \\ &= \frac{c_w^2}{d} \frac{(\Delta t_m - \Delta t_n)}{\left(1 + \frac{c_w \Delta t_m}{d}\right)\left(1 + \frac{c_w \Delta t_n}{d}\right)}.\end{aligned} \tag{2}$$

In order to derive expressions for Δt_n and Δt_m for various marker designations, a model for time-domain signals undergoing frequency-dependent attenuation is taken from a previous paper (Wear, 2000a). Briefly, the water-path calibration signal, $x(t)$, is assumed to be a Gaussian modulated sinusoid, with analytic signal representation given by

$$x(t) = A e^{-t^2/2\sigma_0^2} e^{i2\pi f_0 t}, \tag{3}$$

where A is the pulse magnitude, σ_0 is a measure of the pulse duration, f_0 is the center frequency, and t is time. (Throughout this section, a subscript of 0 will be used for parameters in the calibration measurement while a subscript of 1 will be used for parameters in the sample—e.g., bone—measurement.) The Fourier transform of $x(t)$ is

$$X(f) = A\sigma_0\sqrt{2\pi} e^{-(f-f_0)^2/2\sigma_f^2}, \tag{4}$$

where f is frequency and $\sigma_f = 1/(2\pi\sigma_0)$ is a measure of the bandwidth. A medium with attenuation that varies linearly with frequency can be modeled with the following transfer function:

$$H(f) = e^{-\beta f d} e^{-i2\pi f s}, \tag{5}$$

where βf is the attenuation coefficient, d is the thickness of the sample, and s is the time delay (relative to a water reference signal) given by

$$s = d\left[\frac{1}{c_b} - \frac{1}{c_w}\right], \tag{6}$$

where c_b is the phase velocity in the linearly attenuating medium. (Note that $s<0$ when $c_b>c_w$). It can be shown that the attenuated signal, $y(t)$, takes the form (Wear, 2000a)

$$y(t) = B e^{-(t-s)^2/2\sigma_1^2} e^{i2\pi f_1 (t-s)}, \tag{7}$$

where B is the pulse magnitude, and σ_1 is a measure of the pulse duration. The downshifted center frequency, f_1, of the attenuated wave is given by (Narayana and Ophir, 1983; Wear, 2000a)

$$f_1 = f_0 - \sigma_f^2 \beta d. \tag{8}$$

Locations of zero crossings are determined by the complex exponential factors in Eqs. (3) and (7). The locations of zero-crossing markers for the symmetric pulses shown in Fig. 1 are given in Table II.

Suppose that a marker on the calibration signal, $x(t)$, is chosen at $t=t_{n0}$. Then, by comparing the complex exponentials in Eqs. (3) and (7), it can be seen that the corresponding marker on the attenuated wave form, t_{n1}, will satisfy $f_0 t_{n0} = f_1(t_{n1}-s)$. Therefore, $t_{n1} = (f_0/f_1)t_{n0}+s$. The transit time differential is then

$$\Delta t_n = t_{n1} - t_{n0} = \left(\frac{f_0}{f_1} - 1\right) t_{n0} + s = a t_{n0} + s, \tag{9}$$

where [using Eq. (8)]

$$a \equiv \frac{f_0}{f_1} - 1 = \frac{f_0}{f_0 - \sigma_f^2 \beta d} - 1 = \frac{\sigma_f^2 \beta d}{f_0 - \sigma_f^2 \beta d}. \tag{10}$$

Equation (2) compares the SOS estimate based on marker n located at t_{n0} with the SOS estimate based on marker m located at t_{m0}. Suppose that the second marker (m) is chosen to be the envelope maximum, so that the second SOS estimate corresponds to group velocity. If the reference pulse is assumed to be centered at the origin, then $t_{m0}=0$, $t_{m1}=s$, and $\Delta t_m = s$. The difference of transit-time differentials for two markers required in the numerator of Eq. (2) is then [from Eq. (9)]

$$\Delta t_m - \Delta t_n = -a t_{n0} = -a \frac{\tau}{f_0}, \tag{11}$$

where τ is the interval between calibration marker n and the envelope maximum, measured in units of the calibration wave form period, $T_0 = 1/f_0$, so that $\tau \equiv t_{n0}/T_0 = f_0 t_{n0}$. For example, the central zero crossings on the symmetric calibration wave form in Fig. 1 occur at $\tau = \ldots, -5/4, -3/4, -1/4, 1/4, 3/4, 5/4, \ldots$.

Equation (11) gives the numerator for the right-hand factor in Eq. (2). Now the denominator for the right-hand factor in Eq. (2) must be obtained. Let c_g denote group velocity. Note from Eq. (1) that

$$c_g = \frac{c_w}{1 + \frac{c_w s}{d}}. \tag{12}$$

Now the denominator from Eq. (2) may be simplified as follows:

$$\left(1 + \frac{c_w s}{d}\right)\left[1 + \frac{c_w}{d}\left(\frac{a\tau}{f_0} + s\right)\right] = \frac{c_w}{c_g}\left[\frac{c_w}{c_g} + \frac{c_w a \tau}{d f_0}\right] = \frac{c_w^2}{c_g^2} + \frac{c_w^2 a \tau}{c_g d f_0}. \tag{13}$$

Combining Eqs. (2), (11), and (13) yields

$$\mathrm{SOS}_n - c_g = -c_g \frac{a\tau}{\frac{f_0 d}{c_g} + a\tau}. \tag{14}$$

In the ultrasonic characterization of bone, the first term in the denominator of Eq. (14) will often dominate the second term. For example, using typical values ($f_0 = 500$ kHz, $\sigma_f = 100$ kHz, normalized broadband ultrasound attenuation or nBUA = 30 dB/cm MHz, $\beta = \mathrm{nBUA}/8.68 = 3.5$ 1/cm MHz, $d = 2$ cm, $c_g = 1500$ m/s $\rightarrow a = 0.16$, and $\tau = -1.25$), the two terms become 6.67 and −0.20, respectively. Neglecting the second term in the denominator, Eq. (14) simplifies to

$$\mathrm{SOS}_n - c_g \approx -\frac{c_g^2 a \tau}{f_0 d} = -\frac{\tau c_g^2 \sigma_f^2 \beta}{f_0^2} \frac{1}{1 - (\sigma_f^2 \beta d / f_0)}. \tag{15}$$

Equation (15) shows that the magnitude of the deviation of the SOS estimate from group velocity has a strong dependence on the fractional bandwidth (σ_f/f_0). The disparity of SOS estimates derived from two arbitrary markers, n and m, separated by $\Delta\tau$, is given by

$$\Delta\mathrm{SOS}_{n,m} \approx -\frac{\Delta\tau c_g^2 \sigma_f^2 \beta}{f_0^2} \frac{1}{1 - (\sigma_f^2 \beta d / f_0)}. \tag{16}$$

III. METHODS

A. Bone samples and phantom

Thirty human calcaneus samples (both genders, ages unknown) were defatted using a trichloro-ethylene solution. Defatting was presumed not to significantly affect measurements since SOS of defatted trabecular bone has been measured to be only slightly different from that of bone with marrow left intact (Njeh and Langton, 1997; Alves *et al.*, 1996a). The cortical lateral layers were sliced off leaving two parallel surfaces with direct access to trabecular bone. Since cortical layers are so thin (a few millimeters), their removal does not alter SOS measurements very much; an experimental investigation of SOS from 20 human calcanea, both before and after cortical plate removal, showed that calcaneal SOS is determined mainly by the cancellous bone component (Njeh and Langton, 1997). The thicknesses of the samples varied from 12 to 21 mm. In order to remove air bubbles, the samples were vacuum degassed underwater in a desiccator. After vacuum, samples were allowed to thermally equilibrate to room temperature prior to ultrasonic interrogation. Ultrasonic measurements were performed in distilled water at room temperature. The temperature was not actively controlled but was measured for each experiment and ranged between 19 and 22 °C. The relative orientation between the ultrasound beam and the calcanea was the same as with *in vivo* measurements performed with commercial bone sonometers, in which sound propagates in the mediolateral (or lateromedial) direction.

A quantitative ultrasound phantom (Model 063, CIRS Inc., Norfolk, VA) was also interrogated. The phantom was composed of proprietary urethane and had a thickness of 36 mm (http://cirsinc.com/063_ultra.html). The attenuation coefficient slope (nBUA) was measured to be 15.0 dB/cm MHz (at 19.3 °C). (The model 063 contains two phantoms labeled "normal" and "osteoporotic." The osteoporotic phantom was measured here.)

B. Ultrasonic measurements

Bone samples and the phantom were interrogated in through-transmission in a water tank using a Panametrics (Waltham, MA) 5800 pulser/receiver with pairs of coaxially aligned Panametrics transducers. See Table III. Received signals were digitized (8 bits; 10 MHz for bone data, 25 MHz

TABLE III. Properties of ultrasound transducers.

Nominal center frequency (MHz)	Diameter (mm)	Focal length (mm)
0.5	25.4	38.1
1.0	25.4	50.8

for phantom data) using a LeCroy (Chestnut Ridge, NY) 9310C Dual 400 MHz oscilloscope and stored on computer (via GPIB) for off-line analysis.

The temperature-dependent speed of sound in distilled water, c_w was used as the reference speed and is the given by (Kaye and Laby, 1973)

$$c_w = 1402.9 + 4.835T - 0.047016T^2 + 0.00012725T^3, \tag{17}$$

where T is the temperature in degrees Celsius. Analysis software was written in MATLAB (Natick, MA). Each arrival time was computed as follows. The mean (dc) value of each received signal was subtracted. The signal was filtered with a frequency domain Gaussian filter with center frequency equal to the transducer nominal center frequency and standard deviation equal to 125 kHz (500 kHz data) or 200 kHz (1 MHz data). The envelope was computed using the Hilbert transform. The three zero crossings immediately before (L1, L2, and L3 in Fig. 1) and after (T1, T2, and T3) the radio-frequency maximum were measured using linear interpolation between points at which a change in sign occurred.

"Uncompensated" SOS values were computed directly using Eq. (1). "Compensated" SOS values were computed by subtracting the predicted SOS difference from group velocity—computed using either Eq. (14) or (15)—from the uncompensated SOS values. Spectral parameters (f_0 and σ_f), which are required for Eqs. (14) and (15), were estimated from Gaussian fits to measured water-path spectra. Attenuation coefficients, which are also required in Eqs. (14) and (15), were computed from linear least-squares fits to log spectral differences versus frequency over the range from 300 to 600 kHz (Narayana and Ophir, 1983; Wear, 2000a).

Generally speaking, this substitution technique can exhibit appreciable error if the speed of sound differs substantially between the sample and the reference (Kaufman *et al.*, 1995). However, one study indicates that this diffraction-related error is negligible in calcaneus (Droin *et al.*, 1998). Apparently, the speed of sound in calcaneus, approximately 1475–1650 m/s, (Droin *et al.*, 1998) is sufficiently close to that of distilled water at room temperature, 1487 m/s [from Eq. (17)] that diffraction-related errors may be ignored.

The effect of system spectral properties (f_0 and σ_f) was investigated by simulating four different systems obtained by applying four different Gaussian filters (instead of the filters described earlier) to the radio-frequency data acquired after transmission through the CIRS bone-mimicking phantom (see Table IV). SOS was measured using three zero crossings from the leading half of the pulse (L1, L2, and L3). Uncompensated measurements were compared with measurements compensated using Eq. (14).

TABLE IV. Four systems were simulated by filtering 500 kHz radio-frequency data with Gaussian filters with the characteristics shown.

System	Filter f_0 (kHz)	Filter σ_f (kHz)	Signal f_0 (kHz)	Signal σ_f (kHz)	Fractional bandwidth (σ_f/f_0) (%)
1	600	100	580	84	15
2	600	135	560	102	18
3	500	150	500	110	22
4	500	200	500	126	25

IV. RESULTS

A. Simulation

Estimates for SOS in Eqs. (14) and (15) were compared to estimates generated using a previously published numerical method (Wear, 2000a) in which (1) a simulated Gaussian pulse is generated, (2) a FFT is applied, (3) the dispersive form of the transfer function in Eq. (5) is applied [See Wear, 2000a, Eqs. (3), (4), and (11)], (4) an inverse FFT is applied, and (5) zero crossing locations are located on the resulting time-domain wave form. For this simulation, the following parameter values were assumed: f_0=500 kHz, σ_f=100 kHz, d=2 cm, and c_g=1550 m/s. Figure 3 shows theoretical [Eq. (14)] and numerical results for disparity between SOS and group velocity evaluated at leading edge zero crossings L1, L2, and L3 for the nondispersive case. There is excellent agreement between theoretical and numerical results. Figure 4 shows theoretical results for disparity between SOS and group velocity using Eq. (14) and its approximate form, Eq. (15). Again, excellent agreement is found.

Figure 5 shows numerical results for the effect of dispersion on the disparity between SOS and group velocity evaluated at leading edge zero crossing L3. Mean *in vitro* calcaneal dispersion rates from four published studies (see Table V) range from −15 to −40 m/s MHz, with an average value of −26.25 m/s MHz. According to Fig. 5, a typical

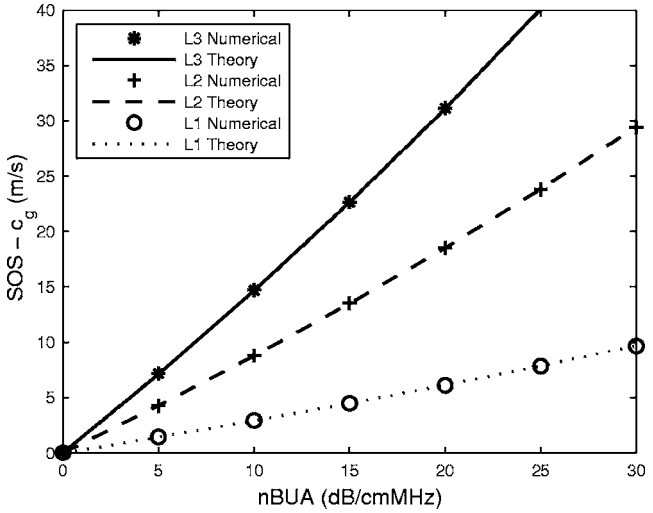


FIG. 3. Theoretical and numerical results for disparity between SOS and group velocity evaluated at leading edge zero crossings L1, L2, and L3. The range of nBUA considered here is similar to that reported from experimental investigations of human calcaneus *in vitro* (Langton *et al.*, 1996).

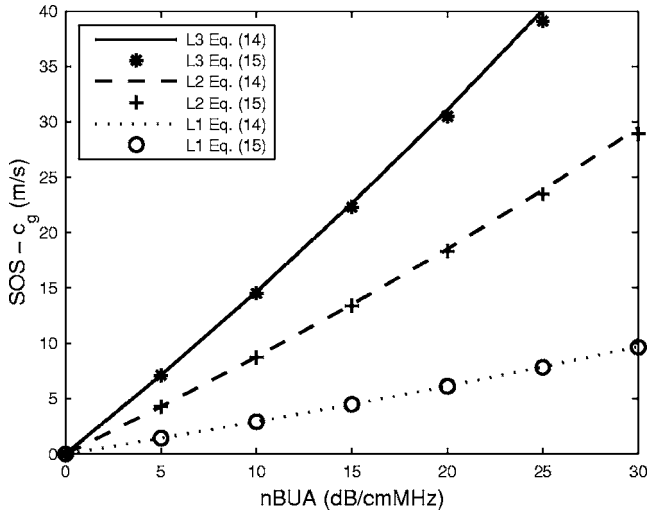


FIG. 4. Theoretical results for disparity between SOS and group velocity using Eq. (14) and its approximate form, Eq. (15).

dispersion rate of −25 m/s MHz has a small effect on the disparity between SOS and group velocity.

B. Experiments

Figure 6 shows the disparity between SOS and group velocity at 500 kHz for 30 bone samples *in vitro* evaluated at zero crossings L3, L2, L1, T1, T2, and T3. The steady decline in uncompensated SOS as marker is moved from the leading edge toward the trailing edge, originally reported by Laugier *et al.* (1993), is evident. The compensation formula [Eq. (14)] effectively suppresses the dependence of SOS on marker location.

Figure 7 shows the disparity between SOS and group velocity at 1 MHz for 24 bone samples *in vitro* evaluated at zero crossings L3, L2, L1, T1, T2, and T3. Again, the compensation formula [Eq. (14)] effectively suppresses the dependence of SOS on marker location.

Figure 8 shows filtered power spectra simulating four different data acquisition systems. Figure 9 shows uncompensated estimates of SOS from the CIRS bone-mimicking phantom evaluated at zero crossings L3, L2, L1 and at the envelope maximum (c_g). Figure 9 shows that system-related variations in SOS decrease as the marker moves toward the pulse center. Relatively subtle changes in system fractional bandwidth can lead to considerable variations in SOS estimate, particularly for markers near the leading edge. Figure 10 shows compensated estimates of SOS from the CIRS bone-mimicking phantom evaluated at zero crossings L3, L2, L1 and at the envelope maximum (c_g). The compensation formula [Eq. (14)] effectively suppresses the dependence of SOS on marker location and system center frequency and bandwidth.

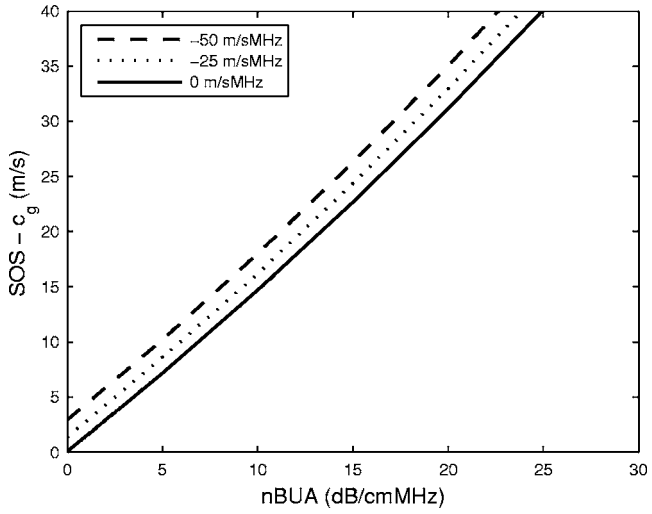


FIG. 5. Effect of dispersion on the disparity between SOS and group velocity evaluated at leading edge zero crossing L3.

TABLE V. Estimates of the first derivative of phase velocity with respect to frequency, dc_p/df, in human calcaneus from Nicholson *et al.* (1996, Table I), Strelitzki and Evans (1996, Table II), Droin *et al.* (1998, Table I), and Wear (2000a, Table I). *N* is the number of calcaneus samples upon which measurements were based.

Author(s)	*N*	Frequency range (kHz)	dc_p/df (mean ± standard deviation) (m/s MHz)
Nicholson *et al.*	70	200–800	−40
Strelitzki and Evans	10	600–800	−32 ± 27
Droin *et al.*	15	200–600	−15 ± 13
Wear	24	200–600	−18 ± 15

V. DISCUSSION

As Njeh *et al.* (2000) recently reported, SOS measurements vary widely among commercial systems (see Fig. 2). Glüer (2007) recently emphasized that improvement in standardization methods will be required before quantitative ultrasound can play a major role in clinical practice. This paper presents a framework for understanding the effects of system parameters (center frequency and bandwidth), algorithm parameters (marker location), and bone properties (attenuation coefficient and thickness) on time-domain SOS estimates. The framework was validated with measurements in a bone-

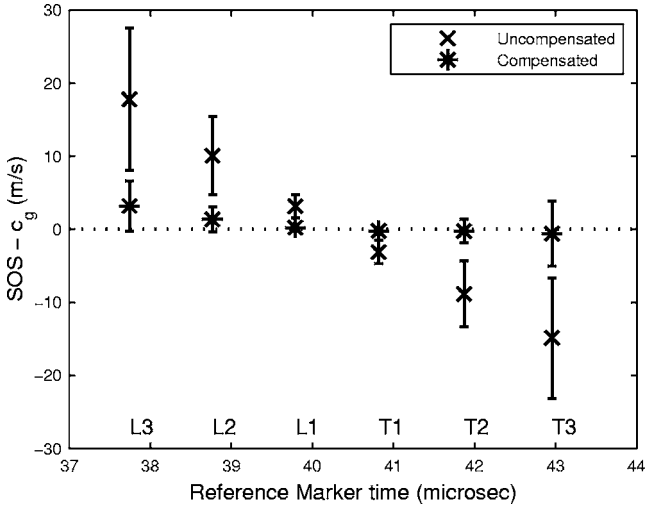


FIG. 6. The disparity between SOS and group velocity at 500 kHz for 30 bone samples *in vitro* evaluated at zero crossings L3, L2, L1, T1, T2, and T3. The compensated measurements were compensated using Eq. (14).

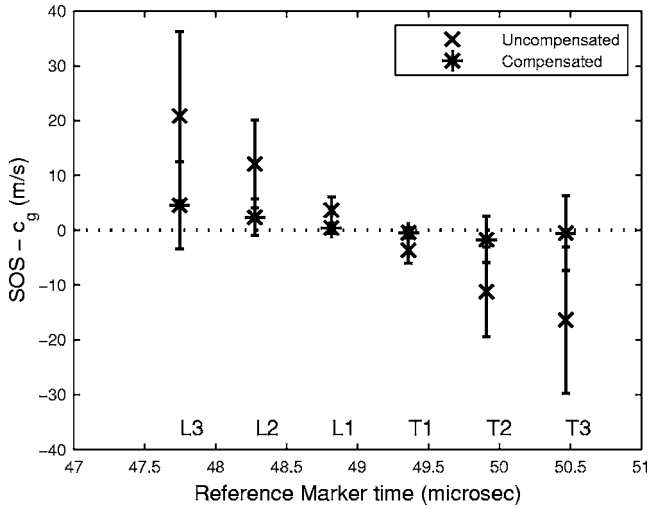


FIG. 7. The disparity between SOS and group velocity at 1 MHz for 24 bone samples *in vitro* evaluated at zero crossings L3, L2, L1, T1, T2, and T3. The compensated measurements were compensated using Eq. (14).

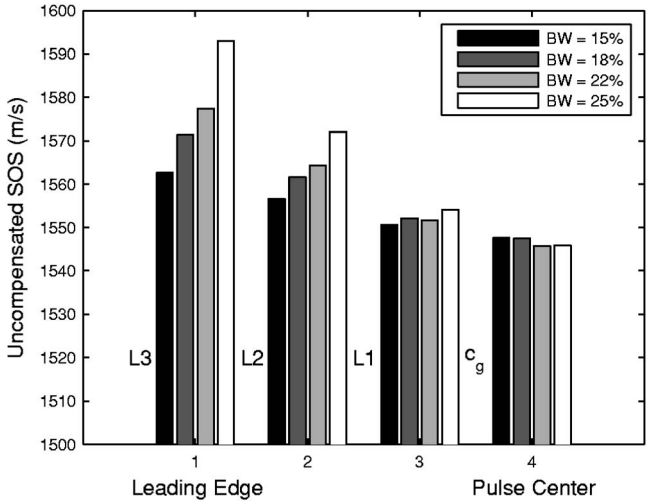


FIG. 9. Estimates of SOS (uncompensated) from the CIRS phantom evaluated at zero crossings L3, L2, L1 and at the envelope maximum (c_g). The standard deviations for all these values are less than 1.5 m/s.

mimicking phantom and in 30 human calcaneus samples *in vitro*. The success of this model suggests that accurate predictions of variations in SOS estimates in human calcaneus *in vitro* may be made even when the effects of dispersion are ignored. Other potential sources of disparity in the SOS measurement—including variations in region of interest location, variations in temperature, assumptions regarding heel thickness—are not considered here.

The variation of SOS with transit-time marker is significant. For example, at 500 kHz, average uncompensated SOS changes by about 8 m/s when transit-time marker shifts from L3 to L2 (see Fig. 6). This is a substantial amount compared with average SOS differences between patients with hip fractures and age-matched controls, which have been reported to be 25 m/s (Schott *et al.*, 1995), 28 m/s (Mautalen *et al.*, 1995), 13.5 m/s (Hans *et al.*, 1996), 20 m/s (Njeh *et al.*, 2000—see Fig. 2), and 20 m/s (Krieg *et al.*, 2006). The variation of SOS with transit-time marker is also substantial when compared to many of the intersystem differences reported by Njeh *et al.* (2000) which are often on the order of 20–40 m/s (see Fig. 2).

Equation (16) provides a means for comparing SOS estimates, SOS_n and SOS_m, based on different marker times separated by $\Delta\tau$. In the case of high $a\tau$, it may be necessary to instead use Eq. (14) twice (first to compute the difference between SOS_n and c_g, and again to compute the difference between SOS_m and c_g) and then to take the difference between the two differences to yield SOS_n-SOS_m. Therefore, data acquired using different systems and computed using different marker times can be compensated to facilitate comparison or pooling provided that the following variables are known: f_0, σ_f, β, d, and c_g. The first two variables may be estimated by performing spectral analysis on a water-path measurement. Attenuation coefficient slope β may be computed from BUA, which is measured by most calcaneal systems, and d, which may be measured *in vitro* or even *in vivo* (Chen *et al.*, 2005; Yi *et al.*, 2007) or if necessary set to an

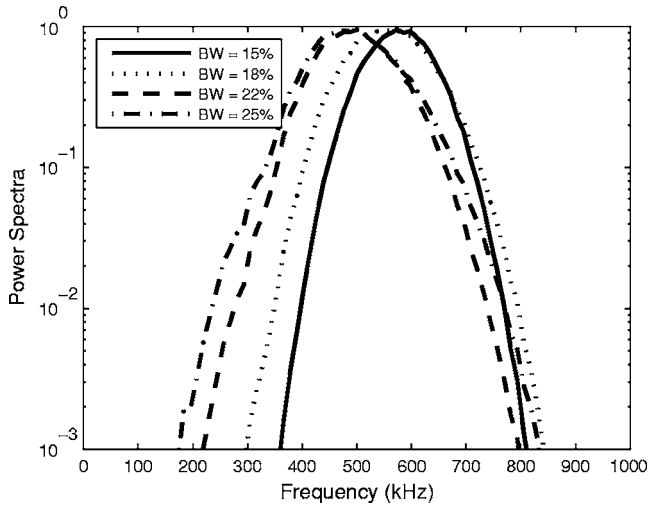


FIG. 8. Filtered power spectra to simulate four different data acquisition systems.

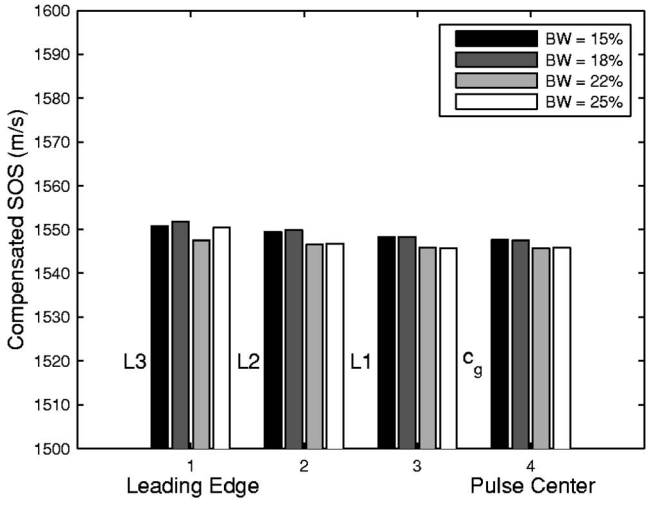


FIG. 10. Estimates of SOS [compensated using Eq. (14)] from the CIRS phantom evaluated at zero crossings L3, L2, L1 and at the envelope maximum (c_g). The standard deviations for all these values are less than 1.5 m/s.

average value [e.g., Nicholson *et al.* (1997) measured an average value of 25 mm from 28 female calcanea]. Finally, although c_g may not be known, a typical value may be assumed. For example, c_g in human calcaneus in vitro has been reported to range from 1475 to 1575 m/s (Wear, 2000b). So a typical value of 1525 m/s may be assumed, producing an error of no more than 7% in Eqs. (14)–(16).

In order to circumvent problems associated with time-domain SOS measures, some authors have suggested the use of phase velocity, which is measured in the frequency domain, to characterize bone (Alves *et al.*, 1996a; Nicholson *et al.*, 1996; Strelitzki and Evans, 1996; Droin *et al.*, 1998). The framework presented in this paper provides a method for retroactively estimating phase velocity from time-domain SOS measurements. First, Eq. (14) may be used to estimate group velocity, c_g. Then, the following relation may be used to estimate phase velocity, c_{pc}, at the center frequency, f_c (Morse and Ingard, 1986; Duck, 1990)

$$c_g = \frac{c_{pc}}{1 - \frac{f_c}{c_{pc}}\left(\frac{\partial c_p}{\partial f}\right)_{f=f_c}}.$$

The quantity in parentheses is the dispersion rate. If independent dispersion measurements are not available, a typical value, such as those listed in Table V may be assumed.

The framework presented here can provide some insight into why many investigators choose markers near the leading edge of the pulse wave form. Attenuation and sound speed tend to increase with density. From Eq. (15), it can be seen that SOS estimates based on markers near the leading edge are boosted by higher attenuation values (recall that τ is negative on the leading half of the pulse). So, for example, in order to increase the sensitivity of the SOS estimate to density, it might be tempting to use a marker near the leading edge. Indeed, Haïat *et al.* (2006) found that SOS measurements based on the first zero crossing have a higher correlation with BMD in femur *in vitro* than group velocity (see Haïat *et al.*, 2006, Table II). Therefore, a marker near the leading edge might be preferable when a single parameter (SOS) is used to assess the bone, as is the case with current cortical bone systems and may be the case with future femur systems. Most calcaneal systems, however, use a linear combination of BUA and SOS to assess the bone. For these systems, the quantitative diagnostic classifier should be able to properly account for attenuation effects directly by optimally adjusting the weighting coefficient for BUA, without the need to enhance SOS measurements by an amount related to BUA. Indeed, Haïat *et al.* (2006) also found that SOS measurements based on the first zero crossing performed comparably to group velocity when either was used *in combination with BUA* in a multiple regression model (again see Haïat *et al.*, 2006, Table II).

As elegantly explored by Haïat *et al.* (2006), zero-crossing-based measures of SOS are prone to discontinuous jumps as BUA or dispersion varies. These jumps are due to changes in the pulse shape that alter reference points along the pulse in such a way to shift the marker locations by one or more zero-crossing intervals. Since group velocity is derived from the envelope, it is more immune to these effects. Haïat *et al.* rightly point out, however, that markers near the leading edge are less sensitive to multipath interference. In the femur (the bone studied by Haïat *et al.*), which has a more complicated shape than the calcaneus, these multipath interference effects may be sufficiently problematic to justify a preference for a leading-edge zero-crossing marker.

Although calcaneal SOS is an effective predictor of fracture risk, its clinical utility is compromised by the high intersystem variability of the measurement. This paper presents a framework for standardization of SOS measurements. System-related variations in SOS decrease as the pulse-arrival-time marker moves toward the pulse center. Therefore, compared with other time-domain measures of calcaneal SOS, group velocity exhibits the minimum system dependence. When SOS is used in combination with BUA in a linear classifier, the choice of group velocity instead of a zero-crossing-based value should not compromise diagnostic performance.

ACKNOWLEDGMENTS

The author is grateful for funding provided by the US Food and Drug Administration Office of Women's Health. The mention of commercial products, their sources, or their use in connection with material reported herein is not to be construed as either an actual or implied endorsement of such products by the Department of Health and Human Services.

Alves, J. M., Ryaby, J. T., Kaufman, J. J., Magee, F. P., and Siffert, R. S. (**1996b**). "Influence of marrow on ultrasonic velocity and attenuation in bovine trabecular bone," Calcif. Tissue Int. **58**, 362–367.

Alves, J. M., Xu, W., Lin, D., Siffert, R. S., Ryaby, J. T., and Kaufmann, J. J. (**1996a**). "Ultrasonic assessment of human and bovine cancellous bone: A comparison study," IEEE Trans. Biomed. Eng. **43**, 249–258.

Bauer, D. C., Gluer, C. C., Cauley, J. A., Vogt, T. M., Ensrud, K. E., Genant, H. K., and Black, D. M. (**1997**). "Broadband ultrasound attenuation predicts fractures strongly and independently of densitometry in older women," Arch. Intern Med. **157**, 629–634.

Bouxsein, M. L., and Radloff, S. E. (**1997**). "Quantitative ultrasound of the calcaneus reflects the mechanical properties of calcaneal trabecular bone," J. Bone Miner. Res. **12**, 839–846.

Chaffai, S., Peyrin, F., Nuzzo, S., Porcher, R., Berger, G., and Laugier, P. (**2002**). "Ultrasonic characterization of human cancellous bone using transmission and backscatter measurements: Relationships to density and microstructure," Bone (N.Y.) **30**, 229–237.

Chen, P., Chen, T., Lu, M., and Yao, W. (**2005**). "The measurements of ultrasound parameters on calcaneus by two-sided interrogation techniques," Meas. Sci. Technol. **16**, 1349–1354.

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).

Glüer, C. C. (**2007**). "Quantitative ultrasound—It is time to focus research efforts," Bone (N.Y.) **40**, 9–13.

Glüer, C. C., Eastell, R., Reid, D. M., Felsenbert, D., Roux, C., Barkmann, R., Timm, W., Blenk, T., Armbrecht, G., Stewart, A., Clowes, J., Thomasius, F. E., and Kolta, S. (**2004**). "Association of five quantitative ultrasound devices and bone densitometry with osteoporotic vertebral fractures in a population-based sample: The OPUS study," J. Bone Miner. Res. **19**, 782–793.

Haïat, G., Padilla, F., Barkmann, R., Kolta, S., Latremouille, C., Glüer, C. C., and Laugier, P. (**2005**). "In vitro speed of sound measurement at intact human femur specimens," Ultrasound Med. Biol. **31**, 987–996.

Haïat, G., Padilla, F., Cleveland, R. O., and Laugier, P. (**2006**). "Effects of frequency-dependent attenuation and velocity dispersion on in vitro ultrasound velocity measurements in intact human femur specimens," IEEE

Trans. Ultrason. Ferroelectr. Freq. Control **53**, 39–51.
Hakulinen, M. A., Day, J. S., Toyras, J., Timonen, M., Kroger, H., Weinans, H., Kiviranta, I., and Jurvelin, J. S. (**2005**). "Prediction of density and mechanical properties of human trabecular bone in vitro by using ultrasound transmission and backscattering measurements at 0.2–6.7 MHz frequency range," Phys. Med. Biol. **50**, 1629–1642.
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., 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 cancellous cubes reflects mainly bone density and elasticity," Calcif. Tissue Int. **64**, 18–23.
Hoffmeister, B. K., Auwarter, J. A., and Rho, J. Y. (**2002a**). "Effect of marrow on the high frequency ultrasonic properties of cancellous bone," Phys. Med. Biol. **47**, 3419–3427.
Hoffmeister, B. K., Whitten, S. A., Kaste, S. C., and Rho, J. Y. (**2002b**). "Effect of collagen and mineral content on the high-frequency ultrasonic properties of human cancellous bone," Osteoporosis Int. **13**, 26–32.
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.
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.
Kaufman, J. J., Xu, W., Chiabrera, A. E., and Siffert, R. S. (**1995**). "Diffraction effects in insertion mode estimation of ultrasonic group velocity," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **42**, 232–242.
Kaye, G. W. C., and Laby, T. H. (**1973**). *Table of Physical and Chemical Constants* (Longman, London).
Krieg, M. A., Cornuz, J., Ruffieux, C., *et al.* (**2003**). "Comparison of three bone ultrasound for the discrimination of subjects with and without osteoporotic fractures among 7562 elderly women," J. Bone Miner. Res. **18**, 1261–1266.
Krieg, M. A., Cornuz, J., Ruffieux, C., *et al.* (**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., Njeh, C. F., Hodgskinson, R., and Currey, J. D. (**1996**). "Prediction of mechanical properties of the human calcaneus by broadband ultrasonic attenuation," Bone (N.Y.) **18**, 495–503.
Laugier, P. (**2004**). "An overview of bone sonometry," International Congress Series **1272**, 23–32.
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.
Laugier, P., Giat, P., Droin, P., Saied, A., and Berger, G. (**1993**). "Ultrasound images of the os calcis: A new method of assessment of bone status," Proc.-IEEE Ultrason. Symp. **2**, 989–992.
Lee, K. L., Roh, H., and Yoon, S. W. (**2003a**). "Correlations between acoustic properties and bone density in bovine cancellous bone from 0.5 to 2 MHz," J. Acoust. Soc. Am. **113**, 2933–2938.
Lee, K. I., Roh, H., and Yoon, S. W. (**2003b**). "Acoustic wave propagation in bovine cancellous bone: Application of the modified Biot-Attenborough model," J. Acoust. Soc. Am. **114**, 2284–2293.
Maggi, S., Noale, M., Gianinini, S., *et al.* (**2006**). "Quantitative heel ultrasound in a population-based study in Italy and its relationship with fracture history: The ESOPO study," Osteoporosis Int. **17**, 237–244.
Mautalen, C., Vega, E., Gonzalez, D., Carrilero, P., Otano, A., and Silberman, F. (**1995**). "Ultrasound and dual x-ray absorptiometry densitometry in women with hip fracture," Calcif. Tissue Int. **57**, 165–168.
Morse, P. M., and Ingard, K. U. (**1986**). *Theoretical Acoustics* (Princeton University Press, Princeton, NJ), Chap. 9.
Narayana, P. A., and Ophir, J. (**1983**). "A closed form method for the measurement of attenuation in nonlinearly dispersive media," Ultrason. Imaging **5**, 17–21.
Nicholson, P. H. F., Lowet, C. 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 cancellous bone," Phys. Med. Biol. **41**, 2421–2435.
Nicholson, P. H. F., Lowet, G., Cheng, X. G., Boonen, S., Van der Perre, G., and Dequeker, J. (**1997**). "Assessment of the strength of the proximal femur in vitro: Relationship with ultrasonic measurements of the calcaneus," Bone (N.Y.) **20**, 219–224.
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., Hans, D., Li, J., Fan, B., Fuerst, T., He, Y. Q., Tsuda-Futami, E., Lu, Y., Wu, C. Y., and Genant, H. K. (**2000**). "Comparison of six calcaneal quantitative ultrasound devices: Precision and hip fracture discrimination," Osteoporosis Int. **11**, 1051–1062.
Njeh, C. F., Hodgskinson, R., Currey, J. D., and Langton, C. M. (**1996**). "Orthogonal relationships between ultrasonic velocity and material properties of bovine cancellous bone," Med. Eng. Phys. **18**, 373–381.
Njeh, C. F., and Langton, C. M. (**1997**). "The effect of cortical endplates on ultrasound velocity through the calcaneus: An in vitro study," Br. J. Radiol. **70**, 504–510.
Njeh, C. F., Kuo, C. W., Langton, C. M., Atrah, H. I., and Boivin, C. M. (**1997**). "Prediction of human femoral bone strength using ultrasound velocity and BMD: An in vitro study," Osteoporosis Int. **7**, 471–477.
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., Weill-Engerer, S., Hans, D., Duboeuf, F., Delmas, P. D., and Meunier, P. J. (**1995**). "Ultrasound discriminates patients with hip fracture equally well as dual energy x-ray absorptiometry and independently of bone mineral density," J. Bone Miner. Res. **10**, 243–249.
Strelitzki, R., Clarke, A. J., and Evans, J. A. (**1996**). "The measurement of the velocity of ultrasound in fixed trabecular bone using broadband pulses and single-frequency tone bursts," Phys. Med. Biol. **41**, 743–753.
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.
Thompson, P., Taylor, J., Fisher, A., and Oliver, R. (**1998**). "Quantitative heel ultrasound in 3180 women between 45 and 75 years of age: Compliance, normal ranges and relationship to fracture history," Osteoporosis Int. **8**, 211–214.
Trebacz, H., and Natali, A. (**1999**). "Ultrasound velocity and attenuation in cancellous bone samples from lumbar vertebra and calcaneus," Osteoporosis Int. **9**, 99–105.
Turner, H., Peacock, M., Timmerman, L., Neal, J. M., and Johnston, C. C., Jr. (**1995**). "Calcaneal ultrasonic measurements discriminate hip fracture independently of bone mass," Osteoporosis Int. **5**, 130–135.
Wear, K. A. (**2000a**). "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. (**2000b**). "Measurements of phase velocity and group velocity in human calcaneus," Ultrasound Med. Biol. **26**, 641–646.
Wear, K. A. (**2001**). "A numerical method to predict the effects of frequency-dependent attenuation and dispersion on speed of sound estimates in cancellous bone," J. Acoust. Soc. Am. **109**, 1213–1218.
Yamoto, Y., Matsukawa, M., Otani, T., Yamazaki, K., and Nagano, A. (**2006**). "Distribution of longitudinal wave properties in bovine cortical bone in vitro," Ultrasonics **44**, e233–e237.
Yi, X., Lin, W., and Qin, Y.-X. (**2007**). "Bone surface topology mapping and its role in trabecular bone quality assessment using scanning confocal ultrasound," Osteo. Int. (in press).
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.

Experimental ultrasound characterization of red blood cell aggregation using the structure factor size estimator

François T. H. Yu[a] and Guy Cloutier[b]
Laboratory of Biorheology and Medical Ultrasonics, University of Montreal Hospital Research Center, Pavilion J.A. de Sève (Room Y-1619), 2099 Alexandre de Sève, Montréal, Québec, H2L 2W5, Canada

(Received 3 October 2006; revised 5 April 2007; accepted 6 April 2007)

The frequency dependence of the ultrasonic backscattering coefficient (BSC) was studied to assess the level of red blood cell (RBC) aggregation. Three monoelement focused wideband transducers were used to insonify porcine blood sheared in a Couette flow from 9 to 30 MHz. A high shear rate was first applied to promote disaggregation. Different residual shear rates were then used to promote formation of RBC aggregates. The structure factor size estimator (SFSE), a second-order data reduction model based on the structure factor, was applied to the frequency-dependent BSC. Two parameters were extracted from the model to describe the level of aggregation at 6% and 40% hematocrits: *W*, the packing factor, and *D* the aggregate diameter, expressed in number of RBCs. Both parameters closely matched theoretical values for nonaggregated RBCs. *W* and *D* increased during aggregation with stabilized values modulated by the applied residual shear rate. Furthermore, parameter *D* during the kinetics of aggregation at 6% hematocrit under static conditions correlated with an optical RBC aggregate size estimation from microscopic images ($r^2=0.76$). To conclude, the SFSE presents an interesting framework for tissue characterization of partially correlated dense tissues such as aggregated RBCs. © *2007 Acoustical Society of America.*
[DOI: 10.1121/1.2735805]

PACS number(s): 43.80.Qf, 43.80.Cs, 43.35.Yb [FD] Pages: 645–656

I. INTRODUCTION

Ultrasonic tissue characterization (UTC) is emerging as a unique noninvasive tool to characterize red blood cell (RBC) aggregation in vivo. UTC is an imaging modality that takes advantage of quantitative ultrasonic parameters (backscattering, attenuation, speed of sound, tissue nonlinearity, statistics) to reveal intrinsic tissue properties. Major advances in this field have been recently accomplished in the areas of prostate cancer diagnosis,[1] cell apoptosis monitoring,[2,3] osteoporosis characterization,[4] rat fibroadenomas and mouse mammary carcinomas characterization,[5,6] and in early Duchenne muscular dystrophy diagnosis.[7] In most cases, the spectral content (integrated backscatter coefficient, spectral slope, *y* intercept, midband fit) of the rf ultrasound signal backscattered by the tissue is used to extract its acoustic properties and to reveal its microstructure and composition.

In the ultrasonic blood characterization (UBC) field, the objective is to obtain quantitative parameters that reflect the aggregation state of blood elements. It is well known that RBCs aggregate to form complex three-dimensional (3D) rouleaux structures depending mainly on an equilibrium between aggregating forces, which are mediated by RBC membrane factors and concentrations of plasmatic high molecular weight proteins such as fibrinogen, and disaggregating forces induced by the shear effects of the flow and the electrostatic and steric interactions between RBCs.[8] This phenomenon is normal, reversible, and it occurs in the circulation of many mammalian species. RBC hyper-aggregation, however, is a pathological state; clinical and epidemiological studies identified it as an independent risk factor of circulatory related disorders such as deep venous thrombosis,[9] atherosclerosis,[10] and diabetes mellitus,[11] to name a few. Because these diseases are characterized by localized blood and blood vessel impairments, this suggests that flow-dependent rheological parameters, such as RBC aggregation, could be involved in their respective pathogenesis. It would thus be of great interest to elucidate the role of RBC aggregation in the etiology of these pathologies *in vivo* and *in situ* with UBC techniques.

One difficulty in UBC resides in the fact that blood is an extremely dense medium (5 million erythrocytes/mm^3) that introduces a nonlinear relationship between the backscattered power and the scatterer concentration. In fact, nonlinear effects become important above 10% hematocrit (*H*, which is the RBC volume concentration), approximately.[12] The physiological hematocrits normally range from 35% to 45%, which is clearly in the nonlinear regime. This particularity is a major difficulty in blood acoustic characterization compared to the other UTC fields, where the number density of scatterers in the targeted tissues rarely exceeds 100 scatterers/mm^3.[13,14] Another major challenge of UBC is to consider clustering particles as RBC aggregates.

The nonlinear hematocrit dependence of the BSC has been thoroughly studied with nonaggregating RBC suspensions. It is well described for Rayleigh scatterers at frequencies up to 90 MHz by the Perkus Yevick packing factor W_{PY},[15,16] and it is supported by quantitative experimental data.[12,17] In the frequency domain, the spectrum of nonaggregating RBCs presents a spectral slope (SS) of 4 (SS is the linear slope of the backscatter as a function of frequency on

[a]Electronic mail: francois.yu@polymtl.ca
[b]Electronic mail: guy.cloutier@umontreal.ca

a log-log scale). When considering aggregating RBCs, an increase in the effective scatterer size (caused by RBC rouleaux formation) or in the insonifying ultrasound (US) frequency (to achieve better resolution) both increase the ka product and restrain the validity of the Rayleigh scattering theory (k is the ultrasonic wave number, and a is the mean radius of individual scatterers or RBCs). Nevertheless, numerous experimental quantification of the effect of RBC aggregation on the backscattering coefficient (BSC) and its frequency dependence can be found in the literature.[18–24] Previously, our group reported different simulation models based on the structure factor, $S(-2k)$, that could explain the frequency dependence of aggregating RBCs in the non-Rayleigh regime.[25–28] It is the purpose here to validate those models with experimental results. The objective of the present study was thus to use an inverse approach of data reduction, using the structure factor size estimator (SFSE), to analyze experimental results obtained from three different wideband focused transducers covering the bandwidth from 9 to 30 MHz. Two parameters that describe the structure of aggregated RBCs were extracted from the second-order Taylor expansion of the structure factor. The first parameter W translates acoustically into variations of the BSC, whereas D reflects reductions of SS from its Rayleigh value of 4.

The theoretical framework behind modeling of the structure factor is first described in Sec. II. Section III explains the experimental setup followed by Sec. IV that gives results on the kinetics of 6% and 40% hematocrit aggregating RBCs, sheared in a Couette flow at residual shear (RS) rates of 0, 2, 10, and 30 s^{-1}. For the static residual shear rate (RS0) at 6% hematocrit, results are compared to aggregate sizes estimated by optical microscopic image segmentation. The validation of the method is discussed in Sec. V.

II. THEORY: THE STRUCTURE FACTOR SIZE ESTIMATOR

Suspended RBCs in a saline solution (no aggregation) can be acoustically considered as a dense collection of partially correlated weak scatterers embedded in a homogeneous medium.[29] Using the Born approximation, the backscattered power is given for Rayleigh scatterers ($ka \ll 1$) by[12]

$$\mathrm{BSC}(-2k) = m\sigma_b(-2k)W, \tag{1}$$

where m is the number density of scatterers, σ_b is the backscattering cross section of a single RBC, W is the packing factor, and k is the wave vector. For suspended RBCs in saline, all parameters in Eq. (1) can be determined analytically. When the hematocrit is known by microcentrifugation, m is given by $m=H/V_s$, where V_s is the volume of a RBC (typically 87 μm^3). The backscattering cross-section σ_b of a Rayleigh scatterer can be estimated by[14]

$$\sigma_b(-2k) = \frac{1}{4\pi^2} V_s^2 \gamma_Z^2 k^4 \left(3\frac{\sin(2ka) - 2ka\cos(2ka)}{(2ka)^3} \right)^2, \tag{2}$$

where γ_Z is the acoustical relative impedance mismatch between the RBC and its suspending medium, hence $\gamma_Z = (Z_{\mathrm{RBC}} - Z_{\mathrm{plasma}})/Z_{\mathrm{plasma}}$ (see Table I). The packing factor W has been extensively studied for nonaggregating scatterers.[30,31] It was shown to reflect the decrease in backscattered power with the increase in spatial correlation among particles with increasing particle crowding. W can be seen as a correction factor that accounts for increasing destructive phase interference (coherent field) introduced by the increase in correlation between particles. The detailed development of this approach can be found in Twersky's paper.[30] In the same manuscript, expressions of the Perkus–Yevick packing factor W_{PY} (hard particles) were proposed for packed spheres ($W_{\mathrm{PY}s}$) and cylinders ($W_{\mathrm{PY}c}$). They are functions of the hematocrit H,

TABLE I. Some physical properties of blood (Ref. 48). The density is denoted ρ, the adiabatic compressibility κ, and the acoustic impedance Z.

	ρ (kg m^{-3})	κ (Pa^{-1})	Z (kg m^{-2} s^{-1})
RBC	1078	3.50×10^{-10}	1.755×10^{6}
Plasma	1021	4.09×10^{-10}	1.580×10^{6}
Relative impedance contrast γ_Z	⋯	⋯	0.11

$$W_{\mathrm{PY}s}(H) = \frac{(1-H)^4}{(1+2H)^2}, \tag{3}$$

$$W_{\mathrm{PY}c}(H) = \frac{(1-H)^3}{(1+H)}. \tag{4}$$

The cylinder prototype [Eq. (4)] provided the best fitting with experimental measurements for suspended RBCs up to physiological hematocrits under Rayleigh conditions.[12] However, this Rayleigh theoretical approach cannot be used when aggregating RBCs are considered and non-Rayleigh scattering occurs. Looking at Eq. (1), one can consider different ways to model the power increase with RBC aggregation. The effect of aggregation can be considered by an increase in the effective scatterer cross section (rouleaux formation) combined to a decrease in the number of scatterers at a constant hematocrit.[32] A comprehensive study of the effect of aggregation on W should then be pursued, as it would not be a simple function of H. A different approach[25–28] introduced the structure factor $S(-2k)$ as an alternative way to model the backscatter power in the non-Rayleigh regime. These last models were based upon the generalized form of Eq. (1), also introduced in Ref. 30, namely:

$$\mathrm{BSC}(-2k) = m\sigma_b(-2k)S(-2k). \tag{5}$$

With this formalism, the number density of scatterers m and the backscattering cross section of a single red cell σ_b remain constant, and changes in backscattered power are entirely caused by variations of the structure factor $S(-2k)$, which is by definition the Fourier transform of the pair correlation function $g(r)$:[30]

$$S(-2k) = 1 + m\int (g(r)-1)e^{-j2kr}dr. \tag{6}$$

In this expression, $g(r)$ represents the probability of finding two particles separated by a distance r (see Figs 9 and 10 of Ref. 26 for a visual interpretation of this parameter). This formula shows that a variation in the microscopic organiza-

tion of particles, and thus of $g(r)$, modifies $S(-2k)$ and therefore also the BSC. It is proper to emphasize here that the low frequency limit of $S(-2k)$ is by definition[30] the packing factor W:

$$S(-2k)|_{k\to 0} = 1 + m\int (g(r)-1)dr = W. \quad (7)$$

Under Rayleigh conditions (nonaggregated RBCs), Eq. (5) therefore directly reduces to Eq. (1).

A second-order Taylor approximation of $S(-2k)$ is proposed here to extract two geometrical parameters from the rf frequency-dependent backscattered US data. The second-order Taylor expansion in k of $S(-2k)$ is of the form:

$$S(-2k) \approx C_0 + (-2k)C_1 + (-2k)^2 C_2, \quad (8)$$

where C_0, C_1, and C_2 are simply the series coefficients. We have shown earlier that $C_0 = S(-2k)|_{k=0}$ is the packing factor W. The second constant C_1 is necessarily equal to zero because of parity (i.e., if the blood sample is turned around 180°, the tissue remains the same). Since $S(-2k)$ is a nondimensional number, the third constant C_2 must be a surface in m^2. This inference is well known in crystallography, using small angle neutron, light, and x-ray scattering to determine polymer and protein radius of gyration. C_2 is negative and $-C_2$ is related to the square of the radius of gyration (R_g) of scatterers. This second-order approximation is valid in the Guinier region,[33] named after the pioneering work of that scientist in x-ray scattering and is discussed in Sec. V. In the current paper, we consider that $C_2 = -R_g^2$, hence

$$S(-2k) \approx W - 4R_g^2 k^2. \quad (9)$$

For identical spherical scatterers of radius a (i.e., individual RBCs), and recombining Eqs.(2), (5), and (9), we find that

$$\mathrm{BSC}(-2k) = \frac{1}{3\pi} H \gamma_Z^2 k^4 a^3 \left(3\frac{\sin(2ka) - 2ka\cos(2ka)}{(2ka)^3}\right)^2 \times (W - 4R_g^2 k^2). \quad (10)$$

By replacing $k = 2\pi f/c$, Eq. (10) becomes a polynomial approximation of the frequency dependence of the backscatter for aggregating RBCs, which can model the decrease in spectral slope observed experimentally in the non-Rayleigh regime. By considering an isotropic 3D aggregate of radius R, R_g is related to R by $R_g = \sqrt{\frac{3}{5}}R$.[33] Hence, $D = \sqrt{\frac{5}{3}}R_g/a$, where D is the isotropic diameter of an aggregate (expressed in number of RBCs).

In the current study, the SFSE was used to achieve data reduction of experimental BSC measurements on aggregating RBCs. Estimated values of W and D were deduced from Eq. (10) by least mean squared polynomial fitting of the BSC as a function of frequency.

III. MATERIALS AND METHODS

A. Blood preparation

Fresh porcine whole blood was obtained from a local slaughter house. It was anti-coagulated with 3 g/L of ethylene diamine tetra acetic acid (EDTA). The buffy coat was removed after centrifugation at 2000g and four 60 mL samples were prepared as described in Table II. These samples were introduced and sheared in a Couette flow system, as in Refs. 23 and 24, where US measurements were performed. All experiments were made at room temperature.

TABLE II. Labeling of the different types of blood samples studied.

Label	Hematocrit (%)	Suspending medium
H6	6	Isotonic saline
T6	6	Porcine plasma
H40	40	Isotonic saline
T40	40	Porcine plasma

B. US setup and normalization algorithm for BSC

The use of focused transducers allows one to overcome the limited signal to noise ratio consequent to increased attenuation in high frequency characterization of blood (>20 MHz), especially at physiological hematocrits. However, it was shown that with a focused transducer, the standard substitution method can yield erroneous BSC results.[34] One normalization technique with focused transducers was suggested using a low density 6% hematocrit RBC suspension, with the condition that the BSC of the reference suspension is known or can be determined.[15] Accordingly, the methodology that follows was utilized to first determine the 6% hematocrit reference $\mathrm{BSC}_{\mathrm{H6}}$, which was then used for assessing the backscattering coefficients of blood with the Couette flow apparatus.

Three broadband transducers were used in these experiments. Their general characteristics are summarized in Table III. The pulse-echo acquisition system was composed of an Avtech pulse generator (model AVB2-TA-C-CRIMA, Ottawa, Canada), a Ritec diplexer (model RDX-6, Warwick, RI), a 10 dB Mitec linear amplifier (model AU-A3-0120, Hauppauge, NY), a Panametric pulser-receiver that was used as a wideband receiver (model 5900 PR, Waltham, MA), and an 8 bit 500 MHz sampling frequency GageScope acquisition board (model 8500CS, Montreal, Canada). The first step consisted in determining the $\mathrm{BSC}_{\mathrm{H6}}$. The H6 sample was stirred in a beaker with a magnetic agitator to avoid sedimentation. One-hundred rf lines were acquired and stored; the procedure was repeated on a stainless steel plane reflector

TABLE III. Characteristics of each transducer used to collect rf US data.

Type (manufacturer)	Center frequency (MHz)	−3 dB bandwidth (MHz)	Transducer radius (mm)	Focal length (mm)
V313-SM (Panametrics, Waltham, MA)	15	9–14[a]	4.5	12
V317-SM (Panametrics)	20	15–26	4.5	12
PVDF (VisualSonics, Toronto, Canada)	35	26–40[b]	1.5	6

[a]The 15 MHz transducer was excited at a lower frequency to cover a different frequency band than the 20 MHz transducer.
[b]For the 35 MHz transducer, according to the Guinier plot of Fig. 11, only the data under 30 MHz were used.

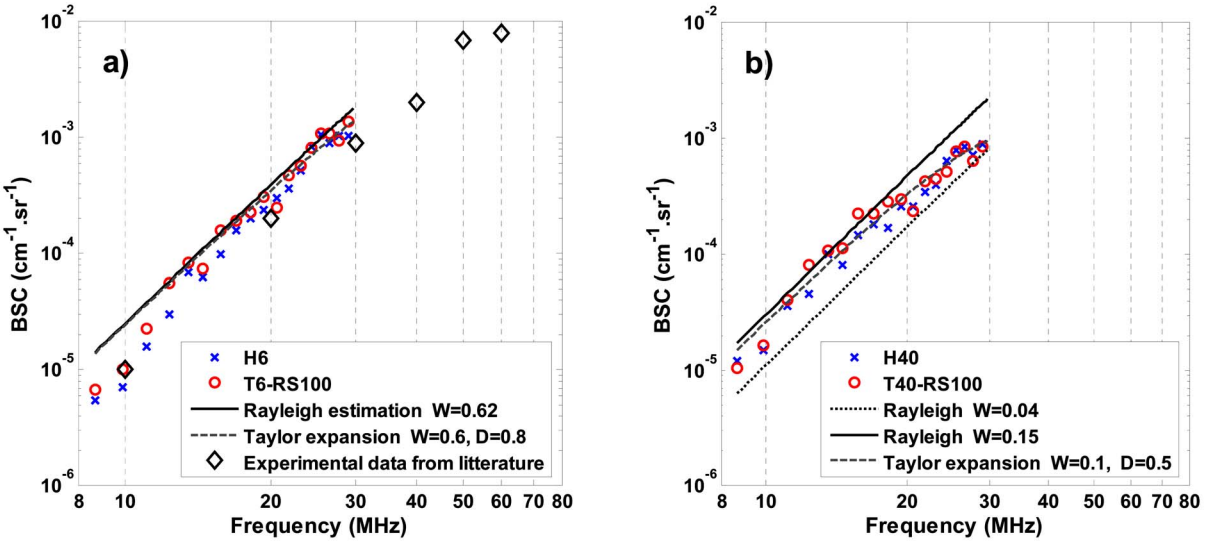


FIG. 1. (Color online) (a) Backscatter coefficient for H6 blood gently stirred in a beaker and T6 blood sheared at 100 s^{-1} in the Couette flow apparatus. The theoretical Rayleigh prediction [Eq. (1)], the second-order Taylor model with $W=0.6$ and $D=0.8$ [Eq. (10)], and H6 experimental data obtained by using nonfocused transducers (Refs. 15 and 16) are also plotted. (b) Backscatter coefficient for H40 blood sheared at 50 s^{-1} and T40 blood sheared at 100 s^{-1} in the Couette flow apparatus. The theoretical Rayleigh prediction with $W=0.04$ and $W=0.15$, and the second-order Taylor model with $W=0.1$ and $D=0.5$ are also plotted (see Table IV). Standard deviations are not shown for clarity.

submerged in degassed water and positioned in the focal plane of each transducer to provide normalization data. A window of 1024 points in each line was selected in the focal zone of the transducer and Fourier transformed; the amplitude was squared to get the power spectrum P of the backscattered signal. These data were then averaged over 100 acquisitions to provide $\overline{P_{\mathrm{H6}B}}$ and $\overline{P_{\mathrm{plane}}}$, where subscripts B and *plane* mean "beaker" and "planar reflector," respectively. The absolute backscatter $\mathrm{BSC_{H6}}$ of Rayleigh diffusers was measured by the substitution method with attenuation compensations, as suggested in[35]

$$\mathrm{BSC_{H6}} = \frac{\overline{P_{\mathrm{H6}B}(f,F)}}{\overline{P_{\mathrm{plane}}(f,F)}} \frac{R_p^2 k^2 r^2}{8\pi d\left[1+\left(\frac{kr^2}{4F}\right)^2\right]} e^{(4\alpha_{\mathrm{H6}}d)}, \qquad (11)$$

where R_p, k, r, d, F, and α_{H6} are, respectively, the reflection coefficient of the planar reflector (assumed to 1), the wave vector, the transducer radius, the inspected depth, the transducer focal length, and the H6 attenuation coefficient. This equation is valid in the focal zone of focused transducers for Rayleigh scatterers.[35] Such is the case for nonaggregated RBCs in our setup in the frequency range 9–30 MHz.[15] Since this step was not performed with nonfocused transducers, the results for $\mathrm{BSC_{H6}}$ are presented in Fig. 1(a) for validation.

Following the beaker and plane reflector measurements, each transducer was then successively placed in the Couette apparatus with its focal plane matching the center of the 2 mm gap between both concentric cylinders filled with blood. An agar gel was used to fill the hole that was made to position the transducer within the static cylinder; the solidified gel was cut to match the curvature of the cylinder in order to minimize any flow disturbance. The nonaggregating H6 sample was filled and gently sheared at 50 s^{-1} while acquiring 100 rf lines for each transducer. The shear rate was precisely controlled by the rotation speed of the moving Couette cylinder. The H6 data provided $\overline{P_{\mathrm{H6}}}$ for the modified substitution method described in Eq. (12). Following these measures, the H6 sample was removed and the Couette apparatus was washed with saline. H40 was then introduced and rotated at the same shear rate of 50 s^{-1} in the Couette system. $\mathrm{BSC_{H40}}$ was computed also using Eq. (12) (with $\overline{P_{\mathrm{blood}}}=\overline{P_{\mathrm{H40}}}$) to provide comparison data for nonaggregating conditions at 40% hematocrit. Then, T6 and T40 whole blood samples were successively introduced in the Couette apparatus. For each aggregation kinetic experiment, the blood was first sheared at 100 s^{-1} for 2 min to disrupt RBC rouleaux. The shear rate was then changed to residual values of 0, 2, 10, and 30 s^{-1} for 3 min. During each experiment, 20 rf lines were acquired every 2 s for 190 s, starting 10 s before the application of the residual shear. At each time instant, a power spectrum was averaged over 20 rf lines to obtain $\overline{P_{\mathrm{blood}}}$ in Eq. (12). The BSC of H6, H40, T6, and T40 samples, sheared in Couette flow, was computed as:

$$\mathrm{BSC_{blood}} = \mathrm{BSC_{H6}} \times \frac{\overline{P_{\mathrm{blood}}}}{\overline{P_{\mathrm{H6}}}} \times e^{4d(\alpha_{\mathrm{blood}}-\alpha_{\mathrm{H6}})}, \qquad (12)$$

where $\mathrm{BSC_{H6}}$ is given by Eq. (11) and α_{blood} is the attenuation coefficient of the investigated samples. Values of $\alpha_{\mathrm{H6}} \approx \alpha_{\mathrm{T6}} = 0.03$ dB/cm/MHz,[15] and $\alpha_{\mathrm{H40}} \approx \alpha_{\mathrm{T40}} = 0.22$ dB/cm/MHz[36,37] were selected for all shear rates.

Finally, since suspended RBCs are Rayleigh scatterers in the range of frequencies considered here,[15,16] theoretical values of $\mathrm{BSC_{H6}}$ and $\mathrm{BSC_{H40}}$ were also computed using Eqs. (1), (2), and (4). This second approach was used to confirm the experimental Couette flow measures of BSC. They are identified as "Rayleigh estimations" in Fig. 1. Using the H6 sample as a reference in the Couette rather than the beaker ensured that the focused beams were rigorously identical for

both spectra in the modified substitution method, as well as providing automatic attenuation compensation for the agar gel.

C. Microscopy analysis of RBC aggregates and segmentation

Aggregation kinetics of whole blood at 6% hematocrit (T6 sample) under static conditions were also quantified using an optical method. Blood was shaken using a pipette by cyclic aspiration and a drop was deposited on a glass lamella. It was covered by a micro lamella for visualization. Images were taken at an optical magnification of 40× (1 pixel $=0.6\times0.6$ μm) at intervals of 15 s during 3 min. This procedure was repeated three times with the same blood sample. A H6 nonaggregating suspension was also imaged for comparison. Grayscale images were processed offline to determine the size of each aggregate. The same algorithm was used on all images. First, a segmentation threshold was determined using images normalized in intensity between [0, 1]. The threshold minimized the intraclass variance of the black and white pixels by using the Otsu method (MATLAB function "Graythresh," version 7.0.1.24704, Natick, MA). From the binary images obtained by thresholding, all elements smaller than 7 pixels in diameter were eliminated since they were smaller than a single RBC (a typical single RBC had a diameter of 12 pixels). The size of each cluster of aggregates was normalized by a 12-pixel-diameter circular prototype to obtain the number of RBCs per aggregate. Following this preprocessing, the histogram of the number of RBCs per aggregate was traced and fitted to an exponential distribution of mean b, the mean number of RBCs per aggregate. Assuming isotropic 2D circular aggregates, as a first approximation, the surface occupied by b RBCs is given by $\pi(D_{\mathrm{OPT}}/2)^2$. An optical equivalent of parameter D, given after Eq. (10), was thus given by

$$D_{\mathrm{OPT}} = \sqrt{\frac{4}{\pi}b}. \tag{13}$$

IV. RESULTS

A. Ultrasonic characterization of RBC suspensions

Figure 1(a) shows BSC obtained with the three transducers on H6, the 6% hematocrit RBC suspension reference medium [Eq. (11)]. The BSC on T6 computed with Eq. (12) at 100 s^{-1} (T6-RS100, where RS means "residual shear") is also presented. The fitted Taylor model expansion [Eq. (10)] and the theoretical Rayleigh estimation [Eq. (1)] are also represented, along with experimental results from Refs. 15 and 16. Figure 1(b) shows BSC obtained on H40, the 40% hematocrit RBC suspension, the theoretical Rayleigh BSC with $W=0.04$ and $W=0.15$ [Eq. (4)] and the Taylor model. Values of W obtained with the Taylor model for suspended nonaggregating RBCs at 6% and 40% hematocrits are very close to the theoretical Perkus–Yevick predictions (see Table IV). In addition, the estimated value of $D=0.78\pm0.09$ for H6 is fairly close to the expected result of 1 RBC/aggregate, whereas the size estimation appears underestimated for H40 ($D=0.5\pm0.1$).

TABLE IV. Experimental and theoretical values of W and D for disaggregated RBCs suspended in an isotonic saline solution at 6% and 40% hematocrits. $W_{\mathrm{PY}s}$ and $W_{\mathrm{PY}c}$ were calculated using Eqs. (3) and (4).

	H6		H40	
	W	D	W	D
Experimental values	0.60±0.03	0.78±0.09	0.10±0.01	0.5±0.1
Theoretical $W_{\mathrm{PY}s}$	0.62	⋯	0.04	⋯
Theoretical $W_{\mathrm{PY}c}$	0.78	⋯	0.15	⋯
Theoretical D	⋯	1	⋯	1

B. Kinetics of RBC aggregation

Figure 2 presents the mean BSC over the respective transducers' bandwidth (see Table III) during the kinetics of rouleaux formation for (a) T6 and (b) T40 blood samples. At both studied hematocrits, the rouleaux formation kinetic profiles had similar shapes. For instance, at all frequencies, BSC first had a low value, when blood was sheared at 100 s^{-1}, which corresponds to the disaggregated state. BSC then gradually increased and stabilized at different levels depending on the applied residual shear rates and on the US frequency. The highest BSC levels were achieved at a residual shear rate of 2 s^{-1} for all experiments (RS2 curves). As expected, higher shearing (RS10 and RS30) partially disrupted RBC aggregates and smaller BSC were thus obtained. Under static conditions (RS0) and for all transducers, the BSC reached an intermediate level near that of RS10. For T40, faster kinetics were observed in the first few seconds at increasing frequencies.

Figure 3 shows the evolution of parameters W and D during the process of aggregate formation. As observed, W and D increased as a function of time for residual shear rates promoting RBC aggregation. The comparison of results at both hematocrits suggests that aggregates formed at 40% hematocrit are smaller in size than those obtained at 6% hematocrit, at each respective residual shear rate. Mean values at the plateau of the kinetic curves for all applied shear rates are summarized in Table V.

C. Frequency dependence of the BSC

Figure 4 shows BSC as a function of frequency for different residual shear rates and at the plateau of the kinetics of aggregation (temporal means between $t=170$ s and $t=180$ s). Standard deviations are not shown for clarity. It can be observed that BSC increases with the level of aggregation promoted by different residual shears. The BSC is Rayleigh at low frequencies and high shear rates, and becomes non-Rayleigh (drop in SS) with decreasing RS. Also represented in Fig. 4 are fitted curves based on the Taylor model. Respective values of W and D were reported earlier in Table V.

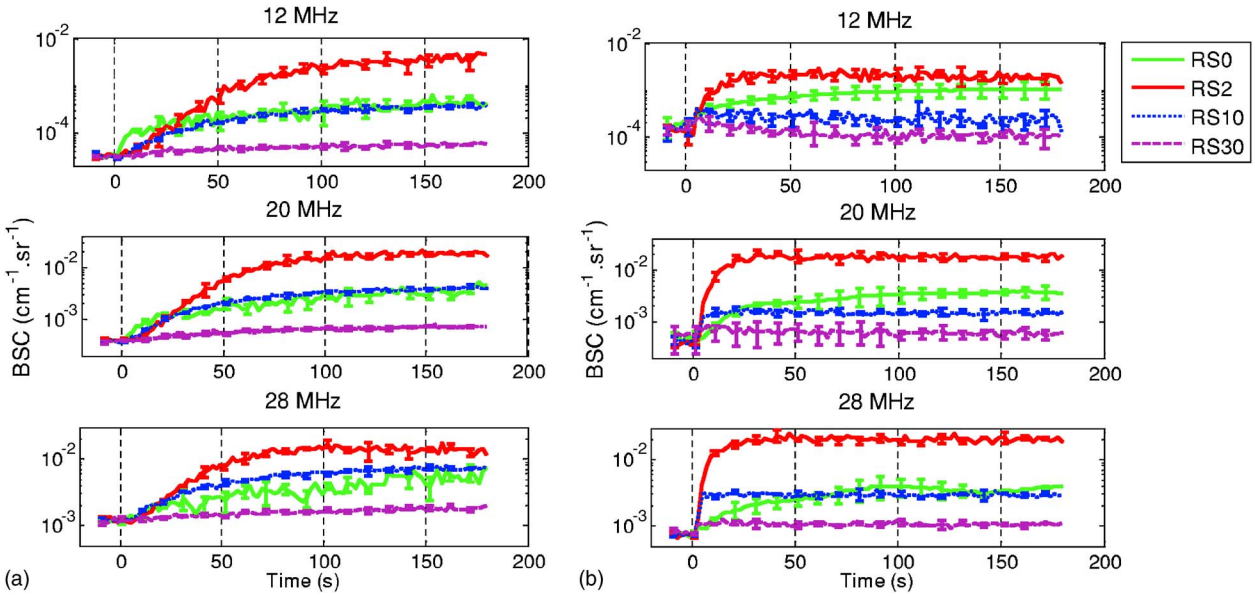


FIG. 2. (Color online) BSC during the kinetics of rouleaux formation for (a) T6 and (b) T40 blood at different residual shear rates of 0, 2, 10, and 30 s^{-1}. A high shear rate of 100 s^{-1} was first applied during the first 10 s (before $t=0$ s). BSC is taken as the mean value over the transducer bandwidth (see Table III). Results are expressed as means ±1 s.d. over three experiments.

D. Comparison of US and optical methods

The US data obtained under static conditions (RS0) are compared in this section with microscopic images acquired at the same hematocrit of 6%. Images were taken with a time resolution of 15 s during 3 min. The first row of Fig. 5 shows typical microscopic images during the aggregation kinetics promoted by Brownian motion. The actual treated images were bigger, only the central 300×400 pixels of the whole 960×1280 pixels are shown in Fig. 5. Binary thresholded images are presented in the second row. Segmented objects are given in the third row. Histograms of the number of RBCs per aggregate, arbitrarily partitioned in 15 logarithmically spaced bins, are displayed in Fig. 6. Each distribution was fitted to an exponential function defined by $P(x)=(1/b)e^{(-x/b)}$, where $b=\text{mean}(x)$ and x is the number of RBCs per aggregate. Parameter b is thus an estimation of the mean number of RBCs per 2D aggregate. D_{OPT} was computed using Eq. (13). Optical and acoustical estimations of D are compared in Table VI. Both methods show an increase of the aggregate size with time under static condition. The linear regression of D_{OPT} with $D_{(\text{US})}$ (see Fig. 7) resulted in an index of correlation $r^2=0.76$, reflecting that both methods followed the dimension of aggregates over time. In Fig. 7, the optical method generally predicted lower values of D.

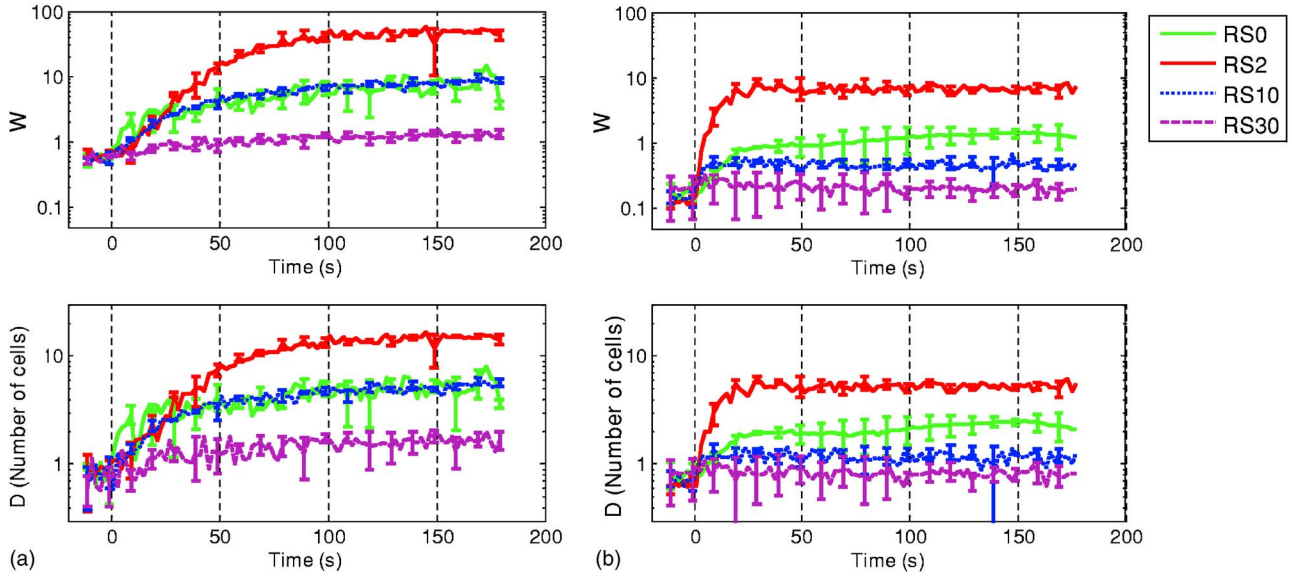


FIG. 3. (Color online) Time variations of fitted parameters W and D during the kinetics of RBC aggregation for experiments with (a) T6 blood and (b) T40 blood at different residual shear rates (RS=0, 2, 10, and 30 s^{-1}). A high shear of 100 s^{-1} was applied during the first 10 s of each acquisition. Results are expressed as means ±1 s.d. over three experiments.

TABLE V. T6 and T40 blood fitting parameters W and D at different residual shear rates (averaged values were computed between $t=170$ and 180 s). RS100 data were averaged between $t=0$ and 10 s. Results are expressed as means ±1 s.d. over three experiments.

Residual shear rate	T6		T40	
	W	D	W	D
RS0	6.8±2.4	4.9±1.4	1.4±0.1	2.4±0.1
RS2	48.9±5.7	15.0±1.0	6.8±1.1	5.3±0.6
RS10	8.4±0.7	5.3±0.4	0.5±0.1	1.1±0.1
RS30	1.3±0.2	1.6±0.3	0.2±0.1	0.7±0.1
RS100	0.60±0.09	0.87±0.27	0.2±0.03	0.8±0.1

V. DISCUSSION

A. On the quality of the data normalization

One concern when pursuing UTC is the quality and reliability of the normalization algorithm. A recent interlaboratory study[38] suggested that improvements in data normalization needed to be pursued, especially for BSC intensity measurements. However, that study also suggested that measurements of backscatter frequency dependent parameters, such as the spectral slope, were more consistent. We therefore tried, in the current study, to take advantage of BSC frequency-dependent information, while emphasizing the validation of the normalization procedure to achieve proper UTC.

First, we used three different transducers to obtain frequency-dependent backscatter measurements. The continuity in frequency can be appreciated in Figs. 1 and 4. Second, values of W obtained with our model for suspended nonaggregating RBCs at 6% and 40% hematocrits are very close to the theoretical Perkus–Yevick predictions. The estimated value of $D=0.78\pm0.09$ for H6 is also fairly close to the expected number of 1 RBC / aggregate, whereas the size estimation seems slightly underestimated for H40 ($D=0.5\pm0.1$, see Table IV). Lastly, as another validation step, one can note that the increases in BSC reported in this study as a function of the level of aggregation are comparable with published data from the literature. For instance, Yuan and Shung[19] reported aggregating pig blood backscattered power as a function of frequency at mean shear rates of 2, 10, and 22 s^{-1} using five unfocused transducers from 3.5 to 12.5 MHz. Hematocrits of 4.5%, 25%, and 45% were investigated in that study. The limitation of the Rayleigh scattering theory could be clearly observed as the spectral slope decreased for strongly aggregating RBCs at 10 MHz and higher, as can be noted with our data (Fig. 4). Our low hematocrit data [Fig. 4(a)] are also in agreement with those shown in Fig. 4 of Ref. 19. In that same study, an increase of 11.5 dB at 12.5 MHz was reported when the shear rate was decreased from 22 to 2 s^{-1} at 45% hematocrit. Our data [T40-RS2—Fig. 2(b)] suggests an increase of 12 dB from 30 to 2 s^{-1} at 12 MHz. In addition, a 13 dB increase at 30 MHz was measured between low and high shearing conditions for whole blood at physiological hematocrits in Refs. 24 and 23. Our data [T40-RS2—Fig. 2(b)] at 28 MHz shows a 14 dB increase with shear rate variations from 100 to 2 s^{-1}. It must be noted, nevertheless, that the former results were obtained on human blood rather than porcine blood. Other experimental studies have also found similar relative increases of the BSC with aggregation.[18,20,21] All these results tend to demonstrate the validity of Eq. (11) (normalization with BSC_{H6}) and of Eq. (12) that were used to obtain BSC_{T6}, BSC_{H40}, and BSC_{T40}.

B. BSC rate of increase during RBC aggregation kinetics

As observed in Fig. 2, the BSC increase in the first few seconds following flow reduction was faster as the frequency was increased; this is especially evident for a hematocrit of 40% and for residual shear rates promoting the largest aggregates. These results confirm previous observations at 40% hematocrit[22] where the rate of increase of the backscattered power as a function of time following flow stoppage was significantly faster at 58 and 36 MHz than at 10 MHz. It was postulated in that study that increasing the ultrasound frequency improved the sensitivity of the method to rapid changes in RBC aggregate sizes. When comparing in Fig. 2 results at 6% and 40% hematocrits, the more rapid increase of BSC at the highest hematocrit is not surprising if one considers that the probability of two RBCs to be in proximity is higher at 40% hematocrit.

C. Frequency-dependent data and second-order Taylor model

The SFSE can be seen as an implementation of form factors used in other tissue characterization fields,[39] adapted to the problem of RBC aggregation. In that sense, it takes into consideration the structure of the tissue rather than the shape of a single scatterer. $S(-2k)=W-4\ R_g^2k^2$ is a decreasing function of k from W to minus infinity.

The parallel is straightforward when illustrated with scarce scatterer concentration: in this case, W is equal to 1 and using $R_g=\sqrt{\frac{3}{5}}R$, then $S(-2k)=1-\frac{12}{5}R^2k^2$. With nonaggregated RBCs and for $f<30$ MHz ($ka<0.33$), $S(-2k)$ behaves similarly to the spherical Gaussian and to the exponential form factor, as shown in Fig. 8 (the notation R is used in this study but the notation a_{eff} is often found in the literature). The structure factor is a decreasing function in ka that translates a variation in scatterer organization but that can have a low frequency limit W different from unity. As Gaussian, fluid sphere, or spherical shell form factors describe the amplitude of the backscattered intensity due to a single scatterer as a function of the frequency, the structure factor describes the effect of aggregation on the backscattered intensity of increasingly organized weak scatterers. The main advantage is that the SFSE can be applied to dense media where the scatterers' positions are partially correlated. If scatterers in the tissue are independently positioned, $W=1$ and the SFSE is very similar to a form factor.

One might wonder why we limited our model to a second-order Taylor expansion in Eq. (10). If it is mathematically true that a higher polynomial order would provide a better fitting with the experimental data, we found that the physical pertinence of the model was lost when a higher

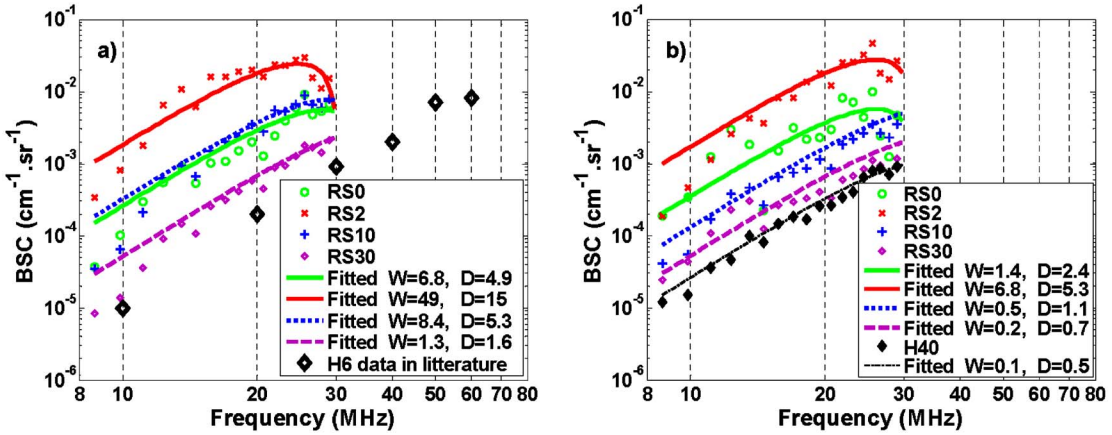


FIG. 4. (Color online) Frequency dependencies of (a) T6 and (b) T40 blood samples sheared at different residual shear rates (RS=0, 2, 10, and 30 s^{-1}), and corresponding fitted models. H6 experimental data taken from Refs. 15 and 16 are also displayed. The standard deviations are not shown for clarity.

order was used. For instance, the second-order model allowed the estimation of W and D, which have a physical meaning.

Also, one might question the physical meaning of $W>1$. In fact, Eqs. (3) and (4) are function of H and are strictly limited to $0<W\leqslant 1$. It is important to emphasize that the above equations were derived for nonaggregating particles. Moreover, our results for nonaggregated RBCs were consistent with these equations. As mentioned earlier, the packing factor W can be generalized for aggregating particles as the low frequency limit of the structure factor. This concept has been used earlier in the field of colloidal suspensions. The Baxter sticky hard sphere model was used to describe the effect of aggregation within the Perkus–Yevick approximation of the structure factor and its low frequency limit W.[40] The adhesive sphere model predicts values of W that decrease from 1 to 0 with increasing hematocrits for low adhesion energy, but presents values of $W>1$ when the adhesion energy is increased. Values of W up to 50 were reported at a volumic fraction of 12%. The structure factor at a frequency of 0 Hz [i.e., $S(0)=W$] is thermodynamically described in this model as $W=mKT\xi_T$ with K being the Boltzmann's constant, T the absolute temperature, and ξ_T the isothermal compressibility, as in Twersky's work.[30] It is also interesting to remark that an increase of W with aggregation is also consistent with an increase in variability of the number of scatterers per elemental scattering voxel, as suggested in Ref. 41.

D. Validation

There is no known way for the authors to experimentally characterize the real size of 3D RBC aggregates at a physiological hematocrit and hence to validate the dimensions reported here. For instance, microscopic observations can only be done with diluted blood; this is why we limited the optical validation to a hematocrit of 6%. Although the correlation between the US and optical methods was good ($r^2=0.76$), it is clear that the comparison has limitations. We can consider the 2D nature of the wet mounts for microscopy versus the 3D structure of real aggregates sensed with US, the isotropic nature of the estimators versus the anisotropic shape of real rouleaux networks, and the limitation of the optical method that only considered a static condition. However, some quantitative information could be deduced with our second-order model. The estimated parameters W and D were modulated in a predictable manner by the applied residual shear rate: Higher shear rates produced smaller values of those measures due to the reversible disruption of RBC clusters. The fact that aggregates were smaller under static conditions is

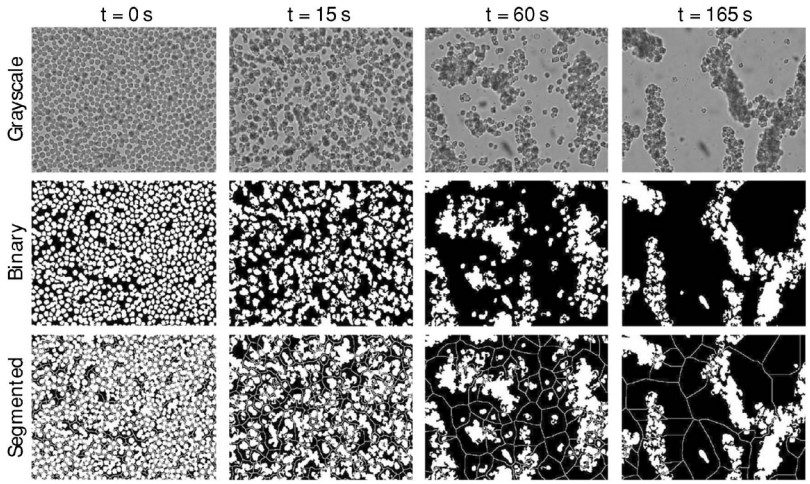


FIG. 5. Columns represent typical microscopic images and image processing at a particular time during the kinetics of aggregation of a T6 sample. Rows represent grayscale images, binary images, and segmented images. Actual processed images (960×1280 pixels) were bigger than those represented here (300×400 pixels). Resolution is 1 pixel=0.6 μm×0.6 μm.

TABLE VI. Estimations of D with the optical and ultrasonic methods. Microscopic images were acquired with a time resolution of 15 s. A time resolution of 30 s was used here to reduce the size of the table. US data taken from Fig. 3 (RS0, H=6%) at corresponding times were used for comparison.

Time (s)	0	30	60	90	120	150	180
D_{OPT}	1.5±0.3	2.4±0.3	3.1±0.5	3.5±0.4	4.2±0.2	4.5±0.1	4.7±0.1
D_{US}	0.8±0.1	3.9±0.1	3.8±0.6	5.3±0.4	4.9±1.0	4.8±1.0	6.2±1.1

not surprising. Others have reported that a minimal level of shear is necessary to promote aggregation, as it increases RBC interactions.[42–44] Further increasing the shear rate results in the breaking of rouleaux. Furthermore, our model showed smaller values of W and D at 40% vs 6% hematocrit. This observation was made earlier in Ref. 45, where the radius of gyration of concentrated cluster suspensions was studied. Stronger aggregate interpenetrations at high hematocrit were hypothesized in that study to explain this observation. Another group reported size estimations from computerized image analysis of RBC aggregates in a small flow chamber.[46] With a diluted 10% hematocrit blood, they measured aggregates of about 6–12 RBCs/aggregate at stasis ($\sim$0 dyn/cm^2) and of about 24–48 RBCs/aggregate at a low shear stress of $\sim$0.1 dyn/cm^2 (corresponding to a shear rate of 10 s^{-1} considering plasma as a Newtonian fluid of viscosity=1 cP).

E. Respective effect of W and D on the second-order Taylor model, relation $(D)^2 \rightarrow W$

To help interpret the acoustical meaning of W and D and the relation between these two physical parameters, we first present in Fig. 9 for T6-RS100 the respective effect of varying W and D on the BSC; we arbitrarily selected a shear rate of 100 s^{-1}, the same interpretation applies to the other shear rates. An increase in W has the effect of increasing the amplitude of the BSC at all frequencies. In terms of standard UTC measures, it can be seen as an analogue to the mean BSC.[47] Increasing D has an effect on the frequency dependence and it thus modulates the spectral slope SS.

Another interesting aspect of parameters W and D, apart from their physical interpretability ("packing factor" and "normalized mean aggregate size"), is the fact that they are not independent from one another. In Fig. 10 are plotted W as a function of $(D)^2$ for all blood samples during the entire kinetics of aggregation for 6% and 40% hematocrits and for all applied shear rates. It can be clearly observed that a quadratic relation exists between W and D. Although this observation could have led us to eliminate one parameter in our data reduction model, we chose not to, as D brings a physical dimension and W is a parameter largely studied in UBC. The fact that the 40% hematocrit scales in the same way as the 6% hematocrit brings further confidence in the second-order approximation model. The relation between both parameters should be further investigated in future studies.

F. Guinier plot

One convenient way to determine the higher frequency limit of our second-order approximation is to use the Guinier plot. This plot is frequently used in x-ray, neutron, and light scattering characterization. It consists in plotting $k^2 \rightarrow \log S(-2k)$. Since we are dealing with a second-order model, the quadratic approximation is valid when the slope is linear in the so-called Guinier region. In Fig. 11 are plotted Guinier results for T6 and T40 blood samples at a residual shear of 2 s^{-1}. The frequency correspondence of the abscissa is indicated by arrows in Fig. 11. Very interesting information can be deduced from Fig. 11. First, the T6-RS2 Guinier plot diverges from a linear slope at frequencies higher than

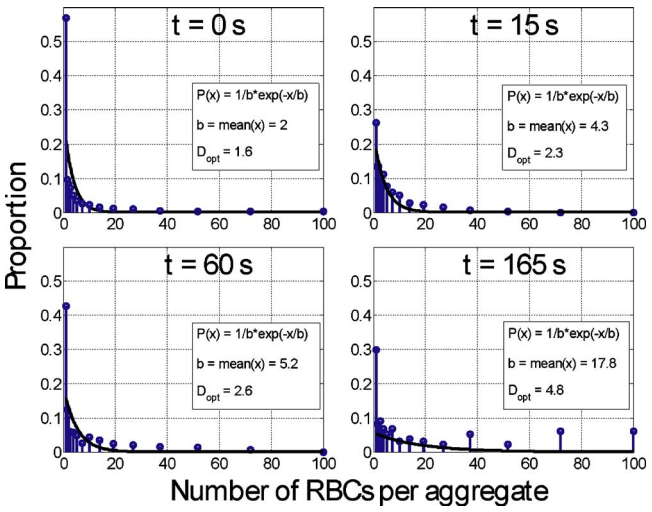


FIG. 6. (Color online) Histograms of the aggregate dimension in number of cells at times 0, 15, 60, and 165 s. Cell count was arbitrarily separated in 15 logarithmically spaced bins to allow better resolution for small sizes. The distributions were fitted by an exponential function of mean b defined by $P(x)=(1/b)\exp^{(-x/b)}$. D_{opt} was determined using Eq. (13).

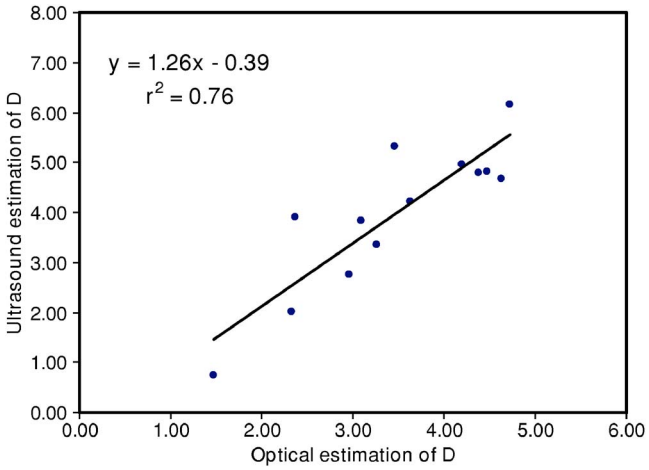


FIG. 7. (Color online) Comparison of D estimated with the optical and US methods. Each point is the mean over three experiments at a particular time during the kinetics of aggregation.

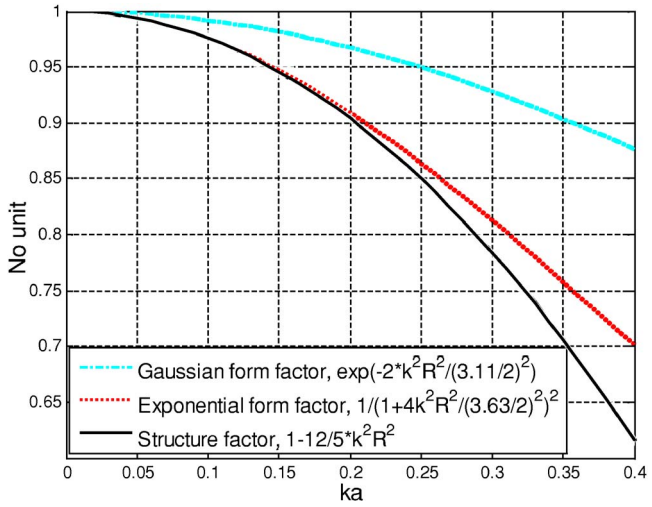


FIG. 8. (Color online) Spherical Gaussian form factor, exponential form factor, and the second-order Taylor expansion of the structure factor as a function of ka, for diluted scatterers ($W=1$). R, the effective radius of the scatterer, is often denoted as a_{eff} in the literature.

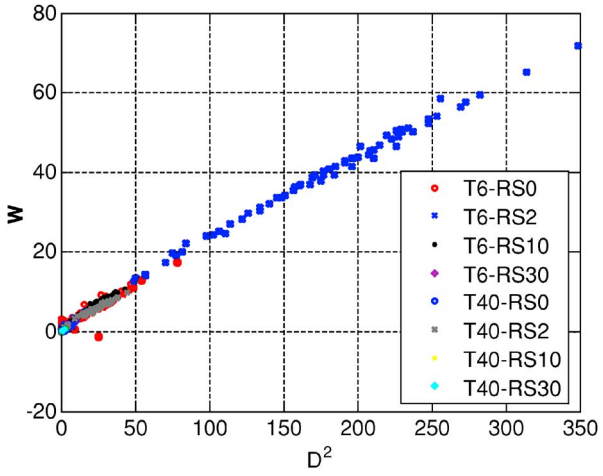


FIG. 10. (Color online) Quadratic relationship between D and W for all shear rates at both 6% and 40% whole blood hematocrits. The high frequency limit was fixed at 30 MHz.

33 MHz. This observation signifies that the second-order approximation is no longer valid for big aggregates at frequencies above 33 MHz. For this reason, we limited our second-order approximation to 30 MHz (see Sec. III). The same cannot be said about the Guinier plot of T40-RS2 that was still linear through 45 MHz. This observation also supports the preceding observation that aggregates formed at 40% hematocrit were smaller that those at 6% hematocrit, for any given residual shear rate.

VI. CONCLUSION

Experimental and theoretical results on BSC variations during the kinetics of RBC aggregation determined by modulating the shear rate in a Couette flow apparatus were presented. Different samples of whole blood and RBC suspensions at 6% and 40% hematocrits were prepared and quantitatively characterized at room temperature using three focused wideband transducers covering the bandwidth from 9 to 30 MHz. A second-order Taylor approximation of the structure factor, the SFSE, was proposed to achieve data reduction of the BSC measurements, to extract two physical parameters, namely the packing factor (W) and the mean normalized dimension of isotropic aggregates (D). The D parameter was validated by an optical imaging method at 6% hematocrit under static conditions. The SFSE presents an interesting framework for ultrasonic characterization of partially correlated tissues, such as aggregating blood. For instance, it more closely mimics the frequency dependence of BSC and it is physically interpretable in a non-frequency-dependent manner.

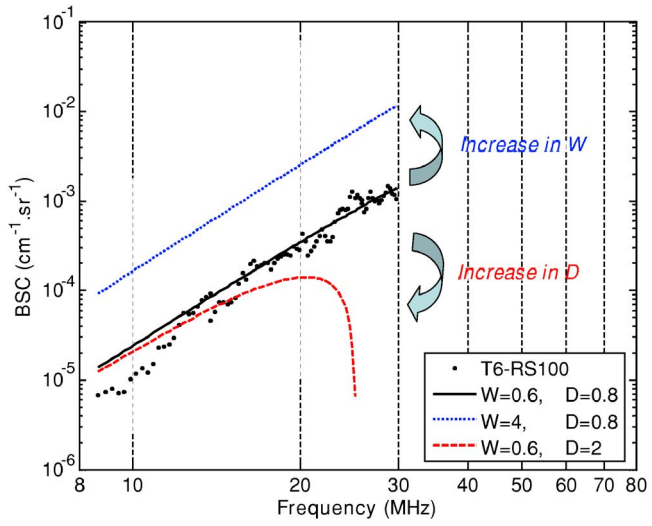


FIG. 9. (Color online) Effect of increasing W and D on the BSC in the second-order model.

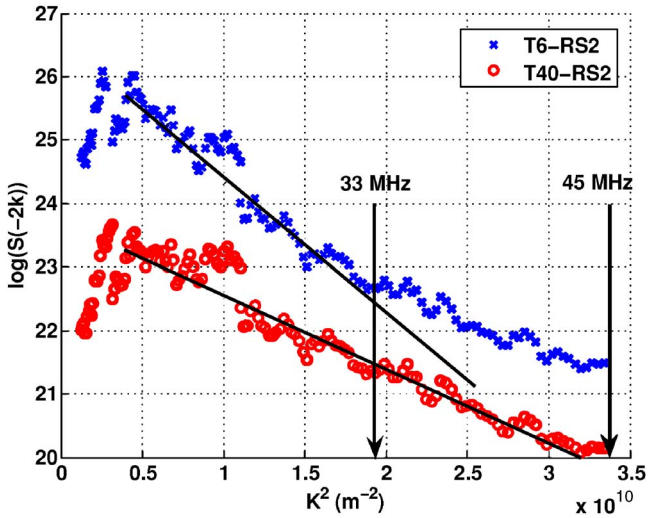


FIG. 11. (Color online) Guinier plots extended to 45 MHz for T6-RS2 and T40-RS2. The linear slope transition indicates the limit of the Guinier domain.

ACKNOWLEDGMENTS

This work was supported by Grant No. MOP-36467 from the Canadian Institutes of Health Research. The salary of F.T.H.Y. is partially supported by a Ph.D. studentship from the Natural Sciences and Engineering Research Council of Canada, and G.C. is recipient of the National Scientist Award from the "Fonds de la Recherche en Santé du Québec." We are also thankful to G. Leblond for his help in the segmentation of the microscopic images, and "Les Viandes ULTRA Meats Inc.," Saint-Esprit, Québec, Canada for blood supply.

[1]E. J. Feleppa, C. R. Porter, J. Ketterling, P. Lee, S. Dasgupta, S. Urban, and A. Kalisz, "Recent developments in tissue-type imaging (TTI) for planning and monitoring treatment of prostate cancer," Ultrason. Imaging **26**, 163–172 (2004).

[2]M. C. Kolios, G. J. Czarnota, M. Lee, J. W. Hunt, and M. D. Sherar, "Ultrasonic spectral parameter characterization of apoptosis," Ultrasound Med. Biol. **28**, 589–597 (2002).

[3]A. S. Tunis, G. J. Czarnota, A. Giles, M. D. Sherar, J. W. Hunt, and M. C. Kolios, "Monitoring structural changes in cells with high-frequency ultrasound signal statistics," Ultrasound Med. Biol. **31**, 1041–1049 (2005).

[4]F. Padilla, L. Akrout, S. Kolta, C. Latremouille, C. Roux, and P. Laugier, "In vitro ultrasound measurement at the human femur," Calcif. Tissue Int. **75**, 421–430 (2004).

[5]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).

[6]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).

[7]V. Giglio, V. Pasceri, L. Messano, F. Mangiola, L. Pasquini, A. Dello Russo, A. Damiani, M. Mirabella, G. Galluzzi, P. Tonali, and E. Ricci, "Ultrasound tissue characterization detects preclinical myocardial structural changes in children affected by Duchenne muscular dystrophy," J. Am. Coll. Cardiol. **42**, 309–316 (2003).

[8]S. Chien, "Biophysical behavior of red cells in suspensions," in *The Red Blood Cell*, edited by D. M. Surgenor (Academic, New York, 1975), Vol. II, Chap. 26, pp. 1031–1133.

[9]E. Alt, S. Banyai, M. Banyai, and R. Koppensteiner, "Blood rheology in deep venous thrombosis—Relation to persistent and transient risk factors," Thromb. Res. **107**, 101–107 (2002).

[10]A. J. Lee, P. I. Mowbray, G. D. O. Lowe, A. Rumley, F. G. R. Fowkes, and P. L. Allan, "Blood viscosity and elevated carotid imtima-media thickness in men and women. The Edinburgh artery study," Circulation **97**, 1467–1473 (1998).

[11]C. Le Devehat, M. Vimeux, and T. Khodabandehlou, "Blood rheology in patients with diabetes mellitus," Clin. Hemorheol Microcirc **30**, 297–300 (2004).

[12]K. K. Shung, "On the ultrasound scattering from blood as a function of hematocrit," IEEE Trans. Sonics Ultrason. **SU-29**, 327–331 (1982).

[13]F. L. Lizzi, M. Ostromogilsky, E. J. Feleppa, M. Rorke, and M. M. Yaremko, "Relationship of ultrasonic spectral parameters to features of tissue microstructure," Ultrasonics **33**, 319–329 (1986).

[14]M. F. Insana, R. F. Wagner, D. G. Brown, and T. J. Hall, "Describing small-scale structure in random media using pulse-echo ultrasound," J. Acoust. Soc. Am. **87**, 179–192 (1990).

[15]S. H. Wang and K. K. Shung, "An approach for measuring ultrasonic backscattering from biological tissues with focused transducers," IEEE Trans. Biomed. Eng. **44**, 549–554 (1997).

[16]S. Maruvada, K. K. Shung, and S. Wang, "High-frequency backscatterer and attenuation measurements of porcine erythrocyte suspensions between 30–90 MHz," Ultrasound Med. Biol. **28**, 1081–1088 (2002).

[17]K. K. Shung, R. A. Sigelmann, and J. M. Reid, "Scattering of ultrasound by blood," IEEE Trans. Biomed. Eng. **BME-23**, 460–467 (1976).

[18]G. Cloutier, Z. Qin, L. G. Durand, and B. G. Teh, "Power Doppler ultrasound evaluation of the shear rate and shear stress dependences of red blood cell aggregation," IEEE Trans. Biomed. Eng. **43**, 441–450 (1996).

[19]Y. W. Yuan and K. K. Shung, "Ultrasonic backscatter from flowing whole blood. II. Dependence on frequency and fibrinogen concentration," J. Acoust. Soc. Am. **84**, 1195–1200 (1988).

[20]V. Rouffiac, P. Péronneau, A. Hadengue, A. Barbet, P. Delouche, P. Dantan, N. Lassau, and J. Levenson, "A new ultrasound principle for characterizing erythrocyte aggregation—In vitro reproducibility and validation," Invest. Radiol. **37**, 413–420 (2002).

[21]F. Yu, D. Savery, A. Amararene, F. S. Foster, and G. Cloutier, "Attenuation compensated spectral slopes during the kinetics of rouleaux formation for porcine whole blood in couette flow at 10–70 MHz," IEEE UFFC Proceedings, 2004, pp. 842–845.

[22]G. Cloutier, M. Daronat, D. Savery, D. Garcia, L. G. Durand, and F. S. Foster, "Non-Gaussian statistics and temporal variations of the ultrasound signal backscattered by blood at frequencies between 10 and 58 MHz," J. Acoust. Soc. Am. **116**, 566–577 (2004).

[23]F. S. Foster, H. Obara, T. Bloomfield, L. K. Ryan, and G. R. Lockwood, "Ultrasound backscatter from blood in the 30 to 70 MHz frequency range," Proc.-IEEE Ultrason. Symp. 1599–1602 (1994).

[24]M. S. Van Der Heiden, M. G. M. De Kroon, N. Bom, and C. Borst, "Ultrasound backscatter at 30 MHz from human blood: Influence of rouleau size affected by blood modification and shear rate," Ultrasound Med. Biol. **21**, 817–826 (1995).

[25]I. Fontaine, M. Bertrand, and G. Cloutier, "A system-based approach to modeling the ultrasound signal backscattered by red blood cells," Biophys. J. **77**, 2387–2399 (1999).

[26]I. Fontaine, D. Savéry, and G. Cloutier, "Simulation of ultrasound backscattering by red cell aggregates: Effect of shear rate and anisotropy," Biophys. J. **82**, 1696–1710 (2002).

[27]I. Fontaine and G. Cloutier, "Modeling the frequency dependence (5–120 MHz) of ultrasound backscattering by red cell aggregates in shear flow at a normal hematocrit," J. Acoust. Soc. Am. **113**, 2893–2900 (2003).

[28]D. Savery and G. Cloutier, "Effect of red cell clustering and anisotropy on ultrasound blood backscatter: A Monte Carlo study," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **52**, 94–103 (2005).

[29]L. Y. L. Mo and R. S. C. Cobbold, "Theoretical models of ultrasonic scattering in blood," in *Ultrasonic Scattering in Biological Tissues*, edited by K. K. Shung and G. A. Thieme (CRC Press, Boca Raton, FL, 1993) Chap. 5, pp. 125–170.

[30]V. Twersky, "Low-frequency scattering by correlated distributions of randomly oriented particles," J. Acoust. Soc. Am. **81**, 1609–1618 (1987).

[31]V. Twersky, "Transparency of pair-correlated, random distributions of small scatterers, with application to the cornea," J. Opt. Soc. Am. **65**, 524–530 (1975).

[32]B. G. Teh and G. Cloutier, "Modeling and analysis of ultrasound backscattering by spherical aggregates and rouleaux of red blood cells," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **47**, 1025–1035 (2000).

[33]A. Guinier and J. Fournet, *Small Angle Scattering of X-Rays* (Wiley Interscience, New York, 1955).

[34]Y. W. Yuan and K. K. Shung, "The effect of focusing on ultrasonic backscatter measurements," Ultrason. Imaging **8**, 121–130 (1986).

[35]M. Ueda and Y. Ozawa, "Spectral analysis of echos for backscattering coefficient measurement," J. Acoust. Soc. Am. **77**, 38–47 (1985).

[36]J. F. Greenleaf, *Tissue Characterization with Ultrasound I*, 1st ed (CRC Press, Boca Raton, FL, 1986), Vol. 1.

[37]J. F. Greenleaf, *Tissue Characterization with Ultrasound II*, 1st ed. (CRC Press, Boca Raton, FL, 1986), Vol. 2.

[38]K. A. Wear, T. A. Stiles, G. R. Frank, E. L. Madsen, F. Cheng, E. J. Feleppa, C. S. Hall, B. S. Kim, P. Lee, W. D. O'Brien, Jr., M. L. Oelze, B. I. Raju, K. K. Shung, T. A. Wilson, and J. R. Yuan, "Interlaboratory comparison of ultrasonic backscatter coefficient measurements from 2 to 9 MHz," J. Ultrasound Med. **24**, 1235–1250 (2005).

[39]K. K. Shung and G. A. Thieme, *Ultrasonic Scattering in Biological Tissues*, edited by K. K. Shung and G. A. Thieme (CRC Press, Boca Raton, 1993).

[40]C. Regnaut and J. C. Ravey, "Application of the adhesive sphere model to the structure of colloidal suspensions," J. Chem. Phys. **91**, 1211–1221 (1989).

[41]L. Y. L. Mo and R. S. C. Cobbold, "A stochastic model of the backscattered Doppler ultrasound from blood," IEEE Trans. Biomed. Eng. **BME-33**, 20–27 (1986).

[42]H. Schmid-Schönbein, P. Gaehtgens, and H. Hirsch, "On the shear rate dependence of red cell aggregation in vitro," J. Clin. Invest. **47**, 1447–1454 (1968).

[43]R. E. N. Shehada, R. S. C. Cobbold, and L. Y. L. Mo, "Aggregation effects in whole blood: Influence of time and shear rate measured using ultrasound," Biorheology **31**, 115–135 (1994).

[44]Z. Qin, L. G. Durand, and G. Cloutier, "Kinetics of the 'black hole' phenomenon in ultrasound backscattering measurements with red blood cell aggregation," Ultrasound Med. Biol. **24**, 245–256 (1998).
[45]P. Snabre and P. Mills, "Rheology of weakly flocculated suspension of rigid particles," J. Phys. III **6**, 1811–1834 (1996).
[46]S. Chen, G. Barshtein, B. Gavish, Y. Mahler, and S. Yedgar, "Monitoring of red blod cell aggregability in a flow-chamber by computerized image analysis," Clin. Hemorheol. **14**, 497–508 (1994).
[47]F. L. Lizzi, M. Greenebaum, E. J. Feleppa, M. Elbaum, and D. J. Coleman, "Theoretical framework for spectrum analysis in ultrasonic tissue characterization," J. Acoust. Soc. Am. **73**, 1366–1373 (1983).
[48]Y. W. Yuan and K. K. Shung, "Ultrasonic backscatter from flowing whole blood. I: Dependence on shear rate and hematocrit," J. Acoust. Soc. Am. **84**, 52–58 (1988).

Boundary element model for simulating sound propagation and source localization within the lungs

M. B. Ozer
Baxter Healthcare Corporation, Deerfield, Illinois 60015

S. Acikgoz and T. J. Royston[a)]
University of Illinois at Chicago, Chicago, Illinois 60607

H. A. Mansy and R. H. Sandler
Rush Medical University, Chicago, Illinois 60612

(Received 19 April 2006; revised 14 February 2007; accepted 15 February 2007)

An acoustic boundary element (BE) model is used to simulate sound propagation in the lung parenchyma. It is computationally validated and then compared with experimental studies on lung phantom models. Parametric studies quantify the effect of different model parameters on the resulting acoustic field within the lung phantoms. The BE model is then coupled with a source localization algorithm to predict the position of an acoustic source within the phantom. Experimental studies validate the BE-based source localization algorithm and show that the same algorithm does not perform as well if the BE simulation is replaced with a free field assumption that neglects reflections and standing wave patterns created within the finite-size lung phantom. The BE model and source localization procedure are then applied to actual lung geometry taken from the National Library of Medicine's Visible Human Project. These numerical studies are in agreement with the studies on simpler geometry in that use of a BE model in place of the free field assumption alters the predicted acoustic field and source localization results. This work is relevant to the development of advanced auscultatory techniques that utilize multiple noninvasive sensors to construct acoustic images of sound generation and transmission to identify pathologies. © *2007 Acoustical Society of America.* [DOI: 10.1121/1.2715453]

PACS number(s): 43.80.Qf, 43.80.Vj, 43.80.Cs [FD] Pages: 657–671

I. INTRODUCTION

A. Background

Passive listening (auscultation) has been used *qualitatively* by physicians for hundreds of years to aid in the monitoring and diagnosis of a wide range of medical conditions, including those involving the pulmonary system (breath sounds), the cardiovascular system (e.g., heart sounds and bruits caused by partially occluded arteries and arteriovenous grafts) and the gastrointestinal system. There may be unique and diagnostically important information in audible frequency sound since characteristic times for many physiological processes and anatomical structural resonances are in that range.[1] This approach offers several potential advantages including noninvasiveness, safety, availability, prompt results, and low cost, making it suitable for in-office check-ups, outpatient home monitoring, the emergency room, and field operations following natural or man-made catastrophes. Simple stethoscopic use is skill-dependent, provides qualitative rather than quantitative information at only a single location, and suffers from inherent limitations of human ability to discern certain acoustic differences. In recent years, many researchers have applied more quantitative measurement and analysis techniques to increase the diagnostic utility of this approach, utilizing electronic sensors and applying computational signal processing and statistical analyses to the measured signals to discern trends or biases correlated with pathologies.[2–14]

To reap the full potential of the inherently rich source of diagnostic information within the audible frequency regime will require a better fundamental understanding of: (1) the acoustic source and its relation to pathology, (2) the acoustic path from the source to the sensor, which can be far more complex at sonic than ultrasonic frequencies due to the potential for multiple reflections, multiple propagating wave types, and multipath behavior, and (3) the use of more accurate and multiple measurement sensors. (4) It could also require more sophisticated and spatially resolved computational processing of the measured signals that considers multipath propagation of the acoustic event from its source to the sensor location to reconstruct a *sonic image* ingrained with quantitative information.

Alterations in the structure and function of the pulmonary system that occur in disease or injury often give rise to measurable changes in lung sound production and transmission. Lung sounds are known to contain spatial information that can be accessed using simultaneous acoustic measurements at multiple locations. It has been shown that lung consolidation, pneumothorax, and airway obstruction, to name a few conditions, alter the production and/or transmission of sound with spectrally and regionally differing effects that, if properly quantified, might provide additional information

[a)]Author to whom correspondence should be addressed. Electronic mail: troyston@uic.edu

about the severity and location of the trauma or pathology.[7–10] Indeed, simultaneous, multisensor auscultation methods have been developed to "map" sounds on the thoracic surface by several groups.[9,11–14]

Beyond this mapping process, Kompis *et al.*[10] attempted to form a three-dimensional (3D) acoustic image of the likely sound source(s) location(s) by using multiple sensors and assuming "ray acoustic," i.e., "free field," models for how sound propagated away from these sources. In their study, they noted that a useful imaging system for the human lung should: (1) be robust with respect to acoustic properties, especially speed of sound which varies and is not precisely known; (2) provide 3D data sets and resulting images that are intuitively interpretable; and (3) be robust with respect to missing sensors or noisy data in individual sensors. The algorithm employed assumed spherically symmetric sound radiation away from a source and propagation throughout the thorax without reflections or standing waves, only dependent on sound speed and a per unit length damping factor; it then triangulated with redundancy on the likely sources of the sound. In five human subject studies the algorithm indicated differences in acoustic source location between inspiration and expiration and in one case, the presence of severe consolidation in the lower left lung. Parametric computer simulation studies showed that their imaging algorithm lacked some robustness with respect to acoustic properties, specifically speed of sound. In their idealized mechanical phantom study the algorithm demonstrated a resolution of nominally 2 cm.

In modeling the transmission of sound throughout the pulmonary system and chest region, the system may be viewed as having two main components: (1) transmission in air through the tracheobronchial tree and (2) coupling to and transmission through the surrounding biological tissues to reach the chest surface—namely the parenchyma, free air, or water/blood region (in the case of a pneumothorax or hydro/hemothorax), surrounding muscle and rib cage regions, and outer soft tissues. Many studies have focused on the transmission of sound in the respiratory tract, the tracheobronchial airway tree, with some also considering coupling to modes of wave propagation in the parenchyma. See Royston *et al.*[15] for a review of this literature. The focus of the present study is related to the second component, transmission through the surrounding biological tissues to reach the chest surface. Previous studies of this part of the problem have assumed simplified geometries and homogenized material properties.[15–17] Wodicka *et al.*[16] assumed an axisymmetric cylindrical geometry, with the outer tissue regions of the chest treated simply as a mass load on the parenchyma. In Vovk *et al.*[17] an axisymmetric layered model for the torso region is used that includes annular regions for the parenchyma, rib cage region, soft outer tissue and skin. In previous work of the authors[15] simplifications of both airway and tissue structures were imposed that resulted in an axisymmetric assumption or two-dimensional planar model assumption that could be easily handled with finite element analysis.

For frequencies above 100 Hz and neglecting the larger segments of the bronchial airway tree, it has been proposed that parenchymal tissue can be modeled as a homogenous isotropic material supporting acoustic compression waves only, i.e., a lossy fluid.[16,18] The parenchymal region is defined by its density ρ_p and a complex wave number k_p, whose real part is linked to phase speed and whose imaginary part defines attenuation. Typical property values provided in the literature are given in Table I and used to generate Fig. 1. Theoretical values of Wodicka *et al.*[16] are based

TABLE I. Acoustic properties of foam phantom and human lung tissue.

	Density (kg/m^3)	Phase speed (m/s)	Attenuation in 10 cm (dB)
Foam at:[a]			
100 Hz	160	27.5	0.053
220 Hz	160	27.5	0.56
460 Hz	160	27.5	5.1
700 Hz	160	27.5	18
Lung theory at[b]			
100 Hz	250	23	0.16
220 Hz	250	23	1.7
460 Hz	250	23	15
700 Hz	250	23	54
Lung model at[c]			
100 Hz	300	32	4.3
220 Hz	300	32	9.4
460 Hz	300	32	20
700 Hz	300	32	30
Lung experiment at[d] 70–140 Hz	⋯	22–24	⋯
Lung experiment at[e] 125–500 Hz	⋯	23–33	⋯
Lung experiment at[f] 75–2000 Hz	⋯	62–82	⋯

[a]Experimentally determined by authors.
[b]Theory of Wodicka *et al.* (Ref. 16).
[c]Empirical model of Vovk *et al.* (Ref. 17).
[d]Experiments of Paciej *et al.* (Ref. 14). (Sound introduced at supraclavicular space, measured on thoracic surface.)
[e]Experiments of Kraman (Ref. 42). (Sound introduced through mouth, measured on thoracic surface.)
[f]Experiments of Mahagnah and Gavriely (Ref. 43). (Sound introduced through mouth, measured on thoracic surface.)

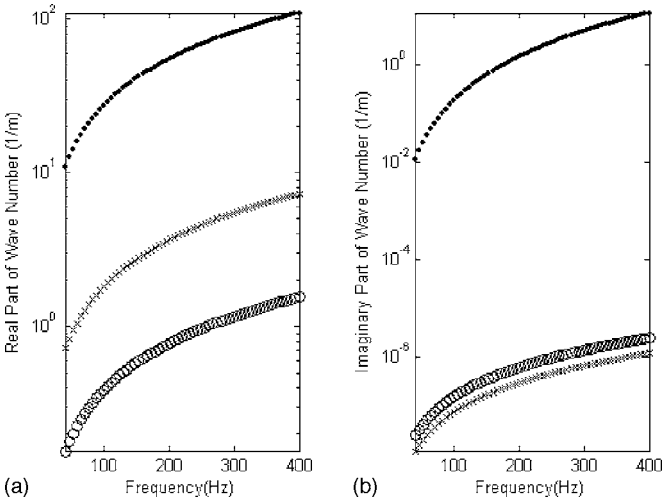


FIG. 1. (a) Real and (b) imaginary parts of complex wave number for: lung parenchyma (Ref. 16) (●); soft tissue (Ref. 15) (○); and air (×).

on calculating effective elasticity, viscosity and density by assuming the parenchyma can be approximated as a closed cell porous material, and by knowing the individual properties of the solid tissue and air that comprise it. Compression wave numbers for the parenchymal material are quite different from those of the two components of which it is comprised, soft tissue and air; correspondingly, in the lung parenchyma sound travels much slower and attenuates rapidly.

To model acoustic compression wave propagation in the surrounding tissues of the torso, the soft tissue regions composed of fat, muscle, and connective/visceral material, can be defined by a density ρ_t and a complex wave number k_t, with $\rho_t \approx 1\ \mathrm{g/cm^3}$, and the real and imaginary parts of the wave number shown in Fig. 1.[15,16] In these nonparenchymal soft tissue regions, shear waves may also be present. Their behavior is primarily governed by the density and the shear viscoelastic moduli. Shear wave lengths and propagation speeds at a given frequency are typically 3 orders of magnitude smaller than that of compression waves in the same soft tissue medium.[19]

Boundary element (BE) or coupled finite element/BE methods have been used by researchers to simulate the inverse problem of magnetocardiography and electrocardiography to aid in diagnosis.[20–25] In other words, given an array of noninvasively recorded torso surface potentials or extracorporal magnetic field measurements, reconstruct epicardial potentials or myocardial activation times. In these and other studies, the theory behind the application of a BE method involves first expressing the governing field equations including source terms, here having to do with electromagnetics, i.e., Maxwell's equations, in terms of boundary integral equations by use of Green's theorems. Then, the boundary integral equations are approximated as a summation of coupled linear algebraic equations by replacing the integral over the boundary surfaces with a finite summation of simplified expressions over a discretized version of the surface (boundary elements). By beginning with the governing acoustic field equations, Helmholtz equations, including source terms, the BE method can be applied to the solution of comparable acoustic problems.

A BE model for lung acoustics is preferred over a finite element model due to the computational burden of meshing the entire parenchyma region. The speed of sound in parenchymal tissue is estimated to be $\sim 23\ \mathrm{m/s}$ in the frequency range of interest. This translates into wavelengths that are roughly ten times shorter than the acoustic waves in air for the same frequency. If the nominal element dimension needs to be $\sim 1/6$ of the wavelength for accuracy, the number of nodes and elements in a 3D finite element solution would be enormous. For example, at 220 Hz, one would need $\sim 120\,000$ degrees of freedom (DOF) in a finite element model for just the left lung of the visible human male (VHM).[26] Unlike the finite element method, a BE analysis requires only the surface of the parenchymal region to be meshed (with comparable demands on nominal element dimension), which drastically reduces the number of elements and degrees of freedom in the analysis. (Only ~ 4000 DOF are needed for 220 Hz for the left lung of the VHM.)

B. Objectives

The first objective of this study is to adapt an acoustic BE model to simulate sound propagation in the lung parenchyma. This BE model is first computationally validated and then compared with experimental studies on lung phantom models with simple geometries. The type of BE formulation that is used is based on the Burton–Miller integral equations.[27] The discretization and the solution of the integral equations are performed as presented in the literature on direct acoustic BE modeling.[28–30]

Another objective of this study is to be able to locate a localized acoustic source within the lung. Acoustic source localization is common to a number of engineering applications, including room acoustics, earthquake epicenter localization, and underwater sonar. Advanced sonar techniques use an analytic acoustic wave propagation model that represents the propagation and reflection of the acoustic waves from the ocean floor and the sea–air interface. The acoustic responses at sensors are predicted using this model and acoustic responses are calculated due to the acoustic source whose position is varied in a grid-like fashion. Sensor predictions for each hypothesized source location are compared with the actual sensor measurements in order to assign a probability value to a particular source location. This method is called matched field processing.[31–35] There are similarities between the lung acoustics problem posed in the present study and the source localization problem in underwater sonar research. There are studies in underwater sonar research which try to quantify and minimize the errors due to incorrect wave number assumptions (due to variations in water temperature and salinity) and imprecise knowledge of the acoustic and mechanical properties of the ocean floor.[33–35] Analogous error sources in the present application are imprecise wave number estimations of the parenchyma tissue and imprecise knowledge of the thickness and content of the biological tissue between the parenchyma and the skin surface. In this study, source localization will be performed using matched field processing principles; but, differing from underwater sonar studies, the field prediction will be calculated using an acoustic BE model.

II. BOUNDARY ELEMENT MODELING

A. Basic theory

The boundary element formulation that was developed and used to generate the numerical results presented in this article is primarily based on the work of Kirkup.[28,29] Using Green's first and second theorems, the Helmholtz equation, which governs compression wave propagation throughout the volume of a finite acoustic domain, can be expressed as a boundary integral equation:

$$\{H\varphi\}_S[\mathbf{p}] + \tfrac{1}{2}\varphi[\mathbf{p}] = \{L\mathbf{v}\}_S[\mathbf{p}] + \varphi[\mathbf{p}]^{\mathrm{inc}}, \tag{1}$$

where

$$\{H\varphi\}_S[\mathbf{p}] \equiv \int_S \frac{\partial G[\mathbf{p},\mathbf{q}]}{\partial n_q}\varphi[\mathbf{q}]dS, \tag{2}$$

$$\{L\mathbf{v}\}_S[\mathbf{p}] \equiv \int_S G[\mathbf{p},\mathbf{q}]\mathbf{v}[\mathbf{q}]dS, \tag{3}$$

$$\mathbf{v}[\mathbf{q}] = \frac{\partial\varphi[\mathbf{q}]}{\partial n_q}. \tag{4}$$

Here, S denotes the surface or boundary of the domain and the integrals are confined to this surface, $\mathbf{p}$ and $\mathbf{q}$ are points on the boundary, $\mathrm{G}[\mathbf{p},\mathbf{q}]$ is the Green's function that relates how the acoustic response at point $\mathbf{p}$ is affected by an acoustic point source (monopole) located at $\mathbf{q}$, and n_q is a unit vector outward normal to the surface at point $\mathbf{q}$. The velocity potential field at a point $\mathbf{p}$ is denoted by $\varphi[\mathbf{p}]$. The term $\varphi[\mathbf{p}]^{\mathrm{inc}}$ denotes the incident field that would exist at point $\mathbf{p}$ due to a distribution of sources within the domain if no boundary were present; i.e., it is the free field solution at point $\mathbf{p}$. And, $\mathbf{v}[\mathbf{q}]$ is the normal velocity component on the surface at point $\mathbf{q}$ directed outwards. The Green's function for the acoustic problem is

$$G[\mathbf{p},\mathbf{q}] = \frac{1}{4\pi R}e^{ikR}, \tag{5}$$

$$R = |\mathbf{p}-\mathbf{q}|, \tag{6}$$

where R is the distance between the point of observation and the source location. Boundary conditions on the domain surface S may be specified as either $\varphi[\mathbf{p}]$ or $\partial\varphi/\partial n_p[\mathbf{p}]$ or a linear combination of these.

The surface S is approximated with a finite set of planar triangles (boundary elements) ΔS_j, $j=1,\ldots,N$, and it is assumed that φ and $\mathbf{v}$ do not change within each triangle; this is referred to as C^0 collocation. Errors due to this approximation of the boundary and property values on the boundary become smaller as the number of elements used is increased. Thus:

$$S \approx \sum_{j=1}^{N} \Delta S_j, \tag{7}$$

$$\varphi[\mathbf{p}] \approx \varphi_j, \tag{8}$$

$$\mathbf{v}[\mathbf{p}] \approx \mathbf{v}_j \quad \text{if } \mathbf{p} \in \Delta S_j. \tag{9}$$

Then, Eq. (1) can be approximated as follows:

$$\sum_{j=1}^{N}\left\{H+\frac{1}{2}I\right\}_{\Delta S_j}[\mathbf{p}]\varphi_j \approx \sum_{j=1}^{N}\{L\}_{\Delta S_j}[\mathbf{p}]\mathbf{v}_j + \varphi[\mathbf{p}]^{\mathrm{inc}} \quad (p \in S). \tag{10}$$

The previous expression is for a single location point $\mathbf{p}$. For the complete solution of the problem one should calculate this for a point on each of the boundary elements, $i=1,2,\ldots,N$. This leads to

$$\left([\mathbf{H}]+\tfrac{1}{2}[\mathbf{I}]\right)\varphi \approx [\mathbf{L}]\mathbf{v}+\varphi^{\mathrm{inc}}, \tag{11}$$

$$[\mathbf{L}]_{ij} \equiv \{L\}_{\Delta S_j}[\mathbf{p}_i] = \int_{\Delta S_j} G_k[\mathbf{p}_i,\mathbf{q}_j]dS, \tag{12}$$

$$[\mathbf{H}]_{ij} \equiv \{H\}_{\Delta S_j}[\mathbf{p}_i] = \int_{\Delta S_j} \frac{\partial G_k[\mathbf{p}_i,\mathbf{q}_j]}{\partial n_j}dS. \tag{13}$$

The subscript k is added to emphasize that the k value in the Green's function is nonzero. It will be shown that the case where k is zero will be used to evaluate certain singular conditions. The previous integrals have Green's function or its derivative in the integrand. From Eq. (5) one can see that the R value in the Green's function will tend to zero as the observation element and source get close to each other. The integrand value will increase and the variations of the integrant within integration limits will be considerable. Therefore, high order quadrature rules should be used for integration on these elements. As the integrand values are big these elements would affect the result much more than the elements where the source and the observation elements are far away from each other. Considerable attention should be paid in using sufficient numbers of quadrature points in boundary element codes.

Further, the R value is exactly zero when evaluating the diagonal elements (source and the observation point are on the same element). The integrals of the diagonal elements become singular and cannot be evaluated using regular quadrature methods. A common approach to this problem is to subtract out the singularity (reducing the strength of the singularity) by subtracting out and adding another integrant. The below equation shows how one can evaluate the diagonal elements in the $[\mathbf{L}]$ matrix:

$$[\mathbf{L}]_{ii} = \int_{\Delta S_i}(G_k[\mathbf{p}_i,\mathbf{q}_i]-G_0[\mathbf{p}_i,\mathbf{q}_i])dS + \int_{\Delta S_i} G_0[\mathbf{p}_i,\mathbf{q}_i]dS. \tag{14}$$

Here, G_0 denotes that the k value in the Green's function is zero. Therefore, G_0 can be expressed as

$$G_0[\mathbf{p}_i,\mathbf{q}_i] = \frac{1}{4\pi R}. \tag{15}$$

This expression will tend to infinity as R gets smaller. But, when two large numbers are subtracted from each other, a smaller value will be obtained and the first integral in Eq. (14) becomes regularly integrable. The second integral though is still singular and is hard to evaluate numerically; but, it has an analytical solution.[28] Therefore, it is possible to evaluate the integrals corresponding to diagonal elements through a combination of numerical and analytical methods.

In the present application a sound generated within the parenchyma region, such as due to breathing, is approximated via a finite number of fundamental acoustic sources. The incident field that would be created by these sources at the locations $\mathbf{p}_i$ of the $i=1,\ldots,N$ centroids of the boundary elements that approximate the actual boundary is denoted by φ^{inc}, a vector of length N. This vector contains the analytically predicted values of the incident field caused by these sources assuming they radiated into an infinite medium. That is, the ith element of this vector represents the incident field on the ith element due to sources in the medium.

In Eq. (11) $\boldsymbol{\varphi}$ and $\mathbf{v}$ are vectors of length N. For the problem to be solvable one needs N boundary conditions

relating φ and $\mathbf{v}$. This can be mathematically stated as follows, where α_i, β_i, and f_i are constants that define the boundary conditions:

$$\alpha_i \varphi_i + \beta_i v_i = f_i, \quad i = 1,2, \ldots, N. \tag{16}$$

Once the boundary condition assignments and the calculations of $[\mathbf{H}]$ and $[\mathbf{L}]$ matrices are completed the problem can be solved using Gauss elimination or any other linear system solving method. Once the velocity potential and normal velocity components are calculated one can also find the acoustic pressure at the surface elements using the relationship: $p = i\rho\omega\varphi$.

Although there is tremendous advantage in not needing the mesh inside of the domain, there is still a computational burden, the majority of it coming from the need to use higher order quadrature methods and fully populated system matrices. For elements that are far from each other single point Gauss quadrature is sufficient; but, for elements that are close to each other as much as 171 Gauss quadrature points have been used for accurate evaluation (asymptotic convergence) of surface integrals in some of the examples cases described below.

B. Coupled boundary conditions for surrounding shell-like structure

A viscoelastic shell-like structure surrounding the BE domain can be approximated in a discretized form using, e.g., a finite element (FE) approach. If the FE degrees of freedom are coincident with the BE degrees of freedom (motion normal to each boundary element at its centroid) then we can write in place of Eq. (16) that:

$$-i\omega\rho\mathbf{A}\varphi + \left(i\frac{\mathbf{K}}{\omega} + \mathbf{C} - i\omega\mathbf{M}\right)\mathbf{v} = \mathbf{F}. \tag{17}$$

Here, $\mathbf{K}$, $\mathbf{M}$, and $\mathbf{C}$ are structural stiffness, mass, and damping matrices, respectively, for the surrounding shell. The $N \times N$ matrix $\mathbf{A}$ is diagonal, with $[\mathbf{A}]_{ii}$ being the area of the ith boundary element. The length N vector $\mathbf{F}$ represents forces applied externally to the shell, such as percussive excitation at the skin surface.

Generally, $\mathbf{K}$, $\mathbf{M}$, and $\mathbf{C}$ can be nondiagonal matrices that couple the N degrees of freedom. Simplifying assumptions though can lead to decoupled boundary conditions. For example, consider the case of an inertia load only on the elements. Suppose a layer of material of density ρ_s and thickness h_i resided on the ith element. Then, $\mathbf{M}$ is diagonal with $[\mathbf{M}]_{ii} = [\mathbf{A}]_{ii} h_i \rho_s$ and $\mathbf{K}$ and $\mathbf{C}$ are neglected. If $\mathbf{F} = 0$, Eq. (17) reduces to:

$$-\rho\varphi_i - h_i\rho_s\mathbf{v}_i = 0, \quad i = 1,2, \ldots, N. \tag{18}$$

III. COMPUTATIONAL VALIDATION STUDY

Comparison of the BE solution to a FE solution was performed for a cylindrical parenchymal region with different boundary conditions. Parenchymal material property values are given in Fig. 1 and Table I. The radius of the cylinder is 41.7 mm and its axial length is 83.3 mm, with 2352 elements used in the BE simulation. An axisymmetric FE model was constructed in ANSYS® 9.0 with 5000 Fluid79 elements. A monopole source of radius 2.4 mm was located on the central axis of the cylinder at a height of 55.5 mm (at 2/3 the length of the cylinder). Figure 2 shows the vertical velocity as a function of frequency for a point located at the bottom center point of the cylinder using the BE and FE methods for the following boundary conditions over the entire cylinder: a free boundary condition; and a 4 mm thick uniform "tissue" layer covering the parenchyma with density of 1 g/cm^3 and Young's modulus of 7.5 or 750 kPa. This density and the smaller modulus approximate biological soft tissue, whereas the larger modulus may approximate a composite of soft and hard (bone) tissue.

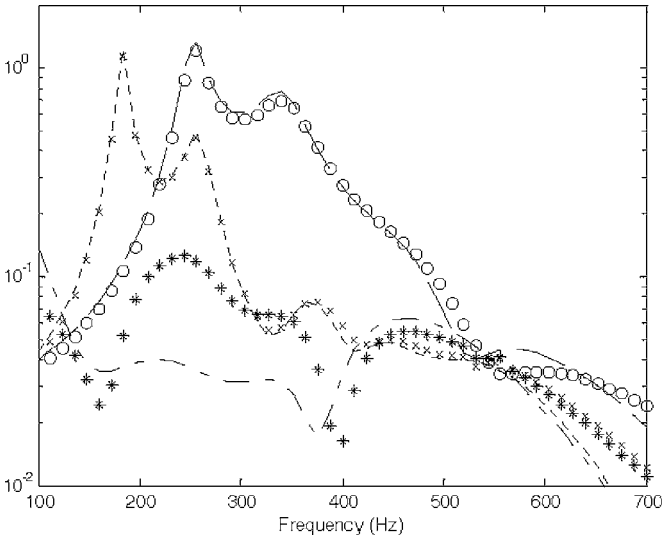


FIG. 2. Comparison of FE and BE results for cylindrical volume with a finite monopole located along the central axis at 2/3 of the height of the cylinder. Vertical velocity (m/s) on surface at bottom center of cylinder is shown for different uniform boundary conditions. Key: BE (– – –) and FE (○) results for free boundary; BE (---) and FE (×) results for 4 mm "soft" tissue layer covering the entire parenchyma surface with $\rho_s = 1$ g/cm^3 and Young's modulus $E = 7.5$ kPa; BE (– - –) and FE (*) results for 4 mm "hard" tissue layer covering the entire parenchyma surface with $\rho_s = 1$ g/cm^3 and Young's modulus $E = 750$ kPa.

For the cases of the tissue layer covering the parenchyma, mass ($\mathbf{M}$), stiffness ($\mathbf{K}$), and damping ($\mathbf{C}$) matrices were developed for Eq. (17) as follows. The mass matrix, $\mathbf{M}$, was diagonal with $[\mathbf{M}]_{ii} = [\mathbf{A}]_{ii} h_i \rho_s$, as described in the text before Eq. (18). Linear viscous proportional damping was used such that $\mathbf{C} = 0.001 * \mathbf{K}$. The stiffness matrix $\mathbf{K}$ was developed as follows. The BE node locations and meshing information were imported into ANSYS 9.0 FE software. A FE mesh was generated using three-node Shell63 shell elements with nodes coincident with the BE nodes. The stiffness matrix for this mesh $\mathbf{K_A}$, which includes three translational DOFs per node, was exported from ANSYS as a text file, which was then read into MATLAB® after some minor format editing. Motion normal to a boundary element at its centroid can be approximated using a linear interpolation of the motion of each of the three nodes that form the element. Motion of each of these nodes in the direction of the element normal can be expressed as a weighted sum of the global Cartesian coordinates with which the Ansys FE model is constructed via direction cosines. Consequently, a transformation matrix

T can be constructed that approximates motion of the N degrees of freedom of the BE model, in terms of the motion of the N_A degrees of freedom of the ANSYS FE model with $N_A > N$. (In this case study $N=2352$ and $N_A=3534$.) The **T** matrix is $N \times N_A$ in size, with potentially nine nonzero entries per row (three for each DOF of each of the three nodes that form an element). Via singular value decomposition, using the "pinv" command in MATLAB a pseudoinverse of **T** is calculated and referred to as $[\mathbf{T}]^{-1}$. It has dimensions $N_A \times N$. The stiffness matrix **K** of dimension $N \times N$ is then approximated as follows:

$$\mathbf{K} = \mathbf{T}\mathbf{K_A}[\mathbf{T}]^{-1}. \tag{19}$$

Note that this same transformation from $\mathbf{K_A}$ to **K** could be applied to approximate the mass matrix **M** given $\mathbf{M_A}$ obtained from ANSYS in the same way as the stiffness matrix $\mathbf{K_A}$ was obtained. In the present case study the diagonal **M** is easily determined directly.

Minor differences between the developed BE code predictions and those of the FE simulation for all cases considered are likely due in part to the fact that the BE method treats the monopole as a point source whereas it has a finite radius of 2.4 mm and represents an additional boundary in the FE case. Discrepancies may also be due to an insufficient number of elements used in the FE and/or BE simulation, though comparable sized elements enabled the BE model to exactly match theory for a case study with a spherically symmetric parenchymal geometry (not shown). FE and BE predictions disagree similarly for the free and covered surface case with 4 mm of "soft tissue" with density 1 g/cm^3 and Young's modulus $E=7.5$ kPa. Indeed, simulations with $E=0$ (mass loading only) agreed with this case indicating that the inertia effect of the 4 mm of soft tissue was much more dominant than the stiffness effect. This was also nearly true for $E=75$ kPa (not shown). For the case where $E=750$ kPa stiffness did have a significant effect and agreement between the BE and FE simulations is not as good. Clearly, the approximation involved in replacing $\mathbf{K_A}$ with **K** per Eq. (19) needs improvement or an alternative approach may be needed to account for the stiffness of the ribcage in the eventual application.

Next, a more complex boundary condition was considered. The surface of the cylindrical parenchymal region was covered with 4 mm of the soft tissue with density 1 g/cm^3 and Young's modulus $E=7.5$ kPa everywhere except along three circumferential bands, each of thickness 4 mm and width ~4 mm at heights on the side of the cylinder of ~10, ~32, and ~52 mm. In these regions bone was simulated with a density of 1.9 g/cm^3 and Young's modulus $E=10$ GPa. An axisymmetric ANSYS FE model was constructed as described earlier with material properties altered for the Shell63 elements within these three bands. For the BE model boundary conditions, the diagonal mass matrix **M** was again easily determined directly based on the density and thickness of the material overlying each boundary element of known area. A very sparse and diagonal stiffness matrix **K** was directly constructed by recognizing axisymmetry, neglecting the stiffness of the soft tissue per the previous discussion and by assuming that the thickness of the boundary layer is small relative to the cylinder radius. Under these conditions the stiffness value associated with a boundary element within one of the rib bands is given by Young's modulus of the bone multiplied by its thickness and area, and divided by the square of the radius of the cylinder. In Fig. 3, the axisymmetric response on the side of the cylinder from 100 to 700 Hz is shown, based on the FE and BE models. Agreement is good though not exact. The three dark bands of neglible motion are where the ribs were located. In the FE model, in addition to the rib regions themselves, there is a discernable attenuating effect near to these regions. In the BE model, the strong attenuating effect of the ribs seems to be confined to the rib region only. Though in both the FE and BE models, the overall acoustic pattern on the surface of the parenchyma is altered by the presence of the ribs.

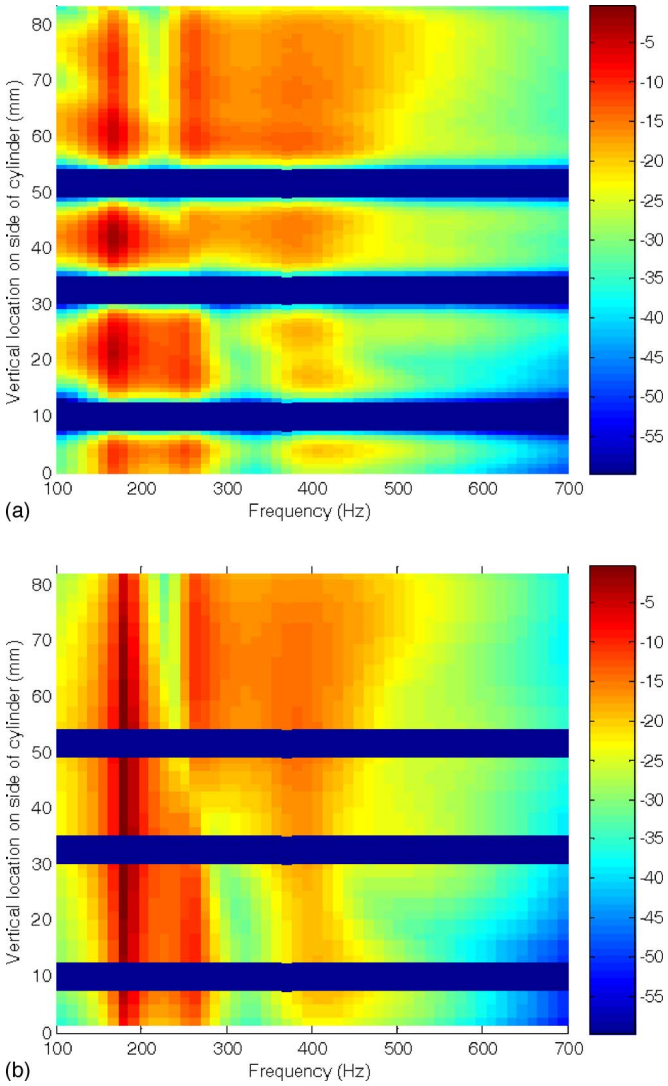


FIG. 3. (Color online) Comparison of (a) FE and (b) BE results for cylindrical volume with a finite monopole located along the central axis at 2/3 of the height of the cylinder and with a composite soft and hard tissue layer. Axisymmetric radial velocity (m/s) on the side of the cylinder as a function of excitation frequency is shown. In both plots the color bar unit is decibels reference 1 mm/s.

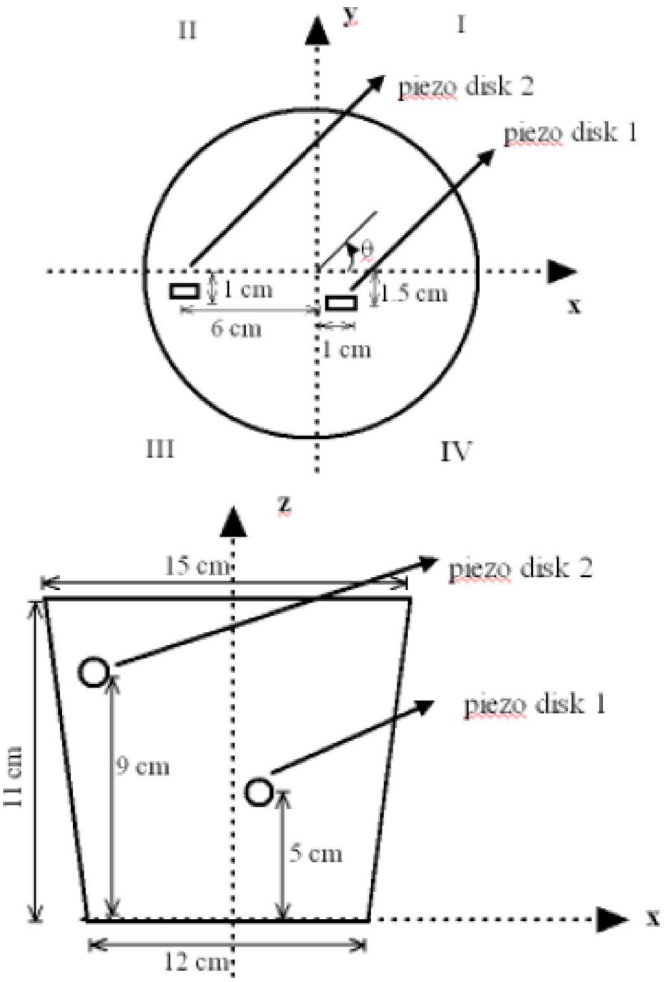


FIG. 4. (Color online) Top and side view diagrams showing the locations of the piezoelectric disk transducers in the lung phantom.

IV. EXPERIMENTAL EVALUATION USING A MECHANICAL PHANTOM MODEL

A. Setup

The acoustic response to an internal source in a mechanical model with acoustic properties similar to the lung parenchyma was investigated in order to: (1) determine if the acoustic field that is generated depends significantly on reflections from the finite boundaries and standing wave patterns or, due to the high rate of attenuation, is this field reasonably approximated by only accounting for outgoing waves from the internal source; (2) determine if the modeling assumptions that were made in formulating the BE modeling approach are reasonable for a compliant porous material like lung parenchyma; and (3) validate the accuracy of the developed BE model experimentally, which is related to and dependent on the second reason.

The experimental phantom setup is shown in Figs. 4 and 5. The foam material is Flex Foam-IT X (Smooth-On Inc., Easton, PA). In terms of phase speed and density, the foam material is comparable to lung tissue.[15,16] A theoretical estimate of the phase speed c_{ph}, which is nondispersive over the frequency range of interest, is based on $c_{\text{ph}}=\sqrt{1/K\rho}$ where ρ is the effective density of the porous foam and K is the volumetric compliance given by $K=(1-f_v)K_g+f_vK_s$, where

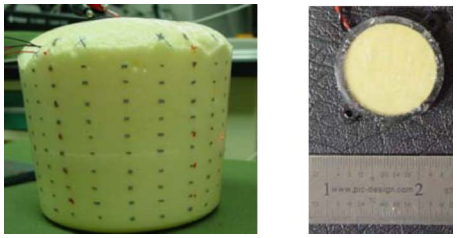


FIG. 5. (Color online) Lung phantom. (Left) Photograph showing LDV measurement points and (right) close-up of a piezoelectric disk transducer.

K_g is the entrapped gas compliance, K_s is the solid material compliance, and f_v is the volume fraction of solid material. The compliance K_s is unknown, but it is reasonably assumed that, like in the lungs, this term can be neglected relative to the gas compliance, $K_g \approx 1/P\kappa$, where P is the atmospheric pressure and κ is the polytropic constant (a value of 1.0 is assumed within the frequency range of interest[16]). The content of the entrapped gas is unknown; it is a by-product of the exothermic process of forming the foam by mixing two components. But, given atmospheric pressure and based on a measured foam density of 160 k/m^3 and volume fraction of solid material of $f_v=1/6$, we have $c_{\text{ph}}=27.5$ m/s. This value was consistent with experimental measurements of phase speed conducted using a 1.5 m length, 10 cm inner diameter rigid tube filled with the foam material with embedded piezoelectric disk sensors along its length at 15 cm intervals and driven by a piston actuator at one end. The resulting real part of the wave number k_R for this material is given by $k_R=2\pi f/c_{\text{ph}}=0.2285\times f$ m^{-1}, where f is the frequency in hertz. Recall, a typical value used for lung parenchyma is $c_{\text{ph}}=23$ m/s.

Regarding attenuation, below the resonant frequencies of the entrapped gas bubbles dissipative losses are expected to roughly be a function of the frequency f cubed and to be dependent on the foam material properties as well as the dimensions and spacing of the entrapped gas bubbles; this is difficult to measure. Consequently, experimental measurements of attenuation as a function of frequency within the same foam-filled tube were used to approximately fit an attenuation curve, yielding an imaginary component to the wave number of $k_i=6\times10^{-8}\times f^3$. For the lung parenchyma, Wodicka *et al.*[16] and Royston *et al.*[15] have used $k_i\approx18.1\times10^{-6}\times f^3$. Thus, the foam damping is about one third that of parenchymal tissue.

In the experimental phantom setup depicted in Figs. 4 and 5, there are two buried piezoelectric disk transducer pairs put in place prior to curing the foam (locations shown in Fig. 4). A single "piezodisk" transducer is shown in Fig. 5. Two of these are glued back-to-back to form each internal acoustic source. During the experiment a band-limited random noise is supplied to the transducers through a high voltage amplifier (P0623A, Trek, Medina, NY). In the BE simulations, it was assumed that these sources radiate with spherical symmetry, like a monopole, into the foam phantom. Normal velocity at the outer surface of the foam was measured at discrete points using a laser doppler vibrometer (LDV) (CLV 800/1000, Polytec, Tustin, CA) with data acquisition and spectral analysis performed using a two-channel spectrum

analyzer (35670A, Agilent, Palo Alto, CA). The geometry of the foam material is a tapered cylinder. The bottom of the cylinder has a diameter of 12 cm and diameter linearly increases to 15 cm over a cylinder height of 11 cm. Note, although there are two acoustic sources buried at two different locations in order to test different methods for source localization of the multiple acoustic sources, the work reported here is confined to localization of and response to a single source. Also, a 2 mm thick layer of CF-11 silicone (Nusil Technologies, Carpinteria, CA) was adhered to the conical portion of the foam phantom and experimental measurements were repeated. Due to the translucence of the silicone, vibrometry measurement points were taken at the same location depicted in Fig. 5(a) on the now-covered foam.

B. Results and discussion

Normal velocity measurements on the radial surface of the tapered foam cylinder were taken at 216 locations. Acoustic source 1 was used (Fig. 4). The BE model predicted the acoustic response due to a sinusoidal acoustic input. The BE model has 2352 elements. A contour plot of the experimental velocity measurements is shown in Figs. 6(a), 7(a), and 8(a) for frequencies of 300, 500, and 700 Hz, respectively. The horizontal axis denotes angular location, as established in Fig. 4. The vertical axis is the height of the measurement point. Corresponding BE predictions at the experimental locations are plotted in contour form in Figs. 6(b), 7(b), and 8(b). Also, predictions based on assuming free field propagation, neglecting boundary reflections, are shown in Figs. 6(c), 7(c), and 8(c). (Note, for this experimental phantom it is not possible to generate a plot of response at a point as a function of frequency that can be directly compared to the BE simulation, as the strength of the piezodisk source is unknown. It will vary with frequency and will be dependent on the coupling of the source to the surrounding foam material. Hence, the piezodisk source cannot be calibrated separately and then implemented in the phantom.)

It is clear from comparing these plots that the assumption of free field propagation is not justified. The frequency-dependent wave number k based on theory and experiments described in Sec. IV A is used for BE and free-field simulations reported here. It is expected that as the excitation frequency increases damping will increase and the velocity distribution will look more similar to the free field propagation predictions. However, lung sounds and other sounds of interest in diagnosis will have a dominant component at the lower end of the audible frequency range that will diminish as the frequency increases beyond 1000 Hz or so. Although neither the free field assumption nor the BE simulation precisely match experiment for any of the frequencies, in all cases (those shown here and at other frequencies) the BE simulation is more accurate, capturing standing wave patterns that tend to increase the spatial variation of the amplitude at the surface as compared to the free field assumption. The wider range of amplitude values predicted by the BE simulation is closer to that of the experiment, as compared to the free field assumption. (Note, the simulated amplitude level depends on an assumed piezoelectric disk source strength; this value is only an approximation with frequency-dependent accuracy. However, the range of amplitude values, maximum minus minimum, is independent of the source strength, as this is a linear system.)

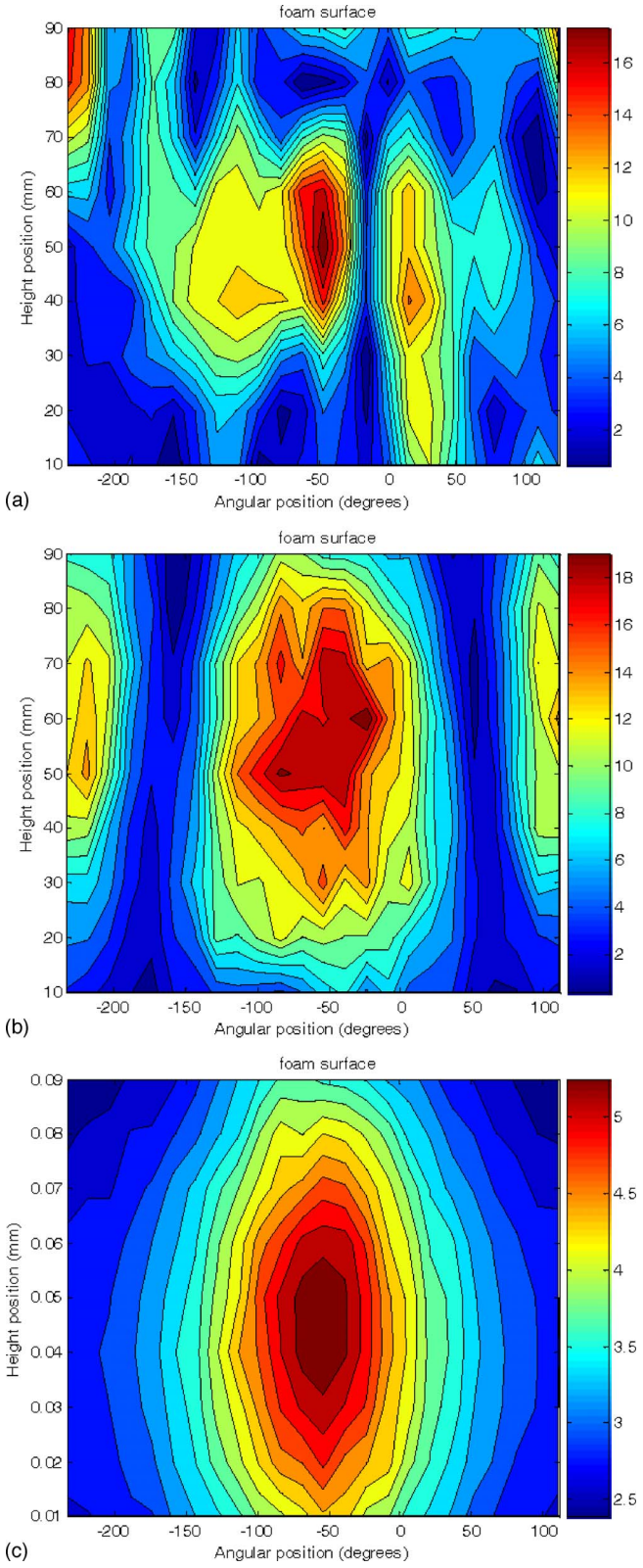
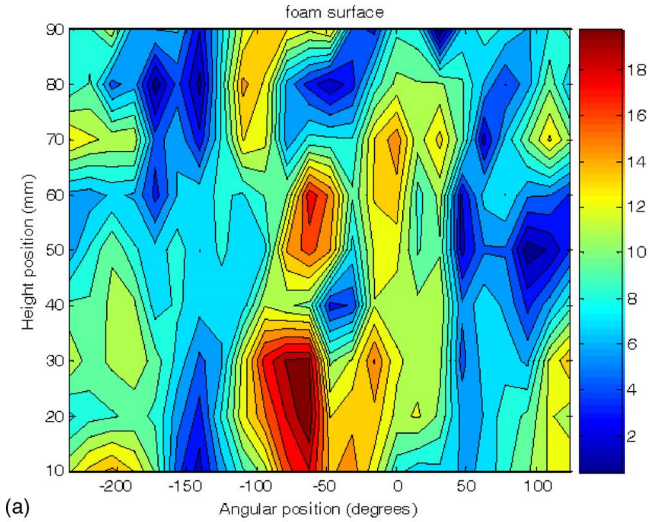
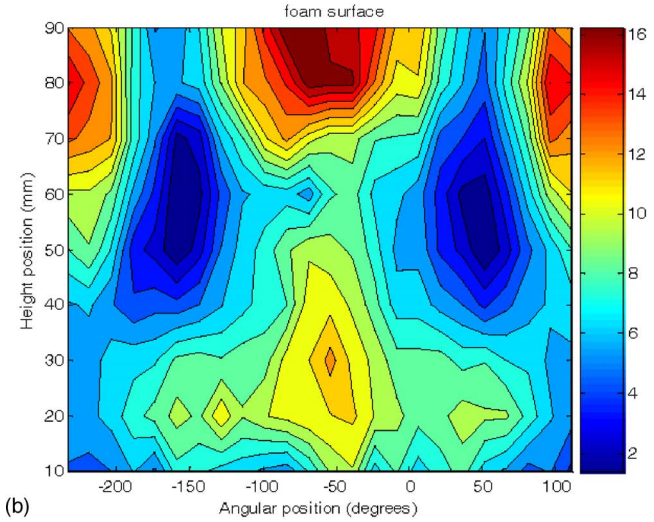


FIG. 6. (Color online) Velocity field amplitude on lateral (curved) surface of lung phantom at 300 Hz. (a) Experiment, (b) simulated using BE model, and (c) simulated using free field assumption. In all plots color bar unit is millimeter per second.





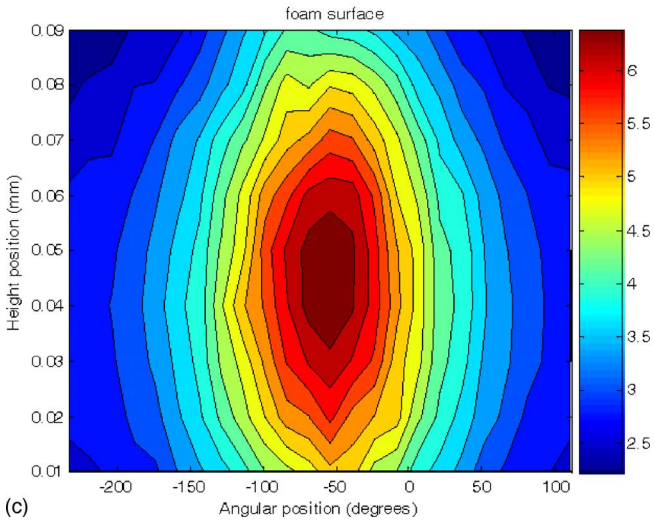
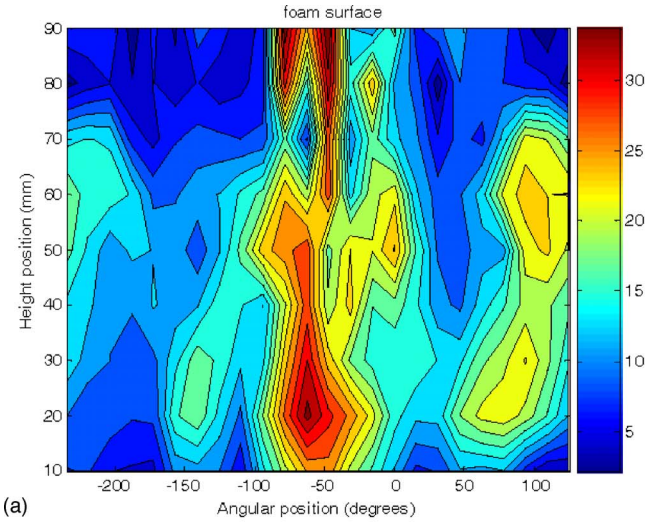
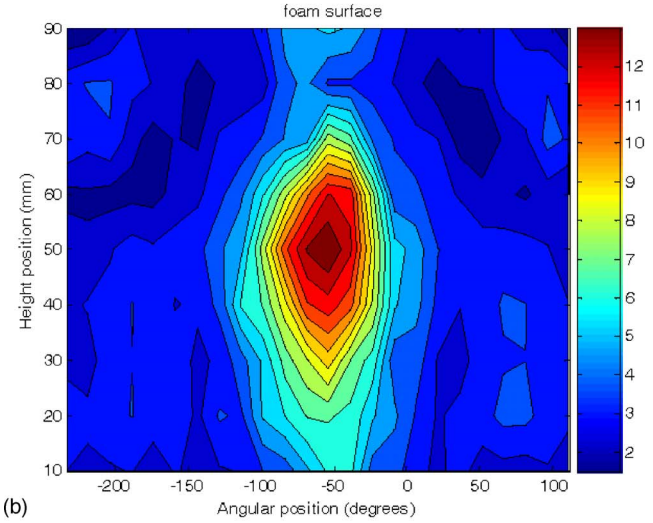


FIG. 7. (Color online) Velocity field amplitude on lateral (curved) surface of lung phantom at 500 Hz. (a) Experiment, (b) simulated using BE model, and (c) simulated using free field assumption. In all plots color bar unit is millimeter per second.





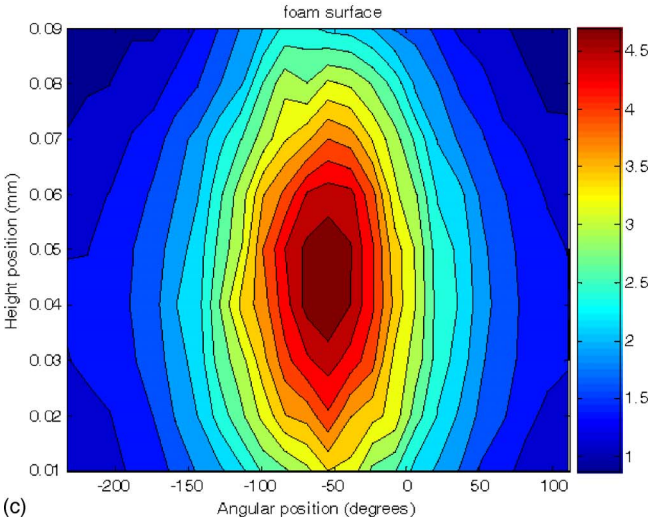
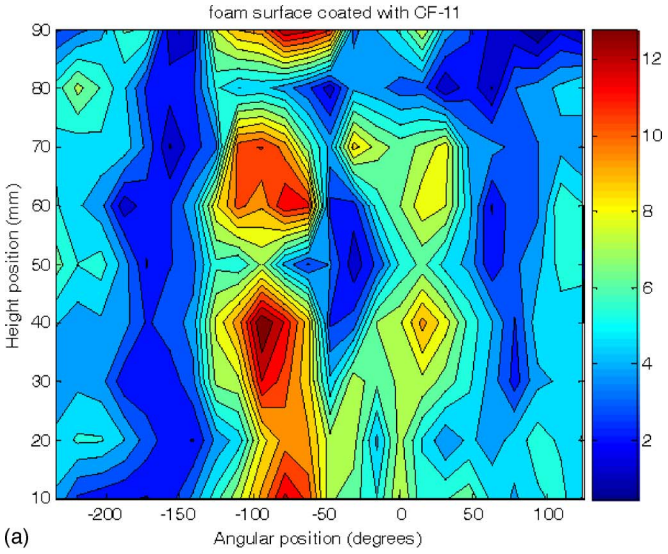


FIG. 8. (Color online) Velocity field amplitude on lateral (curved) surface of lung phantom at 700 Hz. (a) Experiment, (b) simulated using BE model, and (c) simulated using free field assumption. In all plots color bar unit is mm/s.

There are significant differences between the assumptions in the simulations and the experimental system that would account for discrepancies in numerical predictions and measurements. First, the simulations assume a homogeneous medium. In the process of curing the foam, which expands to ~6 times its original volume, it was found that air bubbles generated near the bottom of the mold were smaller in size



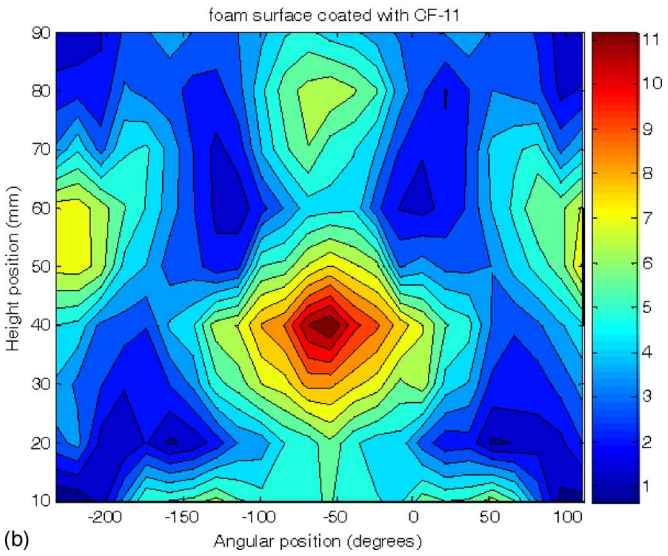


FIG. 9. (Color online) Velocity field amplitude on lateral (curved) surface of lung phantom at 500 Hz with 2 mm silicone coating. (a) Experiment and (b) simulated using BE model. In all plots color bar unit is millimeter per second.

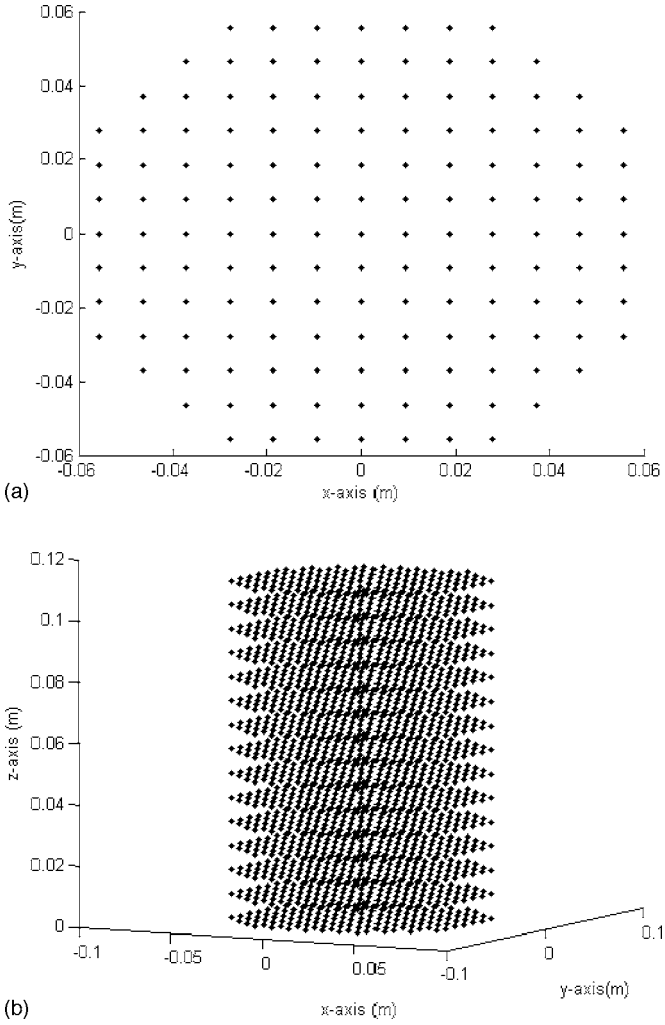


FIG. 10. Hypothetical point source locations using source localization algorithm. (a) Top view and (b) three-dimensional view.

than those generated at the top. Second, the embedded acoustic source is not a point monopole. The piezodisk transducer has a diameter of 24 mm, more than 1/5 of the phantom model diameter. The transducers directivity pattern is not spherically symmetric.

Next, a 2 mm thick layer of CF-11 silicone (Nusil Technologies, Carpinteria, CA) was adhered to the conical portion of the phantom and experimental measurements were repeated as described earlier. A contour image of the experimental measurements is depicted in Fig. 9(a). Results of a BE simulation treating the silicone layer as a thin shell per Sec. II B are shown in Fig. 9(b). Material properties for the silicone material are: density=1049 kg/m^3, Young's modulus=72 kPa, Poisson's ratio=0.495 and linear viscous damping proportional to the stiffness matrix was assumed via a proportionality constant β=0.001. For this 2 mm thick layer, the silicone is expected to primarily behave as a mass load. The BE simulation results match experimental measurements as closely as simulation matched experiment in the previous uncoated cases. Although specific regions of high and low response amplitude are not precisely matched, the range of amplitudes and general distribution of amplitudes are comparable.

V. SOURCE LOCALIZATION

A. Matched field processing

The simple beamformer equation for the Bartlett processor is[36,37]

$$B[\boldsymbol{\theta}] = \frac{|\mathbf{a}^H[\omega,\boldsymbol{\theta}]\mathbf{y}[\omega]|^2}{\mathbf{a}^H[\omega,\boldsymbol{\theta}]\mathbf{a}[\omega,\boldsymbol{\theta}]}. \tag{20}$$

Here, $\mathbf{a}$ is a vector of length N of the predicted responses at N sensor locations due to a monopole source located at θ at frequency ω; $\mathbf{y}$ is a vector of the measured responses at those N sensor locations. Here, superscript H denotes the Hermitian of the vector. First, a grid is created within the volume where the source will be searched. A hypothetical monopole source is placed at N_s grid locations and the Bartlett processor is evaluated. The grid locations are shown in Fig. 10. The BE model is computationally efficient for this type of grid

TABLE II. Distance between the actual source location and estimated source locations at 300 Hz.

	Distance (mm)	
	BE code	Free-field
Baseline case with 216 sensors	18.7	33.8
Baseline case with 9 sensors	18.7	33.8
With −10% error in k and 216 sensors	21	33.8
With −10% error in k and 9 sensors	21	33.8

search algorithm, as it is easy to calculate the incident field created by a source at each grid location, which then only alters φ^{inc} in Eq. (11). The location $\boldsymbol{\theta}$ that results in the maximum value of $B[\boldsymbol{\theta}]$ indicates the most likely location for the source. The case considered here was that of free boundary conditions on the small experimental phantom model of Sec. IV. For this case, Eq. (11) reduces to

$$\mathbf{0} = [\mathbf{L}]\mathbf{v} + \varphi^{\text{inc}}. \tag{21}$$

For a hypothetical source located at θ we then have

$$\mathbf{a}[\omega, \boldsymbol{\theta}] = -[\mathbf{L}]^{-1}|_{\text{sub}}\varphi^{\text{inc}}, \tag{22}$$

where $[\mathbf{L}]^{-1}|_{\text{sub}}$ refers to a subset of N rows of $[\mathbf{L}]^{-1}$ that correspond to N boundary elements that are coincident with the N sensor locations. If a free field assumption is used, then in place of Eq. (22) we simply have

$$\mathbf{a}[\omega, \boldsymbol{\theta}] = \mathbf{v}^{\text{inc}}, \tag{23}$$

where $\mathbf{v}^{\text{inc}}$ is the calculated particle velocity amplitude normal to the boundary surface due to a source located within the boundary at θ, assuming no reflections at the boundary.

B. Experimental results

Likelihood estimates $B[\boldsymbol{\theta}]$ are calculated at each node of the grid geometry based on the BE model and based on a free-field assumption using Eqs. (20), (22), and (23). The vector $\mathbf{y}$ of the measured responses is taken from experimental measurements summarized in Sec. IV [Fig. 6(a)]. The actual source location is at $x=10$ mm, $y=-15$ mm, and $z=50$ mm. Grid spacing is such that the nearest grid point is 4.7 mm from this location; thus, this is the closest possible source location prediction. Table II summarizes case studies done at 300 Hz, in terms of what the distance was from the predicted source location (grid point resulting in maximum value for $B[\boldsymbol{\theta}]$) to the actual source location. Although predicted and actual locations do differ in all cases, the prediction using the BE model consistently outperforms the prediction based on the free field assumption. As the frequency was increased to 700 Hz, differences between predictions using the BE model and free field assumption did become less but were still present. This is due to the fact that the free field assumption becomes more realistic as frequency increases.

Although Table II only indicates the difference between the location of maximum value of $B[\boldsymbol{\theta}]$ and the actual source location, 3D graphical depictions in Fig. 11 provide more detail for two cases. Here, the boundary element mesh is shown, the actual source location is at the intersection of the three thick lines, and isosurfaces are shown for $B[\theta]$ calculated using the BE model and free field assumption. The calculation using the BE model results in the smaller (and darker in grayscale) single isosurface closer to the source. Specifically, the isosurface encompasses the interpolated 3D region of θ within which $B[\theta]$ exceeds 90% of its maximum value. These isosurfaces illustrate the relative levels of precision, as well as accuracy, of the Bartlett beamformer based

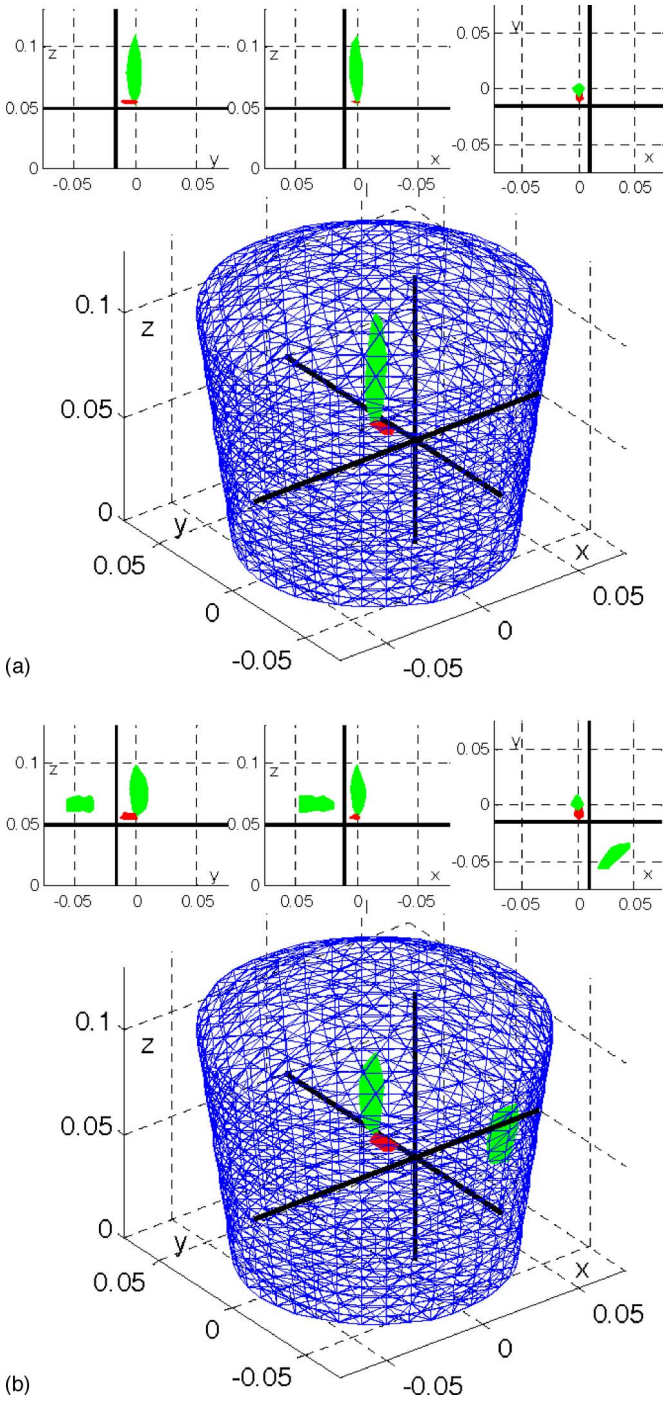


FIG. 11. (Color online) BE surface mesh is shown; actual source location is at the intersection of the three thick (black) lines; iso-surfaces are shown (in red and green) for $B[\boldsymbol{\theta}]$ calculated using the BE model and free field assumption (respectively). The calculation using the BE model results in the smaller (and darker in grayscale) single iso-surface closer to the source in both cases. Each iso-surface encompasses the interpolated 3-dimensional region of theta within which $B[\boldsymbol{\theta}]$ exceeds 90% of its maximum value.

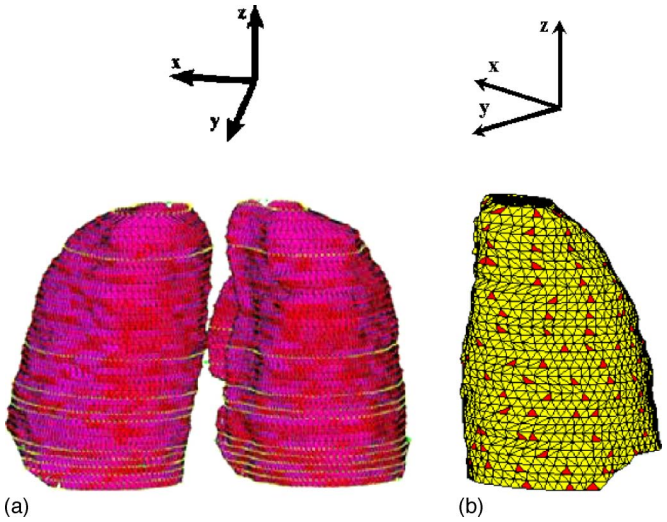


FIG. 12. (Color online) Visible human male. (a) Both lungs meshed with high resolution. Each lung is composed of ~10 000 triangular elements. (b) Left lung meshed with 4052 elements. The location of the 247 sensing locations on the lung parenchyma used in the source localization study are indicated by the darker grayscale marks. Positive y denotes anterior direction.

on the BE or free-field approaches. In general, use of the BE model results in a more precise, or focused, prediction for the source location given the same number of surface sensors. Generally, increasing the number of surface sensors better focuses the predicted source location.

The material properties of the lung parenchyma are frequency dependent, are not known precisely, and will be subject dependent to some degree. Consequently, source localization sensitivity to inaccurate values for the complex wave number was investigated. Real and imaginary parts of the wave number calculated at 300 Hz per Sec. IV A were reduced by 10% and then used to determine $\mathbf{a}[\omega,\theta]$ in Eqs. (22) and (23). Experimental measurements $\mathbf{y}$ at 300 Hz were used with these calculations of $\mathbf{a}[\omega,\theta]$ to determine $B[\theta]$ using Eq. (20). Table II shows the resulting difference in predicted and actual source location values under these conditions for different numbers of sensors. In this case, both BE and free-field approaches seem robust to wave number errors, and the BE approach still provides a better estimate of the source location.

VI. NUMERICAL STUDY USING ACTUAL LUNG GEOMETRY FROM VISIBLE HUMAN PROJECT

A. Meshing and acoustic simulations

Simulation results using the boundary element model are presented for actual lung geometry derived from the visible human male[26] (Visible Human Project, National Library of Medicine). Lung geometry was extracted based on x-ray CT images available online. Image processing was performed using MATLAB. External nodes on the lung surface were selected using a threshold technique and meshed within the MATLAB environment; Fig. 12 shows sample meshes. Acoustic simulations were performed on the left lung only with an internal acoustic source. After the mesh was obtained node and connectivity information were input into the BE code.

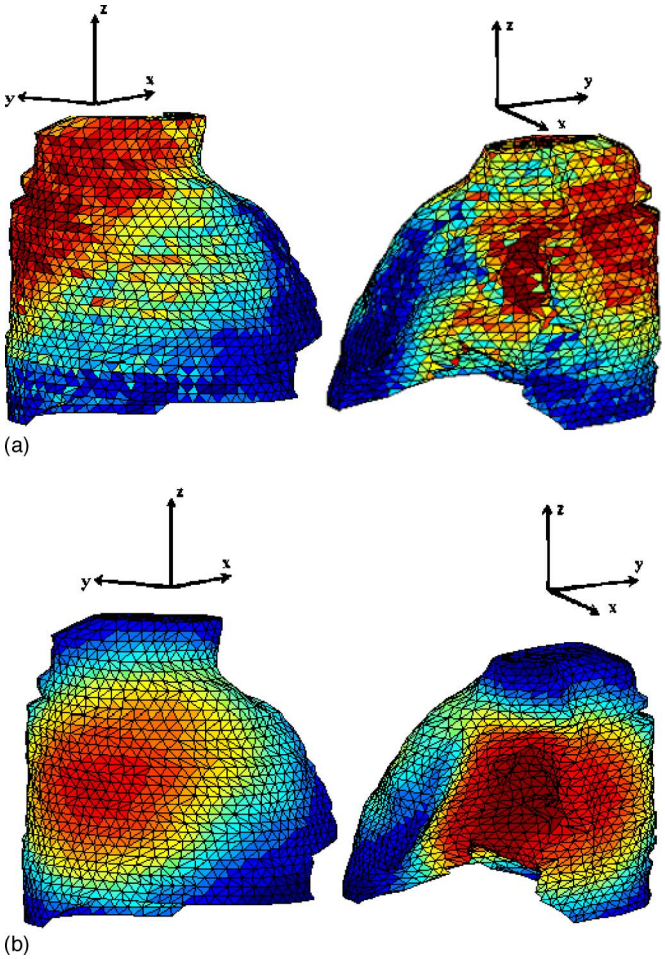


FIG. 13. (Color online) Velocity field due to point acoustic source at 220 Hz calculated using (a) BE model (two different views) and (b) free field model (two different views). Positive y denotes anterior direction.

The wave number value used was $k=60+2i$, which corresponds to 220 Hz per Fig. 1 and Table I.

The main reason for not considering higher frequencies was that a finer mesh would be needed resulting in an increased number of elements. All the simulations were performed in MATLAB 7.0 in Windows XP. The usable memory limit of the 32 bit processor Pentium IV system was 2 GB; however, it was observed that the maximum memory limit was reached when MATLAB variables occupied 800 MB. Therefore, going over 6000 elements was not possible due to computational limitations. However, these memory limitations would not apply for a system with a 64 bit processor. Although 6000 elements is not very many for finite element simulations, the matrices in finite elements are sparse and real. The system matrices in boundary element analysis are fully populated and complex. The number of finite elements that would be needed to obtain the same resolution as ~4000 boundary elements with ~4000 DOF is on the order of 20 000 finite elements with ~120 000 DOF.

BE simulation results in Fig. 13(a) show the velocity field due to a monopole acoustic source located at the centroid of the lung. Figure 13(b) shows the predicted velocity field for the same case but using a free-field assumption; results differ substantially from the BE prediction. The acoustic wavelength at 220 Hz is more than 10 cm, compa-

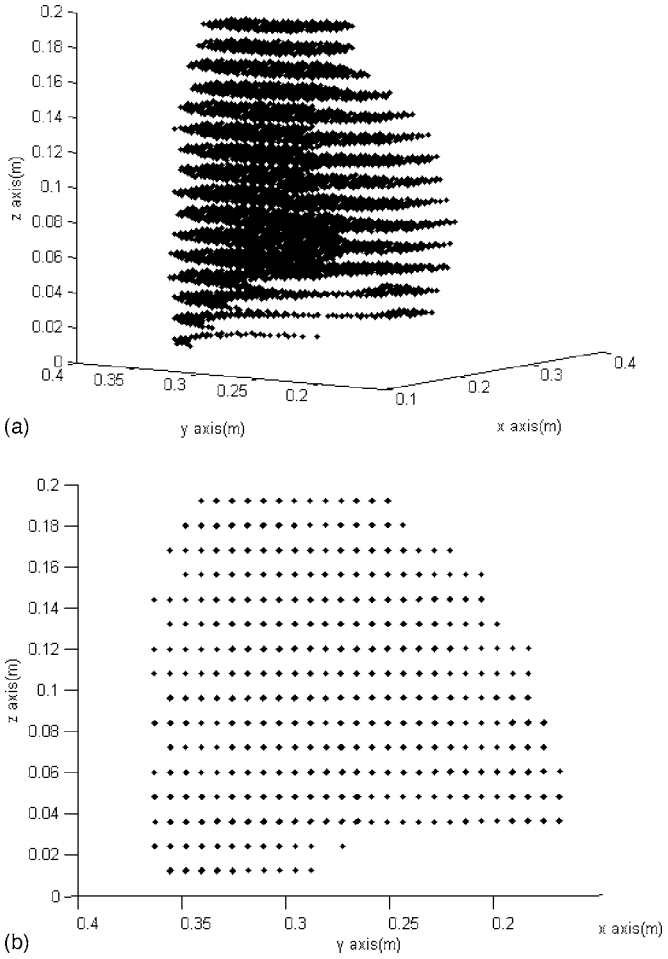


FIG. 14. 3436 hypothetical point source locations using source localization algorithm. (a) 3D view and (b) side view. Positive y denotes anterior direction.

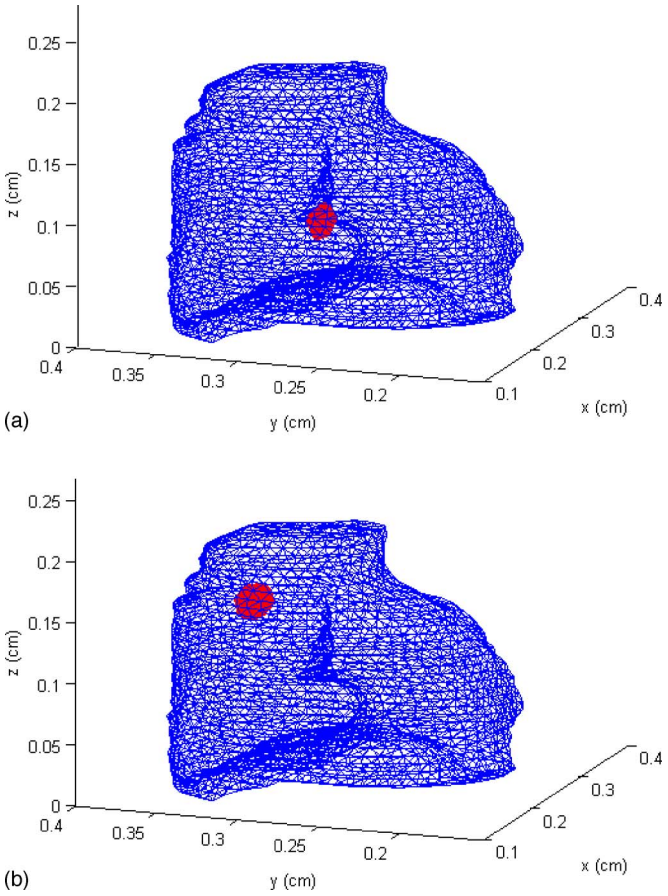


FIG. 15. (Color online) Source localization results at 220 Hz using Bartlett beamformer coupled with (a) BE model and (b) free-field propagation assumption. Positive y denotes anterior direction. Each isosurface encompasses the interpolated three-dimensional region of $\boldsymbol{\theta}$ within which $B[\boldsymbol{\theta}]$ exceeds 87% of its maximum value.

rable to the size of the geometry. As frequency increases, one would first expect the field to become more complex as standing wave shapes become more complicated, further reducing the accuracy of the free field assumption. But, as frequency increases even further, ultimately damping will diminish standing wave patterns and the free field assumption will become more reasonable.

B. Source localization studies

Unlike the previous source localization study on the mechanical phantom model, experimental results are not available here. Consequently, the BE simulation was used to calculate the "actual" acoustic response. In other words $\mathbf{y}=\mathbf{a}[\omega,\boldsymbol{\theta}]$, calculated based on Eq. (22). Figure 14 shows the internal points where hypothetical sources were placed from two different views for the source localization algorithm. A total of 3436 internal points were selected at 16 height levels in the source localization algorithm. Each point is spaced 7.5 mm apart from its nearest node. A total of 247 sensing points are used and their distribution on the lung mesh is shown in Fig. 12(b). The procedure for source localization is the same as in Sec. V. Figure 15(a) shows the result of the source localization process when the BE model is used for predicting the velocity field. Of course, the source location is exact as the boundary element simulation results were also used to generate the "actual" acoustic response. Figure 15(b) shows the results of the source localization process when free-field propagation model was used to simulate the velocity field. The estimate is 68 mm away from the actual source location where the minimum possible error due to discretization is 3.8 mm.

VII. DISCUSSION AND CONCLUSION

An acoustic BE model was used to simulate the propagation of sound waves in compliant, porous material like the lung parenchyma. The BE model compared well with numerical (finite element) simulations for fixed, free and mass-loaded (soft tissue with minimal stiffness) boundary conditions. Initial approaches and case studies for stiffer shell structures at the surface of the BE domain implemented as complex coupled boundary conditions derived from a reduced-order finite element model of the shell compared less favorably with results from a finite element analysis of the entire coupled system. Some improvement is needed for accurate application to human anatomy in order to appropriately account for the rib cage structure.

Experimental studies were performed on a phantom composed of a foam material that has similar acoustic properties to healthy lung parenchyma. Numerical predictions us-

ing the developed BE model had better agreement with experimental measurements as compared to predictions based on a free field assumption. The experimental studies and numerical simulations indicate that the acoustic response field caused by an internal acoustic source is different than that of free field acoustic propagation away from the source, at least in the low audible frequency range, which is the range within which most lung sounds occur. This is relevant to the development of new diagnostic techniques and is an advancement over prior lung acoustic imaging studies that used multiple sensor locations either to map out the sound field that is observed on the torso surface or, by assuming a free field propagation model, to estimate the location of a sound source. Results reported here do support the widely used assumption that acoustic compression waves dominate the response and the shear and surface waves can be neglected for frequencies in the hundreds of hertz. (Note the good comparison between the compression wave only BE model and the FE model simulations, which account for all wave types.)

The source localization studies suggest that the developed BE model coupled with a Bartlett beamformer (matched field processing) yields more accurate results than when the free field assumption is coupled with the same matched field processing algorithm. Source localization results are accurate even when significant errors in the phase speed and attenuation rate are used in the BE-based localization algorithm. Accuracy is dependent on the number of sensors used and becomes more robust to errors in phase speed, attenuation and size when the number of sensors is increased.

The case studies performed with an actual lung geometry that is extracted from the Visible Human Project showed that the velocity field of the boundary element model is different than the velocity field predicted by free-field propagation model. This difference had a significant effect on the source localization problem.

In Sec. IV it was noted that the phase speed and density of the foam material used in the phantom study were similar to those of human parenchymal tissue when compared to what has been reported in the literature based on experiments and theory. In terms of attenuation the foam was about 1/3 as damped as parenchymal tissue based on studies available in the literature. Few researchers have attempted to derive or measure a value for the attenuation rate. *In vivo* measurement of this value, as well as phase speed, is difficult given that it is impossible to know exactly what path the sound may have taken from a source to a receiver. This path may have involved airborne propagation within the bronchial tree at much faster speeds and will be influenced by the soft and hard tissues of the thorax surrounding the lungs and the impedance mismatches at tissue interfaces. Additionally, in many attempted measurements stethoscopes are used, which can be highly nonlinear in their amplitude response and spectral weighting.[3] The model of Wodicka *et al.* derived from first principles[16] and the empirical model of Vovk *et al.*[17] yield nondispersive phase speed calculations for the frequency range of interest. The attenuation rate derived by Wodicka *et al.* based on a bubble swarm analogy increases more rapidly with frequency than the "loss factor" model of Vovk *et al.*, which, according to that article, was based on analysis of experimental data given in another reference (in Russian). Careful measurements of phase speed by Paciej *et al.*[14] that avoid the trachea and bronchial airways by inputting sound at the supraclavicular space with multiple measurement points on the thoracic surface seem to more closely support the bubble swarm model. Experimental work of Wodicka and associates[38–40] do not conflict with their theoretical model's predictions of phase speed and attenuation for the frequency range of 100–600 Hz, though the ability to experimentally determine the attenuation of the parenchymal material alone is difficult, as noted earlier. Extrapolating the theory of Wodicka *et al.* to say 1000 or 2000 Hz results in attenuation rates of 8 and 60 orders of magnitude, respectively, which may seem unlikely given that measurements of lung sounds have been recorded up to at least 2000 Hz.[3,41] In summary, it seems that there is some range of uncertainty as to what the actual damping level within the parenchyma is as a function of frequency. Regardless, in this study it was found that both in terms of predicting the acoustic field and in terms of acoustic source localization, the BE method yielded accurate results even when the assumed attenuation rate was in error. Additionally, simulation studies (not shown) with the geometry of the mechanical phantom but the higher attenuation rate of Wodicka *et al.* for lung tissue indicated that at least up to 700 Hz there was a significant difference between the free field model and the BE model.

Extending the presented technique to the envisioned application, lung diagnosis *in vivo*, will require accounting for the presence of multiple, coherent and distributed sound sources as a part of the source localization problem. It may be necessary to use more sophisticated source localization algorithms. Parametric methods, such as maximum likelihood methods, may be required to deal with this complex source localization problem since simple methods, like the Bartlett beamformer used in this study, are not as effective when dealing with multiple coherent sources.[37]

ACKNOWLEDGMENT

Support for this research from the National Institutes of Health is acknowledged (Grant Nos. HL061108 and EB003286).

[1] A. Sarvazyan, "Audible-frequency medical diagnostic methods: Past, present and future," J. Acoust. Soc. Am. **117**, 2586 (2005).
[2] M. Akay, Y. Akay, W. Welowitz, J. L. Semmlow, and J. B. Kostis, "Application of adaptive filters to noninvasive acoustical detection of coronary occlusions before and after angioplasty," IEEE Trans. Biomed. Eng. **39**, 176–183 (1992).
[3] H. Pasterkamp, S. S. Kraman, and G. R. Wodicka, "Respiratory sounds: Advances beyond the stethoscope," Am. J. Respir. Crit. Care Med. **156**, 974–87 (1997).
[4] H. A. Mansy, T. J. Royston, and R. H. Sandler, "Acoustic characteristics of air cavities at low audible frequencies with application to pneumoperitoneum diagnosis," Med. Biol. Eng. Comput. **39**, 159–167 (2001).
[5] H. A. Mansy, T. J. Royston, and R. H. Sandler, "Use of abdominal percussion for pneumoperitoneum detection," Med. Biol. Eng. Comput. **40**, 439–446 (2002).
[6] H. A. Mansy, S. J. Hoxie, N. H. Patel, and R. H. Sandler, "Computerised analysis of auscultatory sounds associated with vascular patency of haemodialysis access," Med. Biol. Eng. Comput. **43**, 56–62 (2005).
[7] H. A. Mansy, R. Balk, T. J. Royston, and R. H. Sandler, "Pneumothorax detection using pulmonary acoustic transmission measurements," Med. Biol. Eng. Comput. **40**, 520–525 (2002).

[8]H. A. Mansy, R. Balk, T. J. Royston, and R. H. Sandler, "Pneumothorax detection using computerized analysis of breath sounds," Med. Biol. Eng. Comput. **40**, 526–532 (2002).
[9]H. Pasterkamp, R. Consunji-Araneta, Y. Oh, and J. Holbrow, "Chest surface mapping of lung sounds during methacholine challenge," Pediatr. Pulmonol **23**, 21–30 (1997).
[10]M. Kompis, H. Pastercamp, and G. R. Wodicka, "Acoustic imaging of the human chest," Chest **120**, 1309–1321 (2001).
[11]G. Benedetto, F. Dalmasso, and R. Spagnolo, "Surface distribution of crackling sounds," IEEE Trans. Biomed. Eng. **35**, 406–412 (1988).
[12]T. Bergstresser, D. Ofengeim, A. Vyshedskiy, J. Shane, and R. Murphy, "Sound transmission in the lung as a function of lung volume," J. Appl. Physiol. **93**, 667–674 (2002).
[13]S. Charleston-Villalobos, S. Cortes-Rubiano, R. Gonzalez-Camarena, G. Chi-Lem, and T. Aljama-Corrales, "Respiratory acoustic thoracic imaging (RATHI): assessing deterministic interpolation techniques," Med. Biol. Eng. Comput. **42**, 618–626 (2004).
[14]R. Paciej, A. Vyshedskiy, J. Shane, and R. Murphy, "Transpulmonary speed of sound input into the supraclavicular space," J. Appl. Physiol. **94**, 604–611 (2003).
[15]T. J. Royston, X. Zhang, H. A. Mansy, and R. H. Sandler, "Modeling sound transmission through the pulmonary system and chest with application to diagnosis of a collapsed lung," J. Acoust. Soc. Am. **111**, 1931–1946 (2002).
[16]G. R. Wodicka, K. N. Stevens, H. L. Golub, E. G. Cravalho, and D. C. Shannon, "A model of acoustic transmission in the respiratory system," IEEE Trans. Biomed. Eng. **36**, 925–34 (1989).
[17]I. V. Vovk, V. T. Grinchenko, and V. N. Oleinik, "Modeling the acoustic properties of the chest and measuring breath sounds," Acoust. Phys. **41**, 667–76 (1995).
[18]F. Dunn and W. Fry, "Ultrasonic absorption and reflection by lung tissue," Phys. Med. Biol. **5**, 401–410 (1961).
[19]T. J. Royston, Y. Yazicioglu, and F. Loth, "Surface response of a viscoelastic medium to subsurface acoustic sources with application to medical diagnosis," J. Acoust. Soc. Am. **113**, 1109–1121 (2003).
[20]C. P. Bradley, G. M. Harris, and A. J. Pullan, "The computational performance of a high-order coupled FEM/BEM procedure in electropotential problems," IEEE Trans. Biomed. Eng. **48**, 1238–1250 (2001).
[21]L. K. Cheng, J. M. Bodley, and A. J. Pullan, "Comparison of potential-and activation-based formulations for the inverse problem of electrocardiology," IEEE Trans. Biomed. Eng. **50**, 11–22 (2003).
[22]G. Fischer, B. Tilg, R. Modre, G. J. M. Huiskamp, J. Fetzer, W. Rucker, and P. Wach, "A bidomain model based BEM-FEM coupling formulation for anisotropic cardiac tissue," Ann. Biomed. Eng. **28**, 1229–1243 (2000).
[23]J. Lotjonen, I. E. Magnin, J. Nenonen, and T. Katila, "Reconstruction of 3-D geometry using 2-D profiles and a geometric prior model," IEEE Trans. Med. Imaging **18**, 992–1002 (1999).
[24]J. C. de Munck, T. J. C. Faes, and R. M. Heethaar, "The boundary element method in the forward and inverse problem of electrical impedance tomography," IEEE Trans. Biomed. Eng. **47**, 792–800 (2000).
[25]K. Pesola, J. Lotjonen, J. Nenonen, I. E. Magnin, K. Lauerma, R. Fenici, and T. Katila, "The effect of geometric and topologic differences in boundary element models on magnetocardiographic localization accuracy," IEEE Trans. Biomed. Eng. **47**, 1237–1247 (2000).
[26]Visible Human Project, http://www.nlm.nih.gov/research/visible/visible_human.html (2003).
[27]A. J. Burton and G. F. Miller, "The application of integral equation methods to the numerical solution of some exterior boundary value problems," Proc. R. Soc. London, Ser. A **A323**, 201–210 (1971).
[28]S. Kirkup, *The Boundary Element Method in Acoustics* (Integrated Sound Software, West Yorkshire, UK, 1998).
[29]S. M. Kirkup, "Fortran codes for computing discrete Helmholtz integral operators," Adv. Comput. Math. **9**, 391–409 (1998).
[30]S. Amini, P. J. Harris, and D. T. Wilton, *Coupled Boundary and Finite Element Methods for the solution of the dynamic fluid-structure interaction problem*, Lecture Notes in Engineering (Springer, Berlin, 1992).
[31]M. J. Hinich and E. J. Sullivan, "Maximum-likelihood passive localization using mode filtering," J. Acoust. Soc. Am. **85**, 214–219 (1989).
[32]A. N. Mirkin and L. H. Sibul, "Maximum likelihood estimation of the locations of multiple sources in an acoustic waveguide," J. Acoust. Soc. Am. **95**, 877–888 (1994).
[33]B. F. Harrison, R. J. Vaccaro, and D. W. Tufts, "Source localization in an acoustic waveguide with uncertain sound speed," *Proceedings of the IEEE International Conference on Acoustic, Speech, and Signal Processes*, Vol. **5**, 3115–3118 (1995).
[34]B. F. Harrison, R. J. Vaccaro, and D. W. Tufts, "Matched-field localization with many uncertain environmental parameters: Experimental data results," Proceedings of the IEEE International Conference on Acoustic, Speech, and Signal Processes (1996), Vol. **2**, 1193–1196.
[35]B. F. Harrison, "An inverse problem in underwater acoustic source localization: robust matched-field processing," Inverse Probl. **16**, 1641–1654 (2000).
[36]N. E. Collison and S. E. Dosso, "Regularized matched-mode processing for sound localization," J. Acoust. Soc. Am. **107**, 3089–3099 (2000).
[37]H. Krim and M. Viberg, "Two decades of array signal processing research," IEEE Signal Process. Mag. **13**, 67–94 (1996).
[38]G. R. Wodicka and D. C. Shannon, "Transfer function of sound transmission in subglottal human respiratory system at low frequencies," J. Appl. Physiol. **69**, 2126–2130 (1990).
[39]G. R. Wodicka, K. N. Stevens, H. L. Golub, and D. C. Shannon, "Spectral characteristics of sound transmission in the human respiratory system," IEEE Trans. Biomed. Eng. **37**, 1130–1135 (1990).
[40]G. R. Wodicka, A. Aguirre, P. D. DeFrain, and D. C. Shannon, "Phase delay of pulmonary acoustic transmission from trachea to chest wall," IEEE Trans. Biomed. Eng. **39**, 1053–1058 (1992).
[41]H. Pasterkamp, S. Patel, and G. R. Wodicka, "Asymmetry of respiratory sounds and thoracic transmission," Med. Biol. Eng. Comput. **35**, 103–106 (1997).
[42]S. S. Kraman, "Speed of low-frequency sound through lungs of normal men," J. Appl. Physiol. **55**, 1862–1867 (1983).
[43]M. Mahagnah and N. Gavriely, "Gas density does not affect pulmonary acoustic transmission in normal men," J. Appl. Physiol. **78**, 928–937 (1995).

Fundamental study on subharmonic imaging by irradiation of amplitude-modulated ultrasound waves

Norihide Maikusa,[a] Tadanori Fukami, Tetsuya Yuasa, and Yasutaka Tamura
Yamagata University 4-3-16 Jonan, Yonezawa-shi, Yamagata, 992-8510, Japan

Takao Akatsuka
Yamagata College of Industry and Technology, 3-57-4 Kyoda, Sakata-shi, Yamagata, 998-0102, Japan

(Received 8 August 2006; revised 27 April 2007; accepted 2 May 2007)

The second harmonic and subharmonic components, the frequencies of which are twice and one half the fundamental frequency, are included in echoes from contrast agents. An imaging method, which employs a second harmonic (second harmonic imaging), is widely used in medical diagnoses. On the other hand, subharmonic is expected to provide a higher contrast between biological tissues and blood flow because echo signals are generated only from blood containing the contrast agents. However, the subharmonic component echo signal power from contrast agents is relatively low. This has resulted in little progress in the field of subharmonic imaging. In this study, a new imaging method is proposed using amplitude-modulated waves as transmitted waves combined with the pulse inversion method to enhance subharmonic echo signals. Two optimal frequencies are set, including the modulated waves, F_1 and F_2, so that the subharmonic frequency of F_1 and the second harmonic frequency of F_2 may result in the same value. This allows a more powerful signal at the frequency band because the second harmonic and subharmonic components are integrated. Furthermore, a B-mode ultrasound image of an agar phantom that imitated biological tissue and showed the effectiveness of our method was reconstructed. As a result, the echo power of the subharmonic component was enhanced by approximately 11.8 dB more than the conventional method and the signal to noise ratio showed an improvement of 7.6 dB. [DOI: 10.1121/1.2743160]

PACS number(s): 43.80.Qf, 43.80.Vj, 43.80.Ev, 43.80.Jz [FD] Pages: 672–676

I. INTRODUCTION

The applications of ultrasound (US) contrast agents are numerous in medical diagnoses. Recently, these agents have been applied for many modalities and techniques such as contrast imaging, harmonic power Doppler imaging, and the pulse inversion method.

Most contrast agents consist of microbubbles with diameters from 1 to 10 μm to allow for passage of the lung capillaries. These microbubbles are filled with low soluble gases or air and are stabilized with a coating of surfactants or an encapsulating shell. Microbubbles enhance the backscattered signal, as an acoustic impedance of gas differs greatly from those of blood and biological tissue. Moreover, those backscattered signals from the microbubbles contain not only the fundamental driving frequency but also generate significant harmonics and subharmonics.[1] Harmonic imaging is a well-known technique using nonlinear and transient scattering, and the second harmonic from microbubbles is mainly used. This imaging utilizes the characteristic that more harmonics are obtained from microbubbles than from biological tissue, thus emphasizing the contrast between blood flow and biological tissue. A technique using high-intensity echoes obtained instantaneously by intermittent transmission, flash echo imaging, has also been developed.[2]

Shankar *et al.* showed that subharmonic generation from biological tissues is relatively lower than conventional second harmonic components.[3] Therefore, subharmonic imaging is expected to yield more contrast between biological issue and blood flow than conventional harmonic imaging. Recently, there have been many reports associated with subharmonic imaging. Chomas *et al.* reported an optimal transmission mode for subharmonic imaging based on nondestructive contrast agents with modified Rayleigh-Plesset analysis and optical experimental analysis of the microbubble oscillation and destruction.[4] They believe that an optimal nondestructive imaging technique should minimize relative expansion and wall velocity of the bubble, transmit a minimum number of cycles, and maintain a satisfactory contrast signal to noise ratio (S/N). Subharmonic imaging is a potential nondestructive technique that may satisfy these constraints. Bhagavatheeshwaran *et al.* reported the detection of subharmonic signals from contrast agents flowing through a tube, simulating tumor neovascularity, with internal diameters of less than 300 μm.[5] Forsberg *et al.* investigated the mechanism of subharmonic generation from contrast agents experimentally with a view towards practical use in US imaging.[6] Shankar *et al.* simulated the subharmonic oscillation of stationary state microbubbles with various surface properties using the RPNNP (Rayleigh, Plesset, Noltingk, Neppiras, and Poritsky) model.[7] However, Shi *et al.* believe that the new

[a]Electronic mail: aimiteno@eatpost.yz.yamagata-u.ac.jp

 0001-4966/2007/122(1)/672/5/$23.00

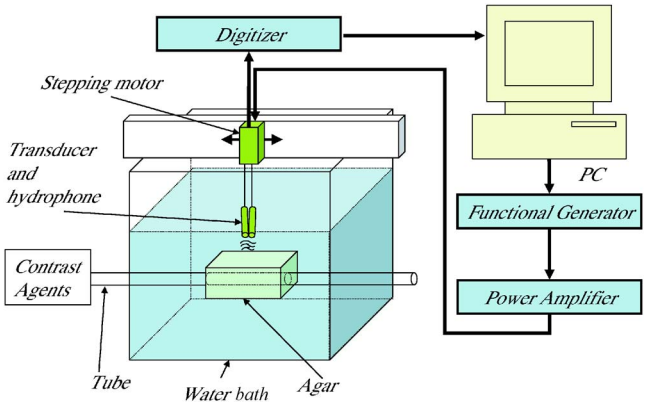


FIG. 1. (Color online) Imaging system.

technique should be developed to enhance subharmonic signals from contrast agents for maintaining a sufficient signal to noise ratio (S/N).[8]

To solve this problem, we proposed a new method to enhance signal intensity, which extracted not only subharmonics but also second harmonics in the frequency space and integrated two harmonic frequency components accordingly. We constructed a mechanical control B-mode imaging system and performed imaging of an agar phantom that imitated biological tissue and blood flow. As a result of comparing conventional subharmonic, second harmonic, and fundamental imaging with the proposed method, we showed that our method was superior in contrast to the conventional method.[9]

In this study, we used amplitude-modulated waves as transmitted pulses to combine subharmonic and second harmonic components. Moreover, we reconstructed B-mode US images of agar phantom, which imitates biological tissue, and showed the effectiveness of our method.

II. METHODS

A. Experimental setup

We constructed an imaging system to compare S/N between the proposed and conventional methods in B-mode US imaging (Fig. 1). We acquired US echo data from agar phantom in a water bath by probe. Amplitude-modulated waves were generated by a function generator (YOKOGAWA, WE-7121) and transmitted by a single-element transducer (Krautkramer, GAMMA-ISS) with a diameter of 0.50 in., −6-dB bandwidth of 64.7%, and a beamwidth of 2.5 mm under the condition that acoustic pressure was 550 kPa. Acoustic pressure was calibrated by needle hydrophone (GENEREX) before the experiment. An echo signal was received by wideband hydrophone (TORAY-Engineering, H9C) and sampled with 20 MHz and quantized with 12 bits per sample point by digitizer (YOKOGAWA, WE-7116). The position of the receive hydrophone was set for two echoes at maximum from two layers of an agar phantom. We extracted the part of the echo corresponding to the agar phantom in response to each transmitted pulse with the proposed and conventional methods. These data were analyzed by computer (PC). We then analyzed the data by fast-Fourier transform (FFT) with a window size of 1024 points. In frequency space, a sixth-order Butterworth filter with cutoff frequencies from 1.6 to 1.8 MHz, 3.3 to 3.5 MHz, and 6.7 to 6.9 MHz was employed to extract the three frequency components, subharmonic, fundamental, and second harmonic, respectively. We subsequently transformed them to an A-mode waveform by inverse FFT, and reconstructed B-mode US images normalized to 8 bits grayscale.

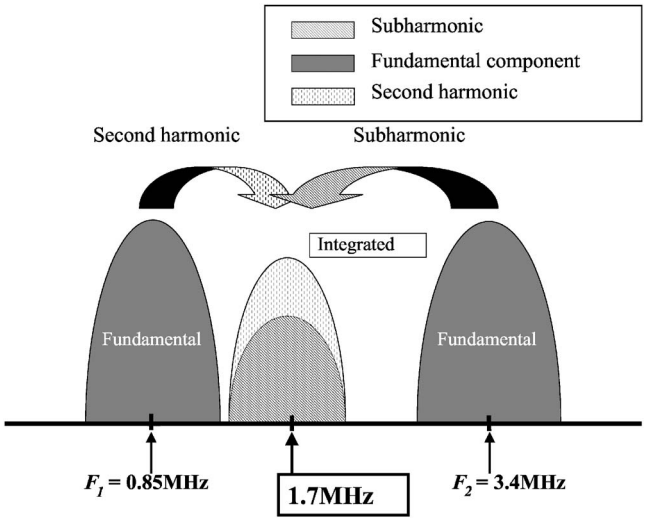


FIG. 2. Explanatory diagram of the proposed method.

B. Transmitter pulses

1. Amplitude-modulated wave

In this study, we used amplitude-modulated waves to enhance subharmonic signal intensity by integrating a second harmonic component. We assumed that f_1 and f_2 ($f_1 > f_2$) are frequencies of carrier and modulation, respectively. This modulated wave includes two frequency components f_1+f_2 and f_1-f_2. We defined f_1+f_2 as F_1 and f_1-f_2 as F_2. We set f_1 and f_2 so that the subharmonic frequency of F_1 and the second harmonic frequency of F_2 may result in the same value as shown in Fig. 2. Our method will enable a more powerful signal at the frequency band of $F_1/2$ because the second harmonic and subharmonic component are integrated. In this frequency band, not only the subharmonic but also the second harmonic was included. The second harmonic was also generated from biological tissue. However, we believe that we can create an image with good contrast between biological tissue and blood flow because second harmonic generation from biological tissue is relatively low, and we combined the subharmonic component in this frequency, which is generated only from contrast agents. In this study, we set f_1 and f_2 to 2.125 and 1.275 MHz, respectively, corresponding to F_1 for 3.4 MHz and F_2 for 0.85 MHz. The reason why we set those values is that the resonance frequency of the contrast agent, Optison® (Molecular Biosystems), is approximately 2.0 MHz, and it is known that the subharmonic echo can acquire enough subharmonic signal when microbubbles are irradiated at twice the resonance frequency.[4]

2. Pulse-inversion method

A scattered signal includes various frequency components such as the subharmonic frequency of F_1 fundamental and second harmonic frequencies F_2 and more high-order harmonic components of F_1 and F_2. We set the fundamental component of F_2 and subharmonic frequency of F_1 to 0.85 and 1.7 MHz, respectively. These frequencies are close to each other and each frequency component shows a broadening in the spectrum. Thus, we have a problem that the two spectrums overlap around 1.7 MHz. To solve this problem, we used the pulse inversion method to remove the fundamental frequency component of F_2, 0.85 MHz.

The pulse inversion method uses a sequence of two ultrasound pulses. The second pulse is transmitted after a suitable delay and is an inverted replica of the first pulse. The fundamental component of the second pulse is inverted compared to the first pulse, and therefore the sums of the two responses are canceled out. The second harmonic, however, will not be inverted copies. The sum is not zero and the remainder is related to the degree of nonlinearity.[2]

In this study, we used the first pulse, y_1, and the second pulse, y_2, as follows:

$$\begin{cases} y_1(t) = \sin(2\pi f_1 t) \cdot \cos(2\pi f_2 t) \\ y_2(t) = \cos(2\pi f_1 t) \cdot \sin(2\pi f_2 t) \end{cases} \quad (0 \leqslant t \leqslant n\lambda/f_1), \tag{1}$$

where λ is wavelength and n is burst count. We set n to 12. We could expand Eq. (1) to Eq. (2),

$$\begin{cases} y_1(t) = \dfrac{1}{2}\sin(2\pi(F_1)t) + \dfrac{1}{2}\sin(2\pi(F_2)t) \\ y_2(t) = \dfrac{1}{2}\sin(2\pi(F_1)t) - \dfrac{1}{2}\sin(2\pi(F_2)t). \end{cases} \tag{2}$$

From Eq. (2), the fundamental component of F_1 in the second pulse is inverted compared to the first pulse. Therefore, we removed the fundamental component of F_1, 0.85 MHz, by the sum of the two responses.

We defined the conventional method and proposed method as an irradiation of a sinusoidal wave without the pulse inversion method and an irradiation of a modulation wave, respectively. Moreover, subharmonic, fundamental, and second harmonic imaging was achieved using 1.7-, 3.4-, and 6.8-MHz frequency components, respectively. In the proposed method, the pulse inversion method was carried out only at the 1.7-MHz frequency component.

3. Structure of the agar phantom

We constructed an agar phantom for the simulated experiments to compare signal intensity of various frequency components from contrast agents and imitated biological tissues by the conventional and proposed methods (Fig. 3). The phantom penetrated 5.0-mm holes at 20.0- and 45.0-mm depths. We created the agar mixed with injected polyvinyl chloride powder (Aldrich) at a concentration of 1.0 mg/ml, as imitation biological tissue, into the upper hole. The contrast agent, Optison®, was diluted to 100 μl/ml into the lower hole (Fig. 3). The reason for this placement was the consideration that blood flow is located within a biological tissue, and thus we set the contrast agent layer under the polyvinyl chloride layer. In addition, the above-mentioned acoustic pressure calibration was carried out in a contrast agent layer. The attenuation and velocity of mixed agar were −3.02 dB/cm and 1440 m/s, respectively, at 20 deg as measured by the transmission method. These values were similar to biological tissue, and thus we believed that it could be used to imitate biological tissue characteristics. Moreover, the subharmonic and second harmonic generations compared to the fundamental component from polyvinyl chloride were −40.0 and −20.6 dB, respectively.

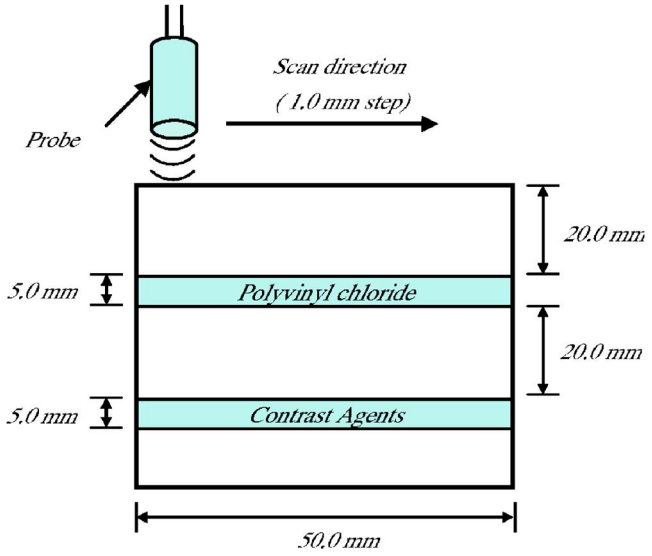


FIG. 3. (Color online) Structure of the phantom for B-mode ultrasound imaging.

The transducer was used to transmit US pulses and we scanned the agar 50 times by stepping motor (Sigma-Koki, Mark-102). We set the scanning interval to 1.0 mm because the beamwidth of the transducer that we used was 2.5 mm, and thus we set the scanning interval smaller than the beamwidth. The US echo signals from the phantom were acquired by using a wideband hydrophone and averaged ten times.

We reconstructed B-mode US images and compared the signal intensity of various frequency components between the conventional and proposed methods in the region of the contrast agent and imitated biological tissues. We evaluated the performance of our proposed method with S/N. This S/N was defined as the ratio of the echo signals from the contrast agent (signal) and the polyvinyl chloride layer (noise).

III. RESULTS AND DISCUSSION

A. Enhancement of the subharmonic component

First, we showed the power spectrum in the region of the contrast agent to compare enhancement of the subharmonic component in the proposed method to that in the conventional method (Fig. 3). In Fig. 3, we normalized two power spectrums to 0 dB. We assumed that the subharmonic component was 1.7 MHz, the fundamental component was 3.4 MHz, and the second harmonic component was 6.8 MHz. It was found that the fundamental component of

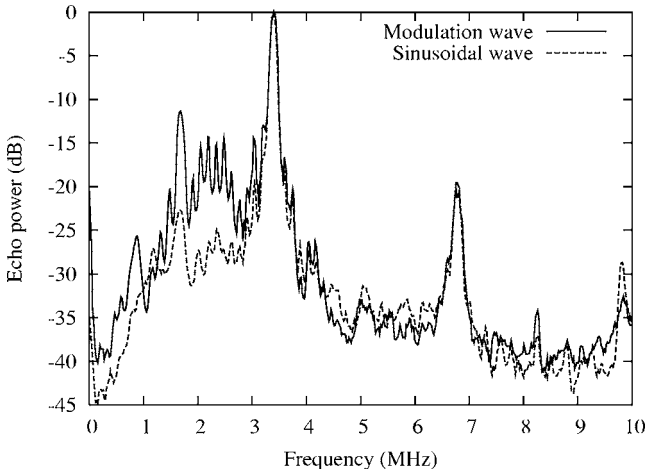


FIG. 4. Power spectrum of the ultrasound echo from the phantom. The solid line indicates the results of the proposed method and the dashed line represents the conventional method.

F_2, approximately 0.85 MHz, was effectively canceled. We also compared the power spectrum of the subharmonic component between the proposed and conventional methods. After use of the proposed method, the fundamental and second harmonic components showed a similar value to that obtained with conventional methods. Nevertheless, the value of the subharmonic component increased approximately 11.8 dB, as shown in Fig. 4.

B. Two-dimensional US imaging

We compared the results of B-mode US images from the agar phantom between the conventional and proposed methods. The results of B-mode US images of subharmonic, fundamental, and second harmonic imaging using the proposed method are shown in Figs. 5(a)–5(c), and Fig. 5(d)–5(f) are the results of the conventional method.

We could ascertain that the B-mode image shown in Fig. 5 reflected the structure of the phantom. From fundamental and second harmonic images shown in Figs. 5(b), 5(c), 5(e), and 5(f), we could clearly visualize the polyvinyl layer. This result was not desirable because we assumed that the polyvinyl layer was an imitation of biological tissue. Thus, it deteriorates contrast between blood flow and biological tissue. The polyvinyl layer did not appear using subharmonic imaging as shown in Figs. 5(a) and 5(d). However, enough signal intensity of the subharmonic component could be obtained when we used the conventional method. As shown in Fig. 5(a), we could clearly recognize the region of the contrast agent layer. These results showed that the proposed method can effectively increase the signal intensity of the subharmonic component. As shown in Fig. 5(a), a better contrast between biological tissue and blood flow was acquired in the subharmonic image obtained by the proposed method.

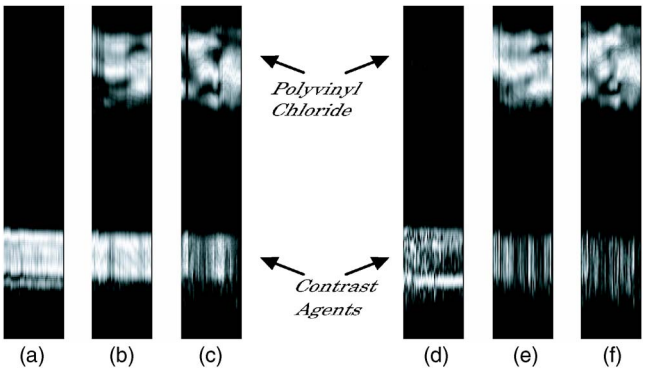


FIG. 5. (Color online) Results of each imaging method. (a) Subharmonic imaging; (b) fundamental wave imaging; and (c) second harmonic imaging with the proposed method; (d) subharmonic imaging; (e) fundamental wave imaging; and (f) second harmonic imaging with the conventional method.

TABLE I. Signal-to-noise ratio of each imaging method.

Imaging method	S/N [dB]
Proposed method (subharmonic)	31.38
Proposed method (fundamental wave)	6.06
Proposed method (second harmonic)	0.15
Conventional method (subharmonic)	23.75
Conventional method (fundamental)	8.74
Conventional method (second harmonic)	3.53

Finally, we calculated the S/N of subharmonic imaging for comparing the efficiency of the conventional and proposed methods quantitatively (Table I). After use of the proposed method, the S/N of the subharmonic component showed a 7.6-dB improvement compared to the conventional method.

From our results, the signal intensity of the subharmonic component was increased by the proposed method. Therefore, the proposed method improved the S/N of the subharmonic image more than the conventional method.

IV. CONCLUSIONS

We proposed an imaging method using modulated waves, which included two frequency spectra. By using the proposed method, we combined subharmonic and second harmonic components in echo. The results showed that the echo power of the subharmonic component increased approximately 11.8 dB more than the conventional method. Using the proposed method, the S/N of the subharmonic B-mode imaging was enhanced by 7.6 dB more than the conventional method. These results demonstrated that our method improved image contrast between biological tissue and blood flow compared to the conventional method. Our experiments were carried out at an acoustic pressure of 550 kPa, and we believe that using higher pressures could have interesting results. However, we could not perform a higher acoustic pressure because of the limitation of the device. Therefore, we hope to perform this condition in the future.

[1]A. Eller and H. G. Flynn, "Generation of subharmonics of order one-half by bubbles in a sound field," J. Acoust. Soc. Am. **46**, No. 3(part 2), 722–727 (1969).
[2]P. J. A. Frinking, A. Bouakaz, J. Kirkhorn, F. J. Ten Cate, and N. De Joung, "Ultrasound contrast imaging: Current and new potential methods," Ultrasound Med. Biol. **26**, No. 6, 965–975 (2000).
[3]P. M. Shankar, P. D. Krishna, and V. L. Newhouse, "Advantages of subharmonic over second harmonic backscatter for contrast to tissue echo enhancement," Ultrasound Med. Biol. **24**, No. 3, 395–399 (1998).
[4]J. Chomas, P. Dayton, D. May, and K. Ferrara, "Non-destructive subhar-

monic imaging," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **49**, No. 7, 883–892 (2002).
[5]G. Bhagavatheeshwaran, W. T. Shi, F. Forsberg, and P. M. Shankar, "Subharmonic signal generation from contrast agents in simulated neovessels," Ultrasound Med. Biol. **30**, No. 2, 199–203 (2004).
[6]F. Forsberg, W. T. Shi, and B. B. Goldberg, "Subharmonic imaging of contrast agents," Ultrasonics **38**, 93–98 (2000).
[7]P. M. Shankar, P. D. Krishna, and V. L. Newhouse, "Subharmonic backscattering from ultrasound contrast agents," Phys. Med. Biol. **44**, 681–693 (1999).
[8]W. T. Shi, F. Forsberg, A. L. Hall, R. Y. Chiao, J. Liu, S. Miller, K. E. Thomenius, M. A. Wheatley, and B. B. Godberg, "Subharmonic imaging with microbubble contrast agents: Initial results," Ultrason. Imaging **21**, 79–94 (1999).
[9]N. Maikusa, A. Yagi, T. Fukami, T. Yuasa, Y. Tamura, and A. Takao, "Subharmonic imaging considering the second harmonic component," Med. Imaging Technol. **23**, No. 4, 203–210 (2005).

Erratum: Scattering and active acoustic control from a submerged spherical shell [J. Acoust. Soc. Am. 111(2), 893–907 (2002)]

Clyde Scandrett
Naval Postgraduate School, Monterey, California 93943

(Received 26 March 2007; accepted 2 April 2007)

[DOI: 10.1121/1.2735103]

PACS number(s): 43.40.Fz, 43.40.Rj, 43.40.Vn, 43.10.Vx [DF]

The original paper has an error in one of the intermediate terms for expressions involving the piezoelectric material. It is the β term found in Eq. (22) of the original paper. The expression for β used a "difference of terms" rather than a sum. Its correct value should be

$$\beta = (c_{13}\epsilon_{33} + e_{31}e_{33})/\alpha.$$

Correcting the sign error leads to small changes in Tables II and IV, as well as in all figures that involve the material PZT4 (Figs. 6 though 11).

In addition, revised numbers for modal voltages in the fifth paragraph of the results section become: $V_0=0.0075$, $V_1=0.0150$, $V_2=0.4237$, $V_3=0.3650$, $V_4=0.1705$, $V_5=0.0803$, $V_6=0.0400$, $V_7=0.0136$, $V_8=0.0060$, and $V_9=0.0059$. Note also that Fig. 9 displayed results of scattering from a point source while this paragraph refers to scattering from a plane wave.